

An abstract painting with a complex, layered composition. The background is a mix of dark, muted colors like purples, blues, and greys. Overlaid on this are various shapes and textures in brighter colors: a large, bright pink shape in the upper left, a yellow shape in the upper right, and a large, light-colored, almost white shape in the center. A jagged, orange line runs diagonally across the image, starting from the bottom left and ending near the top right, creating a sense of movement and division. The overall style is expressive and painterly.

Outils d'estimation démographique

**Tom Moultrie, Rob Dorrington, Allan Hill,
Kenneth Hill, Ian Timæus and Basia Zaba**

AAKING

Outils d'estimation démographique

Outils d'estimation démographique

Tom Moultrie, Rob Dorrington, Allan Hill,
Kenneth Hill, Ian Timæus and Basia Zaba



© 2017 International Union for the Scientific Study of Population (IUSSP)
3-5 rue Nicolas, 75980 Paris Cedex 20, France
www.iussp.org

Le contenu de ce volume est distribué sous licence Creative Commons (Paternité-Pas d'Utilisation Commerciale-Partage des Conditions Initiales à l'Identique 3.0 Unported).

Le texte complet de cette licence figure en anglais sur la page <https://creativecommons.org/licenses/by-nc-sa/3.0/legalcode>. Un résumé en français des principales dispositions est disponible sur la page <https://creativecommons.org/licenses/by-nc-sa/3.0/deed.fr>. En résumé, cette licence implique que vous êtes libre d'adapter, de copier, de distribuer ou de transmettre cet ouvrage, à condition de le créditer de la manière indiquée ci-dessous, et sans suggérer que les auteurs ou détenteurs du copyright vous soutiennent ou soutiennent la façon dont vous avez utilisé leur ouvrage ; que vous ne pouvez pas utiliser cet ouvrage à des fins commerciales (c'est-à-dire le revendre, en entier ou en partie, pour un profit) ; et que si vous modifiez, transformez ou créez à partir de cet ouvrage, vous ne pouvez distribuer l'ouvrage résultant que sous la même licence ou une licence similaire à celle-ci.

La citation appropriée pour ce travail est la suivante :

Moultrie TA, RE Dorrington, AG Hill, K Hill, IM Timæus et B Zaba (eds). 2017.

Outils pour l'estimation démographique. Paris: International Union for the Scientific Study of Population. demographicestimation.iussp.org

Publié en anglais pour la première fois en 2013.

ISBN: 978-2-87108-073-2

Image de couverture : « The Journey ». 2012. Makiwa Mutomba. Huile sur toile. 56 x 71 cm.
Reproduit avec la permission de l'artiste.

Conception et composition : User Friendly, Le Cap, Afrique du Sud, et pour la version française Desk, Saint-Berthevin, France.

Imprimé et relié par Paarl Media Paarl, 15 janvier van Riebeeck Drive, Paarl, Afrique du Sud

SOMMAIRE

| | |
|---|------------|
| Introduction | iii |
| ESTIMATION ET ÉVALUATION DE LA QUALITÉ DES DONNÉES | 1 |
| Chapitre 1 Évaluation générale de la qualité des données par âge et sexe | 3 |
| FÉCONDITÉ | 23 |
| Chapitre 2 Introduction à l'analyse de la fécondité | 25 |
| Évaluation et correction des données de fécondité | 29 |
| Introduction | 30 |
| Chapitre 3 Évaluation de la qualité des données sur les descendance | 31 |
| Chapitre 4 La correction d'El-Badry | 37 |
| Chapitre 5 Évaluation de la qualité des données sur la fécondité récente tirées des recensements ... | 44 |
| Méthodes fondées sur un seul recensement | 53 |
| Chapitre 6 Vue d'ensemble sur les méthodes d'estimation de la fécondité fondées sur le quotient P/F .. | 54 |
| Chapitre 7 Le modèle relationnel de Gompertz | 56 |
| Chapitre 8 Probabilités d'agrandissement des familles | 72 |
| Chapitre 9 Estimation de la fécondité par la survie inverse | 86 |
| Méthodes fondées sur plusieurs recensements | 99 |
| Chapitre 10 Modèles relationnels synthétiques de Gompertz | 100 |
| Chapitre 11 Estimation de la fécondité à partir des accroissements de parité des cohortes | 107 |
| Méthodes fondées sur des données d'enquêtes | 113 |
| Chapitre 12 Estimation directe de la fécondité à partir de données d'enquêtes contenant des histoires génésiques | 114 |
| Chapitre 13 Fécondité par cohorte et période | 122 |
| Méthodes fondées sur les données d'état civil | 133 |
| Chapitre 14 Comparaison des parités par cohorte et des données d'état civils | 134 |
| LA MORTALITÉ DES JEUNES ENFANTS | 145 |
| Chapitre 15 Introduction à l'analyse de la mortalité des jeunes enfants | 147 |
| Méthodes fondées sur un seul recensement | 155 |
| Chapitre 16 Estimation indirecte de la mortalité des jeunes enfants | 156 |

| | |
|---|------------|
| Méthodes fondées sur des données d'enquêtes | 175 |
| Chapitre 17 Estimation directe de la mortalité des jeunes enfants à partir d'histoires génésiques .. | 176 |
| Chapitre 18 Estimation de la mortalité des jeunes enfants à partir des données des établissements de santé : la technique de la naissance précédente | 189 |
| LA MORTALITÉ ADULTE | 201 |
| Chapitre 19 Introduction à l'analyse de la mortalité adulte | 203 |
| Méthodes fondées sur un seul recensement | 207 |
| Chapitre 20 La méthode de la balance de l'accroissement démographique de Brass | 208 |
| Chapitre 21 La méthode de Preston et Coale | 223 |
| Chapitre 22 Estimation indirecte de la mortalité adulte à partir des proportions d'orphelins | 238 |
| Chapitre 23 Estimation indirecte de la mortalité adulte à partir des données sur les fratries | 261 |
| Méthodes fondées sur plusieurs recensements | 277 |
| Chapitre 24 La méthode généralisée de la balance de l'accroissement démographique | 278 |
| Chapitre 25 Méthode de l'extinction des cohortes synthétiques | 297 |
| Chapitre 26 Estimation indirecte de la mortalité adulte à partir des proportions d'orphelins issues de plusieurs enquêtes ou recensements | 317 |
| Les données d'enquêtes et les méthodes directes | 335 |
| Chapitre 27 Estimation de la mortalité adulte à partir de la survie des frères et sœurs | 337 |
| La mortalité maternelle | 351 |
| Chapitre 28 Introduction à l'analyse de la mortalité maternelle | 353 |
| Chapitre 29 Estimation de la mortalité liée à la grossesse à partir de la survie des frères et sœurs .. | 358 |
| Chapitre 30 Estimation de la mortalité liée à la grossesse à partir des décès survenus dans les ménages | 364 |
| UTILISER DES MODÈLES POUR OBTENIR DES TABLES DE MORTALITÉ À PARTIR DE DONNÉES INCOMPLÈTES | 369 |
| Chapitre 31 Introduction aux tables-types de mortalité | 371 |
| Chapitre 32 Ajuster une table-type de mortalité à une estimation de la mortalité des enfants et une estimation de la mortalité des adultes | 378 |
| Chapitre 33 Combiner des estimations indirectes de la mortalité des enfants et de la mortalité adulte pour produire une table de mortalité | 387 |
| MIGRATION | 399 |
| Chapitre 34 Introduction à l'analyse des migrations | 401 |
| Chapitre 35 Estimation de la migration à partir de données de recensement | 406 |
| Chapitre 36 Le modèle multi-exponentiel de migration | 421 |
| Chapitre 37 Les modèles log-linéaires de flux migratoires | 435 |

INTRODUCTION

OUTILS D'ESTIMATION DÉMOGRAPHIQUE résulte d'un projet, financé par le Fonds des Nations Unies pour la Population (Fnuap) et conduit sous les auspices de l'Union Internationale pour l'Étude Scientifique de la Population (UIESP), afin de regrouper en un lieu unique, et dans un style convivial, les principales méthodes utilisées par les démographes à travers le monde pour mesurer les paramètres démographiques à partir de données limitées et défectueuses.

L'idée d'un tel ouvrage a émergé lors d'une réunion conjointe UIESP/Fnuap sur la « Formation démographique appliquée et technique dans les pays en développement » à La Haye en mars 2009, où une préoccupation est apparue concernant le déclin de la formation des démographes à l'usage et l'application des techniques d'estimation indirecte dans la plupart des institutions universitaires du monde.

Plusieurs facteurs ont contribué à cette situation. En premier lieu, la modification des priorités concernant la population mondiale, en particulier l'ordre du jour révisé adopté par la Conférence Internationale sur la Population et le Développement au Caire en 1994, avait profondément transformé le panorama financier, davantage de ressources étant consacrées aux domaines émergents de la santé reproductive et sexuelle plutôt qu'à la démographie technique nécessaire à l'étude des schémas de croissance et à la gestion de la croissance des populations. En parallèle, la génération des démographes qui avaient été formés aux méthodes et aux techniques classiques vieillissait rapidement et rares étaient les jeunes démographes formés à la science ou à l'art de l'estimation démographique à partir de données limitées et défectueuses.

En deuxième lieu, le programme des Enquêtes Démographiques et de Santé (EDS), en collectant des histoires fécondes complètes qui permettent le recours à des méthodes d'estimation directe de la fécondité et de la mortalité, a donné l'impression que les outils et les techniques d'estimation de la mortalité et de la fécondité à partir des données de recensements ou d'autres enquêtes

n'étaient plus aussi importants qu'ils l'avaient été dans le passé. Il est certain que les EDS ont énormément contribué à la discipline démographique et à sa réorganisation, mais la marginalisation croissante de l'analyse démographique des données de recensement et des autres matériaux démographiques réduit notre capacité à comprendre la dynamique démographique des pays en développement. Le rôle du recensement comme base de sondage pour les enquêtes démographiques est souvent oublié. En outre, la taille de l'échantillon dans la plupart des EDS est telle que des estimations précises sont rarement possibles avec de telles enquêtes à des niveaux géographiques plus fins que les régions ou les provinces, et l'information recueillie sur des événements relativement rares (par exemple les décès d'adultes) est généralement trop pauvre pour que des estimations robustes puissent en être tirées.

En troisième lieu, dans la plupart des régions en développement (l'Afrique subsaharienne étant une exception notable), des améliorations dans les systèmes d'enregistrement de l'état civil et dans le recueil de données démographiques par les recensements font que les techniques existantes d'estimation démographique à partir de données limitées et défectueuses sont considérées comme obsolètes. C'est certainement le cas dans les pays où l'enregistrement des événements d'état civil est complet et précis et où existe une série de recensements fiables : une estimation directe et continue des paramètres démographiques y devient possible. Mais dans de nombreux pays ayant un revenu bas ou moyen, aucune de ces conditions ne prévaut encore et il reste donc important d'examiner de façon critique la qualité des statistiques fondées sur l'enregistrement à l'état civil et de les recouper avec des questions sur la fécondité et la mortalité tirées des recensements.

Une autre raison du recul de la priorité accordée à l'enseignement des techniques indirectes d'estimation démographique est l'évolution naturelle des populations où, même dans les pays les plus pauvres, la fécondité baisse après plusieurs décennies de progrès dans la lutte contre la mortalité.

Les distributions par âge de ces populations sont donc loin du modèle théorique des populations stables ou même quasi-stables, de sorte que bien des techniques s'appuyant sur ces modèles et publiées dans Division de la Population des Nations Unies (1967) ne peuvent plus avoir cours aujourd'hui. La disparition de ce qu'on appelait l'analyse de la population stable a conduit certains analystes à préférer des méthodes d'estimation directe de style EDS plutôt que l'ensemble des méthodes développées initialement par Ansley J. Coale et William Brass, les auteurs des premiers volumes des Nations Unies.

Dans bien des cas, une estimation démographique directe à partir des données de recensement, d'enquête ou d'état civil reste impossible ou problématique. Il subsiste donc un besoin d'estimations indirectes fondées sur le recensement ou autres sources. La réunion de 2009 a aussi noté que le manuel de référence pour l'estimation démographique à partir des données de recensement, *Manuel X* (Division de la Population des Nations Unies 1984) date de plus d'un quart de siècle et que diverses méthodes et techniques nouvelles ont été développées depuis sa publication. Deux autres manuels ont été préparés depuis, *Estimating Demographic Parameters from Census Data* (Sloggett, Brass, Eldridge et al. 1994) et *Methods for Estimating Adult Mortality* (UN Population Division 2002), mais aucun ne fait une révision et une mise à jour complètes et approfondies du *Manuel X*.

C'est pourquoi la réunion de La Haye a décidé de lancer un projet de révision et de mise à jour du *Manuel X*. À la suite d'un appel concurrentiel à propositions évalué par l'UIESP, un contrat a été passé avec un consortium de démographes de l'Université du Cap et de la London School of Hygiene and Tropical Medicine et de démographes indépendants associés à l'Université de Harvard pour réaliser le projet. Les *Outils d'estimation démographique* en sont le résultat.

L'ouvrage présenté ici descend en ligne directe du *Manuel X* et la logique sous-tendant le travail est fondamentalement la même – exposer les méthodes d'estimation des paramètres démographiques à partir de données limitées et défectueuses. C'est pourquoi nous conseillons instamment aux utilisateurs des *Outils d'estimation démographique* de lire le chapitre introductif du *Manuel X* (disponible sur le site internet de la Division de la Population des Nations Unies), à la fois pour son histoire des méthodes d'estimation indirecte, pour son plaidoyer en leur faveur et pour sa discussion des limitations inhérentes à des ouvrages de référence de cette sorte.

Les *Outils d'estimation démographique* divergent de leurs précurseurs sur plusieurs points importants. Les différences proviennent en partie de l'énorme accroissement de la puissance de calcul à la disposition des analystes depuis que le *Manuel X* a été publié. Elles reflètent aussi des avancées dans les approches de l'estimation démographique, des nouvelles méthodes et une évolution des idées sur le bon fonctionnement des différentes méthodes et les conditions nécessaires à ce bon fonctionnement. Les méthodes décrites dans les *Outils d'estimation démographique* et dans les manuels précédents ne sont donc pas les mêmes. Des méthodes qui ont été développées depuis la publication du *Manuel X* sont présentées ici pour la première fois. D'autres méthodes, qui étaient présentées dans le *Manuel X*, ont été exclues parce qu'il a été constaté depuis qu'elles fonctionnaient mal ou parce que des méthodes plus élaborées ou plus innovantes les rendent obsolètes.

Deuxièmement, à la différence de ses précurseurs, les *Outils d'estimation démographique* sont avant tout une ressource électronique disponible sur internet. La version imprimée représente le contenu du site internet (demographicestimation.iussp.org) à la date d'impression. Le site internet est organisé pour être dynamique, mis à jour et évolutif au fil du temps. C'est pourquoi, chaque fois que possible, le point de référence du lecteur devrait être le site internet, plutôt que la version imprimée du manuel. Le site internet, hébergé par l'UIESP, est librement et gratuitement accessible à toute personne enregistrée.

Troisièmement, le site internet inclut des feuilles de calcul téléchargeables qui mettent en œuvre les méthodes décrites, de sorte à faciliter leur application et leur utilisation. La décision de mettre en œuvre les méthodes par l'utilisation de feuilles de calcul plutôt que sous forme de programmes exécutables téléchargeables (tels que *MortPak* par exemple) est destinée à assurer un degré maximal de transparence. Les formules et les calculs sont visibles pour l'utilisateur final, et les feuilles de calcul peuvent être modifiées par les utilisateurs si elles ne correspondent pas exactement aux données disponibles. Les feuilles de calcul sont dans le format de Microsoft Excel, mais elles ont été conçues pour être compatibles avec d'autres tableurs en libre accès. Ce n'est qu'exceptionnellement qu'on a utilisé des commandes propres à Excel (par exemple *Solveur*).

Une quatrième différence avec les précédents manuels d'estimation indirecte est que, alors que *Outils d'estimation démographique* adopte à peu près la même démarche que ses

précurseurs en décrivant étape par étape la manière d'appliquer les méthodes analysées, un accent plus fort a été mis sur l'exposé des hypothèses qui sous-tendent chaque méthode, ainsi que sur les situations et conditions dans lesquelles les méthodes peuvent être contre-indiquées ou peuvent donner des résultats non fiables. Pour aider les utilisateurs intéressés à comprendre comment les méthodes fonctionnent, nous nous sommes efforcés de présenter la formalisation mathématique des méthodes dans un style aussi accessible que possible.

Cinquièmement, *Outils d'estimation démographique* inclut un développement sur l'évaluation et la mesure de la migration à partir des données de recensement, un domaine que le *Manuel X* n'avait pas du tout couvert, et qui n'avait pas été traité dans un ouvrage de ce type depuis le *Manuel VI* (Division de la population des Nations unies 1970).

Malgré ces avancées, le présent ouvrage souffre sur bien des points des mêmes limitations que ses précurseurs. La présentation de chaque méthode séparément fait perdre trop souvent la vue d'ensemble de l'estimation démographique à partir de données limitées et défectueuses. Une composante importante de ce travail consiste à assembler les pièces d'un puzzle composé de paramètres démographiques à l'aide de méthodes et de sources multiples pour en faire un tout doté de cohérence et de consistance interne. Le type d'estimation démographique présenté ici relève, en fin de compte, autant de l'art que de la science. Autant que possible, nous avons cherché à donner un sens à l'art en question. Pour faciliter et encourager une application attentive des méthodes décrites ici, le site internet inclut aussi un forum de discussion, dont nous espérons qu'il fournira un lieu où seront discutés les résultats tirés de l'application des méthodes, suggérées des

modifications ou des corrections à apporter aux méthodes existantes et proposées de nouvelles approches de l'estimation démographique à partir de données limitées et défectueuses.

Outils d'estimation démographique est le fruit d'une longue préparation. Les éditeurs remercient les nombreuses personnes et organisations qui ont aidé à amener ce projet à son terme. Notons les contributions de Ralph Hakkert (FNUAP) et Mary Ellen Zuppan (UIESP) à l'obtention des financements nécessaires et à la supervision du projet ; des lecteurs anonymes choisis par l'UIESP qui ont fait de nombreux commentaires utiles sur la première version du texte ; des concepteurs du site internet (Charles Oertel) et du livre (Jo-Anne Friedlander) ; et de la correctrice des épreuves (Debbie Budlender). Nous sommes aussi extrêmement reconnaissants aux responsables des Manuels des Nations Unies et de l'Institut Statistique pour l'Asie et le Pacifique d'avoir renoncé à leurs droits d'auteur et de nous avoir autorisés à reproduire des matériaux tirés de ces sources quand c'était nécessaire.

Tom Moultrie, Rob Dorrington, Allan Hill, Kenneth Hill, Ian Timæus and Basia Zaba

Le Cap, juillet 2013

La traduction française des *Outils pour l'estimation démographique* a été rendue possible grâce à des financements de la Fondation William et Flora Hewlett et de l'UNFPA. Nous remercions également Patrick Festy, Eric Vilquin, Michel Garenne et Richard Dackham Ngatchou pour leurs traductions ainsi que Bruno Schoumaker et Bruno Masquelier pour leur relecture des chapitres traduits.

RÉFÉRENCES

Division de la Population des Nations Unies. 1967. *Manuel IV : Méthodes permettant d'estimer les mesures démographiques fondamentales à partir des données incomplètes*. New York : Nations Unies, Département des affaires économiques et sociales internationales, ST/ESA/SER.A/42. version anglaise : <http://www.un.org/esa/population/techcoop/DemEst/manual4/manual4.html>

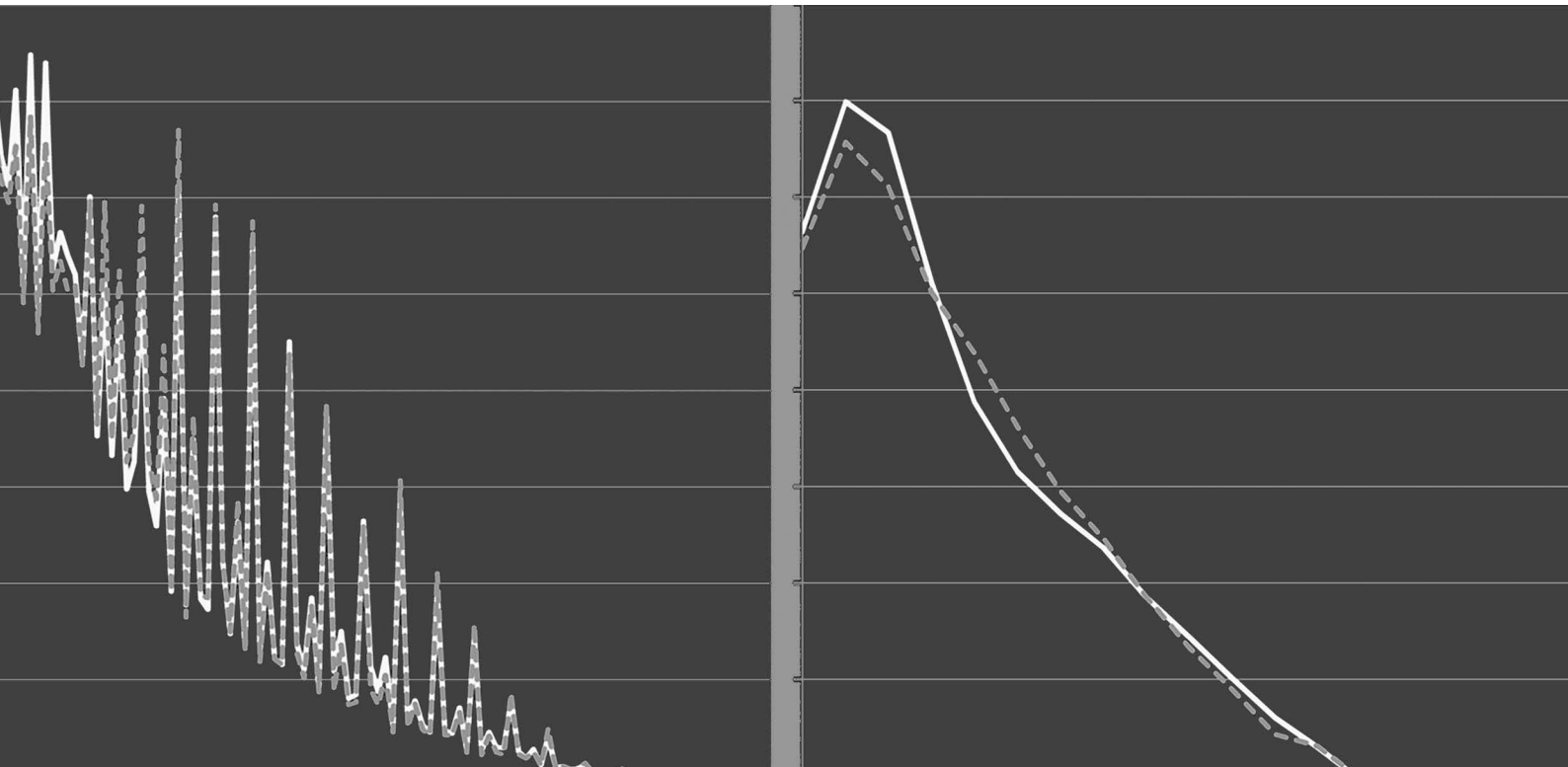
Division de la Population des Nations Unies. 1971. *Manuel VI : Méthodes de mesure de la mobilité interne*. New York : Nations Unies, Département des affaires économiques et sociales internationales, ST/ESA/SER.A/47. version anglaise : <http://www.un.org/esa/population/techcoop/IntMig/manual6/manual6.html>

Division de la Population des Nations Unies. 1984. *Manuel X. Techniques indirectes d'estimation démographique*. New York : Nations Unies, Département des affaires économiques et sociales internationales, ST/ESA/SER.A/81. http://unstats.un.org/unsd/demographic/standmeth/handbooks/Manuel_X-fr.pdf

Sloggett A, W Brass, SM Eldridge, IM Timæus, P Ward and B Zaba (eds). 1994. *Estimation of Demographic Parameters from Census Data*. Tokyo: Statistical Institute for Asia and the Pacific.

UN Population Division. 2002. *Methods for Estimating Adult Mortality*. New York: United Nations, Department of Economic and Social Affairs, ESA/P/WP.175. http://www.un.org/esa/population/techcoop/DemEst/methods_adultmort/methods_adultmort.html

Estimation et évaluation de la qualité des données



CHAPITRE I ÉVALUATION GÉNÉRALE DE LA QUALITÉ DES DONNÉES PAR ÂGE ET SEXE

Tom A Moultrie

INTRODUCTION

« Dans un monde idéal, les données seraient toujours complètes, exactes, actuelles, pertinentes et sans ambiguïtés. Dans le monde réel, elles sont généralement imparfaites sur l'une ou sur l'ensemble de ces dimensions » (Feeney 2003 : 190). Estimer et évaluer la qualité des données, c'est pour l'essentiel identifier la nature, la direction, l'ampleur et la signification vraisemblable de ces imperfections. L'estimation et l'évaluation de la qualité des données débute juste après que celles-ci ont été élaborées, mais c'est une activité récurrente – à chaque étape de l'analyse, l'utilisateur des données démographiques devrait regarder les résultats obtenus d'un œil sceptique, à l'affût d'éventuelles indications d'erreur ou de biais introduits par les données dans les résultats.

Nous exposons ici l'essentiel des recherches qu'un analyste devrait réaliser comme une évidence avant de s'embarquer dans un processus d'analyse démographique. Les principes de base selon lesquels conduire une évaluation et estimation démographiques ont à peine changé dans les cinquante dernières années. C'est pourquoi certains aspects de ce qui est présenté dans cette section sont tirés du Manuel II des Nations Unies, *Évaluation de la qualité des statistiques de base pour les estimations de la population* (UN Population Branch 1955), mis à jour et modifié en fonction des nécessités, ainsi que d'un autre guide pour l'évaluation des données de recensement, plus récent, proposé par le Bureau du recensement américain, *Evaluating Censuses of Population and Housing* (U.S. Bureau of the Census 1985). Ce dernier ouvrage est un guide utile et complet pour notre sujet ; les chapitres 4 et 5 sont en particulier fortement recommandés à tous les analystes entreprenant un processus d'estimation et d'évaluation de la qualité des données démographiques.

La section qui suit explique pourquoi il est nécessaire d'évaluer la qualité des statistiques démographiques. Elle offre aussi un survol des principes et des pratiques utilisés.

NÉCESSITÉ D'APPRÉCIER LA QUALITÉ DES STATISTIQUES DÉMOGRAPHIQUES

Les statistiques de population, comme toutes les autres statistiques démographiques, qu'elles soient obtenues par dénombrement, par exploitation d'un registre ou par d'autres moyens, sont sujettes à erreur. Les erreurs peuvent être larges ou minimes, selon les obstacles à un enregistrement exact dans la zone étudiée, selon les méthodes utilisées pour compiler les données et selon l'efficacité relative avec laquelle ces méthodes sont appliquées. Leur ampleur étant donnée, les erreurs ont une importance plus ou moins grande selon les usages auxquels les données sont destinées. Certaines applications restent valides même si les statistiques sont sujettes à d'assez larges erreurs ; d'autres nécessitent des données plus précises. Quand on traite un problème spécifique, il est important de savoir si les données sont assez exactes pour fournir une réponse d'une précision acceptable.

Pour les estimations de population, l'évaluation de la qualité des statistiques tirées des recensements ou d'enregistrements administratifs sur lesquelles les estimations sont fondées est doublement importante. En premier lieu, une recherche sur la précision des données de base doit précéder tout essai en vue de déterminer la fiabilité des estimations. Les erreurs d'estimation proviennent à la fois d'imprécisions dans les statistiques de base sur la population et d'erreurs dans les hypothèses introduites dans le calcul des estimations (par exemple dans les hypothèses sur les variations de population entre la date des dernières statistiques et la date à laquelle l'estimation s'applique). Il faut prendre en compte les deux sources d'erreur si on veut connaître le degré de

confiance à mettre dans les estimations. En second lieu, lorsqu'une recherche sur la précision des données de base révèle des erreurs, dont la direction et l'ampleur peuvent être estimées, il est possible de faire des ajustements compensateurs explicites ou implicites lors de la préparation des estimations de population. Il est fréquent également que des mesures démographiques raisonnablement fiables (par exemple des taux de fécondité) puissent être obtenues, même lorsque les données sous-jacentes ne sont pas fiables sur certains aspects.

Notre propos ici est de décrire les principales méthodes pour apprécier la précision des données de recensement dans leurs aspects les plus couramment utilisés, comme base des estimations de la population actuelle et des projections de la population future. On suppose que les résultats ont été tirés d'au moins un recensement et que l'analyse doit déterminer la précision de celui-ci et des autres données démographiques, sans qu'il soit possible de ré-dénombrer l'ensemble de la population ni de recommencer une quelconque partie importante de l'entreprise de recensement.

Il n'est pas possible de voir en détail toute l'information qu'on peut utiliser, dans un pays donné, pour apprécier la qualité de ses données démographiques. Par exemple, des données d'enquête peuvent fournir des estimations des paramètres démographiques utiles pour l'évaluation de la qualité des données tirées d'un recensement. Les exemples présentés ici sont des illustrations des méthodes et les résultats ne doivent pas être pris comme des évaluations définitives de la qualité des données employées.

Les résultats des tests décrits dans le présent manuel sont de divers types. Parfois, un test révélera seulement que les statistiques ont toutes chances d'être « raisonnablement précises » ou « suspectes » ; si elles sont « suspectes », d'autres recherches seront nécessaires avant qu'un jugement définitif soit porté. D'autres tests ne se contenteront pas d'indiquer qu'il y a des erreurs, mais ils conduiront aussi à estimer la direction et l'ampleur probable de l'erreur. Dans ce dernier cas, il sera souhaitable d'ajuster ou de corriger les statistiques fautives et de réviser les estimations fondées sur elles. La description des procédures à utiliser pour réviser des estimations est toutefois hors du champ du présent manuel.

Une distinction est souvent faite dans les textes démographiques entre les erreurs de couverture (qui résultent de disparités dans le dénombrement entre régions, groupes ethniques, âges, etc. et qui conduisent les données à un défaut de représentativité de l'ensemble statistique qu'elles sont censées décrire) et les erreurs de contenu (résultant d'erreurs

du répondant ou de l'agent recenseur, ou de fausses déclarations), mais cette distinction n'est pas très utile pour déterminer les stratégies d'évaluation de la qualité des données. Il est fréquent que les défauts des données ne puissent être attribués exclusivement à l'un ou l'autre type d'erreur. Mais en cherchant à expliquer et à comprendre les erreurs détectées, il peut être utile de considérer où dans le processus de recensement l'erreur peut s'être introduite. Procéder ainsi aide à déterminer les actions appropriées pour corriger les données, si c'est possible. Répétons que la description de ces remèdes est hors du champ du présent manuel.

DOCUMENTATION DE BASE À RECHERCHER

La conduite d'un recensement est difficile et compliquée – on dit par exemple que le recensement décennal aux États-Unis est l'entreprise du gouvernement fédéral la plus grande et la plus complexe en temps de paix (National Research Council 2004). Il en va sans doute de même dans tout autre pays réalisant un recensement. Pour aider dans cette tâche, des normes et procédures recommandées ont été rédigées par la Division statistique des Nations Unies. Nombre des manuels pertinents sont disponibles en ligne : les deux les plus intéressants pour les démographes qui analysent et évaluent la qualité des données de recensement sont les *Principes et recommandations concernant les recensements de la population et des logements. Deuxième révision* (Division statistique de Nations Unies 2008) et le *Manuel de vérification des recensements de la population et de l'habitation, Révision 1* (Division statistique des Nations Unies 2010). Le premier propose des conseils pour la logistique de l'organisation d'un recensement, de la planification de l'ensemble de l'opération à la dissémination des résultats ; le second porte sur le traitement des données après les collectes de terrain en vue de la sortie des résultats.

La nature et la qualité des données démographiques disponibles varient largement selon les pays. Les recensements de population sont entrepris avec une fréquence et une précision variables, et les données tirées de l'état civil contiennent des niveaux de détail fortement divergents, et varient énormément en qualité entre pays et à l'intérieur de chacun d'eux. La migration internationale peut être relativement importante ou pas. En conséquence, l'appréciation de la précision des statistiques doit être faite par des méthodes différentes, en fonction de situations différentes ; c'est pourquoi on ne peut pas considérer tous les tests détaillés auxquels peuvent être soumises toutes les sortes concevables

de données sur les sujets traités ici. C'est pourquoi aussi les méthodes présentées ici ne peuvent pas toujours être directement applicables à un problème spécifique ; des modifications doivent être introduites pour répondre à des exigences particulières.

Quand c'est possible, l'analyste doit chercher à obtenir autant d'informations pertinentes que possible de la part de l'organisme responsable de la conduite du recensement ou de l'enquête, concernant les pratiques opératoires et les difficultés rencontrées, ainsi que les politiques et les pratiques en matière d'apurement et de correction des données avant leur sortie. Quand une étude post-censitaire a été réalisée, il est nécessaire d'en connaître les résultats.

En plus des sources de données qui peuvent ne pas être dans le domaine public, la qualité des idées qu'on peut se faire sur la nature des données dépendra de la capacité qu'aura l'analyste de mettre à profit autant de matériau que possible susceptible d'être pertinent, pas seulement dans le domaine démographique, mais aussi social, économique, historique et politique. À titre d'exemple (développé plus loin), la baisse spectaculaire des probabilités de survie adulte dans les années 1990, telle qu'indiquée par les données des recensements de 1992 et 2002 au Zimbabwe, peut tenir dans une large mesure aux effets du VIH/sida sur la mortalité adulte à cette époque.

TYPES DE TESTS UTILISÉS

Selon qu'on traite de données de recensement, de statistiques d'état civil ou d'enregistrement des migrations, les mêmes types de tests peuvent être appliqués. Cette similitude vient du fait que les phénomènes démographiques sont liés à la fois entre eux et avec les autres phénomènes sociaux et économiques. Certaines de ces relations sont directes et nécessaires. Par exemple, l'accroissement de population dans un intervalle de temps donné est précisément déterminé par les nombres de naissances, de décès et de mouvements migratoires nets survenant dans cet intervalle. D'autres relations sont moins précises et moins définies. Par exemple, dans certains pays, une dépression économique est susceptible d'entraîner une baisse du taux de natalité – et la prospérité une hausse – mais l'ampleur exacte de la variation du taux de natalité ne peut pas être connue, même à partir d'une connaissance détaillée de la situation économique.

Les types principaux de tests possibles peuvent être résumés ainsi :

- a) tests de cohérence, fondés sur un ou plusieurs recensements ;
- b) comparaison des données observées avec une configuration théoriquement attendue, par exemple l'utilisation d'équations d'équilibre et les modèles de projection de population ;
- c) comparaison des données observées dans un pays avec celles observées ailleurs ;
- d) comparaison avec des données semblables obtenues pour des usages non démographiques ; et
- e) vérifications directes (ré-dénombrement d'échantillons de la population, etc.).

Le premier type de procédure de contrôle examine la cohérence des données, soit de façon interne (par exemple, la distribution de la population par âge et/ou sexe est-elle conforme aux attentes), ou de façon externe par comparaison avec des données antérieures du même pays. La théorie de la transition démographique nous conduit à penser – typiquement – que les taux de natalité et de mortalité (et donc les taux de croissance de la population) vont baisser de manière cohérente, ordonnée, sans discontinuité majeure. (Une exception est la possibilité qu'au tout début de la transition les taux de natalité s'accroissent). En l'absence de facteurs exogènes clairement identifiables (comme la guerre, la famine ou les épidémies), des écarts à ces régularités suggèrent fortement l'existence de problèmes dans les données.

Les comparaisons du deuxième type ont fortement varié avec les années. Historiquement, les tests les plus courants consistaient à comparer les données avec celles qu'implique la population stable équivalente du pays en question. Avec le début du déclin de la fécondité dans presque tous les pays du monde, il est devenu de plus en plus rare que soient remplies les hypothèses nécessaires pour que des comparaisons de ce type soient pertinentes. Aujourd'hui, les comparaisons cherchent plus souvent à confronter les taux de mortalité des hommes et des femmes et les rapports de masculinité par âge avec ceux qu'on attendrait dans des contextes semblables à ceux de la source de données analysées. En outre, la comparaison avec les résultats de calculs modélisés (par exemple les Perspectives de la population mondiale des Nations Unies ou les projections du Bureau du recensement américain) peut être utilisée pour révéler d'éventuelles incohérences dans les données.

Des « équations d'équilibre » peuvent aussi être utilisées pour tester la cohérence de l'accroissement des populations

révélé par deux dénombrements à différentes dates, en utilisant l'accroissement révélé par les statistiques des diverses composantes de la variation de la population – naissances, décès et migrations – pendant le même intervalle. Si toutes les données étaient exactes, les deux mesures de la croissance (ou de la décroissance) seraient équilibrées. Le test ne vaut pas seulement pour la population totale, mais aussi pour les groupes d'âge et sexe et pour d'autres catégories de population identifiables dans les statistiques. En outre, en réarrangeant et en redéfinissant les composantes de cette équation, des appréciations séparées peuvent être portées sur la précision des statistiques de naissances, de décès et de migrations.

Le troisième type de tests repose sur la connaissance préalable d'un pays qu'on peut considérer comme démographiquement semblable au pays étudié. Ce peut être, par exemple, un pays voisin. Mais si cette démarche est adoptée, il faut prendre grand soin de s'assurer que les similitudes entre les deux pays sont suffisamment grandes (pas seulement démographiquement, mais aussi socialement, économiquement, culturellement, etc.) pour autoriser l'extrapolation des données d'un cadre démographique à un autre.

Les deuxième et troisième types de contrôle sont semblables. Les variations démographiques observées dans un autre pays où les conditions sont présumées être semblables peuvent parfois être substituées à une configuration théoriquement attendue. Dans les deux cas, les comparaisons révéleront des différences fortes ou faibles. L'essence du test repose alors sur la réponse à la question : l'écart entre les valeurs observées et attendues peut-il être expliqué par des événements historiques ou les conditions actuelles dans le pays dont les données sont testées ? Sinon, il faut en conclure que les données observées sont « suspectes ». Des recherches supplémentaires peuvent apporter une explication de la différence, ou elles peuvent donner des indications claires selon lesquelles les données « suspectes » sont effectivement erronées. Ce genre de méthode est souvent appliqué comme étape préliminaire, pour suggérer les directions dans lesquelles devraient s'engager des tests supplémentaires.

Le quatrième type de test s'appuie sur la disponibilité de statistiques administratives ou d'autres statistiques sociales pouvant éclairer la démographie du pays étudié. Des estimations de la taille de diverses composantes de la population nationale peuvent être tirées de listes électorales, de statistiques d'inscriptions scolaires ou de populations particulières telles que celles des sites du Système de surveillance démographique (SSD), etc. Si de telles estimations diffèrent

des données du recensement de la population, la question est de savoir si la différence peut être expliquée de façon satisfaisante. Étant donné la spécificité des données localement disponibles et la nature des comparaisons qui peuvent en être tirées, les tests de ce type ne font pas l'objet d'une discussion ultérieure. Mais il faut prendre soin de ne pas supposer que ces sources alternatives sont inévitablement meilleures que le recensement testé.

Enfin, les vérifications directes impliquent des recherches de terrain, comme une enquête post censitaire. L'avantage d'une vérification directe tient au fait que les individus dénombrés ou les événements individuels enregistrés peuvent être identifiés, de sorte que la cohérence n'est pas seulement vérifiée sur les totaux, mais qu'elle peut aussi porter sur les erreurs spécifiques d'omission ou de double compte. Les contrôles directs sous la forme d'une enquête post censitaire permettent aussi de corriger la population dénombrée pour tenir compte du sous-dénombrement estimé.

Les quatre premiers types de procédure ne donnent qu'une indication de précision relative car les deux séries de données peuvent être sujettes à erreur. Si différentes procédures sont utilisées, ou s'il existe une forte présomption selon laquelle une série de données utilisée dans la comparaison est très précise, la démonstration donne une indication majeure sur les inexactitudes des données testées. Dans de nombreux autres exemples, la comparaison peut seulement révéler qu'au moins une, si ce n'est les deux, séries de données sont erronées.

Les recherches rapportées ci-dessous se concentrent sur le premier et le deuxième type de test. (Les vérifications directes sont présentées brièvement ailleurs, dans la section sur les contrôles de cohérence post censitaires). Chaque fois que ce sera possible, des exemples spécifiques seront inclus. Les données utilisées dans ces exemples seront tirées des données de recensement diffusées par IPUMS (Minnesota Population Center 2010). Mais seule une fraction des données et des connaissances disponibles pour chaque pays a été utilisée pour élaborer ces exemples. De nombreuses données plus pertinentes, dont certaines ne sont publiées nulle part, existent dans ces pays.

Une observation finale avant de décrire les divers tests présentés ici : la plupart des tests (mais pas tous) peuvent être appliqués à des subdivisions géographiques plus réduites, à la réserve près que la migration joue un rôle croissant dans la détermination de la taille et de la forme des populations à des niveaux plus fins de désagrégation. Ici aussi nous nous attendons à trouver des schémas « ordonnés » de variations de popu-

lation, à la fois au sein de la même subdivision à des périodes intercensitaires successives, et parmi différentes subdivisions lors d'une période donnée. Toute dissemblance devrait pouvoir être expliquée par des différences de condition connues. En pratique, on sait qu'il peut y avoir une diversité considérable dans les taux de variation de la population dans les différentes régions de toute nation. En conséquence, le problème revient à essayer de distinguer entre les variations qu'on peut expliquer dans des termes autres que des erreurs dans les statistiques, et celles où on ne peut pas. Il faut noter que, bien que ces procédures puissent révéler la présence d'erreurs et dans certains cas indiquer leur ordre de grandeur, elles n'offrent pas une base pour des estimations exactes de la taille des erreurs.

CONTRÔLES PRÉLIMINAIRES

Avant d'essayer d'évaluer la qualité des données, l'analyste doit :

- Examiner les procédures de dénombrement et l'information sur la qualité de sa réalisation. Ceci inclut de vérifier si une enquête post-censitaire a eu lieu, et si les données doivent être pondérées et, le cas échéant, comment. Quand c'est possible, l'accès doit être cherché aux données non encore corrigées, ou corrigées légèrement, ainsi qu'aux manuels et aux algorithmes utilisés pour corriger les données.
- Vérifier comment les données ont été collationnées dans une forme lisible par une machine. La saisie manuelle a l'inconvénient d'être lente ; la lecture optique – technique utilisée dans de nombreux recensements autour de 2000 et dans les vagues suivantes – est plus rapide, mais elle est sujette à de nombreux autres défauts (par exemple des difficultés à distinguer les 1 des 7 dans de nombreuses écritures), ainsi qu'à des problèmes de lecture des dernières pages des formulaires des recensements, exposées au risque de salissures.
- Comparer les résultats de recensement avec toute autre donnée disponible dans des sources non démographiques se rapportant aux effectifs de la population ou à des parties de celle-ci.
- Comparer la distribution de population révélée par les résultats du recensement avec les caractéristiques connues des subdivisions ; par exemple, la densité de population dans les zones rurales devrait être moindre que dans les zones urbaines.
- Comparer le nombre d'individus et le nombre de ménages (ainsi que le nombre moyen de personnes par ménage et

le nombre de ménages d'une seule personne) au niveau national et régional, et par subdivision rurale/urbaine, pour voir si ces résultats sont raisonnables.

Dans le compte du nombre total d'habitants d'un pays, le degré de précision est directement lié à la précision avec laquelle l'ensemble de l'opération de recensement est conduit. Le dénombrement peut être plus ou moins précis que le décompte de certaines composantes de la population, par exemple par groupe d'âge ou d'état matrimonial, mais si toutes les procédures de recensement sont défaillantes et que les caractéristiques de la population n'ont pas été déterminées précisément, il y a peu de chances pour que l'effectif total soit correct. En fait, une des façons d'apprécier la qualité du dénombrement consiste à analyser la précision des données relatives aux différentes caractéristiques de la population. Cette analyse risque de révéler non seulement une classification incorrecte des individus dénombrés mais aussi une tendance à omettre certaines catégories de population. Des efforts particuliers devraient être consacrés à l'appréciation de la complétude des décomptes de recensement dans les zones et les groupes de populations connus pour être difficiles à recenser. Il y a eu par exemple une longue tradition d'omission des très jeunes enfants en Afrique subsaharienne.

Une présentation détaillée des facteurs qui contribuent à la complétude d'un décompte de recensement dépasse les objectifs de ce manuel. On trouvera une discussion approfondie de ces facteurs dans les textes démographiques classiques (par exemple, Shryock and Siegel (1976) ; UN Population Branch (1955)).

Données manquantes et corrigées

Il est improbable que tous les recensés aient répondu aux questions à la fois sur l'âge et le sexe. S'il n'y a pas de données manquantes sur ces variables, c'est presque certainement que les données ont été corrigées. Toute correction n'est pas mauvaise. Cependant, comme la détermination de la fiabilité globale d'un ensemble de données repose, pour une part cruciale, sur la cohérence interne de la structure par âge et sexe de la population, il vaut mieux pouvoir déterminer quelles variables ont été apurées ou corrigées, et pouvoir évaluer la pertinence des règles mises en place à cet effet. Ceci est parfois indiqué sous la forme d'une étiquette attachée aux données, qui peut aussi préciser le type de correction ou d'imputation réalisée pour cette variable. Si c'est le cas, les distributions des données corrigées selon la méthode utilisée pour obtenir les données finales peuvent mettre en lumière les défauts ou les

anomalies dans les règles de correction. Quand c'est possible, il est souhaitable d'avoir accès aux données avant correction et apurement (ou très légèrement corrigées/apurées). Malheureusement, peu de pays diffusent des données « étiquetées », encore moins des versions de données avant correction.

Il est également important de connaître la proportion des données d'une variable qui ont fait l'objet de correction ou d'imputation. Si une proportion trop élevée de données ont été déterminées par une procédure de correction ou d'imputation, la distribution qui en résulte reflètera davantage les règles de correction des données que la réalité.

Quand des données sur l'âge manquent pour certaines personnes, ces enregistrements doivent faire l'objet d'une décision. Il n'est pas recommandé de se contenter de les exclure de l'analyse : en procédant ainsi, on réduit la taille absolue de la population et on suppose que la distribution par âge de ceux dont l'âge manque est la même que celle des personnes qui ont déclaré leur âge. Si on pense que c'est le cas, les âges manquants dans les tableaux doivent être répartis proportionnellement à la distribution par âge de la population dont les âges sont connus. Ainsi (en distinguant éventuellement les données par sexe), si nous désignons par N_x la population dénombrée âgée de x et par N_m la population dénombrée dont l'âge manque, nous distribuons ces cas proportionnellement par âge :

$$N_x^* = N_x + N_m \frac{N_x}{\sum_{x=0}^{\omega} N_x} = N_x \frac{\sum_{x=0}^{\omega} N_x + N_m}{\sum_{x=0}^{\omega} N_x}.$$

Mais s'il y a de fortes raisons de penser que les âges manquants sont regroupés dans une fraction spécifique de la population, la répartition doit être modifiée pour en tenir compte. Par exemple, il peut être souvent raisonnable de penser que les répondants connaissent les âges de leurs enfants en dessous d'un certain âge, disons 20 ans.

Quand il s'agit de répartir des données selon deux dimensions (par exemple l'âge et la région), il convient de suivre la démarche exposée par Arriaga (US Census Bureau 1997). La méthode consiste à mettre à l'échelle de façon itérative les colonnes puis les lignes pour que leurs sommes égalent les totaux marginaux désirés. La convergence est généralement obtenue après quelques itérations. La feuille de calcul disponible sur internet à <http://demographicestimation.iussp.org/fr> suit cette démarche et permet de traiter jusqu'à 20 lignes et 30 colonnes.

CONTRÔLES FONDÉS SUR LA DISPONIBILITÉ DES DONNÉES D'UN SEUL RECENSEMENT

Les contrôles fondés sur un seul recensement doivent être réalisés dans tous les cas, même s'il s'avère qu'il existe aussi des données tirées de recensements ou d'enquêtes antérieurs. Ces contrôles donnent une première idée sur les données collectées dans le recensement et reposent largement sur une évaluation de la cohérence et de la régularité des données par âge et sexe.

Distributions par âge et sexe

Étant donné le rôle central de l'âge et du sexe dans la détermination des trois composantes des variations démographiques, les recherches sur les distributions de la population par âge et sexe sont fondamentales dans tout processus d'estimation et d'évaluation de la qualité des données. Des recherches de ce type peuvent fournir une information essentielle sur

- la structure par âge et sexe de la population,
- les différences de couverture ou d'omission,
- la précision des âges déclarés ainsi que d'éventuelles attractions par certains âges, et
- si les données ont été corrigées ou non.

Pyramides des âges et autres évaluations graphiques

Nous ne recommandons pas la construction de pyramides des âges comme outil d'évaluation de la qualité des données démographiques, même si elles peuvent être utiles pour un certain nombre d'autres applications et si les pyramides animées peuvent être un outil éducatif précieux pour montrer comment les populations se modifient au fil du temps (voir les exemples du Canada ou de l'Allemagne). Historiquement, les pyramides des âges ont été utilisées pour se faire une idée de la structure globale de la population dénombrée par le recensement. Bien que la construction de pyramides rudimentaires soit relativement simple dans Excel, leur bon formatage est laborieux. Plus important : le jugement visuel sur les données est difficile quand les données par âge et sexe sont présentées sous cette forme. Une représentation graphique simple de la population dénombrée par âge et sexe sur le même système d'axes de coordonnées peut donner bien plus facilement la même information (et davantage). La première évaluation des données doit être faite par année d'âge, après quoi on peut passer à l'examen des distributions par groupes quinquennaux.

Identification des âges attractifs

La représentation graphique de la population par année d'âge permet de repérer d'emblée les âges qui attirent des déclarations plus fréquentes que les âges adjacents. Le constat visuel fournit probablement une indication aussi bonne que des mesures comme l'indice de Myers, l'indice de Whipple ou l'indice de régularité sexe-âge des Nations Unies. Ces indices peuvent être utiles pour des fins comparatives mais leurs échelles sont au mieux indicatives et l'information additionnelle qu'ils fournissent par rapport à une simple observation graphique ne justifie pas leur emploi. Le manuel du Bureau du recensement américain conclut de la même façon : « Ces procédures sont utiles comme mesures synthétiques ou pour des fins comparatives, mais elles ne donnent généralement pas d'idée sur les schémas d'erreur dans les données qui ne puisse être tirée d'analyses des données par des méthodes graphiques ou par le calcul de ratios. » (U.S. Bureau of the Census 1985 : 140)

L'attraction de certains âges prend généralement – mais pas toujours – la forme de concentrations de la distribution par âge de la population sur des âges terminés par 0 ou 5. Selon la façon dont la variable âge est recueillie ou calculée au recensement, la concentration peut se faire à d'autres âges. Par exemple, si l'âge au recensement est calculé à partir de la déclaration par le répondant de son mois et de son année de naissance, l'attraction peut être exercée par des années de naissance se terminant en 0 ou 5 (1920, 1925, etc.) et la concentration qui en résultera sur les années d'âge révolu dépendra de la date du recensement. En outre, d'autres formes d'attraction peuvent être difficilement visibles – par exemple celle occasionnée par un enregistrement massif en un point du temps ou par des événements d'importance historique majeure, conduisant à des préférences pour des âges se terminant en 0 ou 5 à cette date.

En l'absence d'événements extérieurs significatifs, on peut s'attendre à une évolution démographique régulière, et donc à une progression harmonieuse des nombres de personnes dénombrées au fil des âges. Dans les pays en développement où la fécondité est restée élevée, on s'attend à ce que la taille de la population décroisse de façon monotone à mesure que l'âge augmente. Si le nombre absolu des naissances a baissé récemment, on s'attend à avoir moins d'enfants aux jeunes âges qu'à des âges un peu plus élevés.

Une limitation à la mise en graphique de la population par âge et sexe est que les distorsions et les erreurs dans les données aux âges avancés risquent d'être cachées par des tailles de population (beaucoup) plus importantes aux jeunes âges.

On peut recourir à des ratios ou des taux relatifs peuvent être utilisés pour étudier d'éventuelles distorsions et erreurs aux âges avancés. En l'absence de comparateur, les âges avancés devraient être étudiés séparément.

Rapports d'âge

Alors que l'attraction par certains âges peut être détectée plus facilement par une représentation graphique que par le calcul d'indices, le calcul de rapports d'âge peut fournir des indications utiles concernant d'éventuels sous-dénombrements ou des déplacements entre groupes d'âge. Le rapport d'âge pour un groupe d'âge donné est le ratio du double de la population dans ce groupe d'âge à la somme de la population dans chacun des groupes d'âge adjacents. Algébriquement,

$${}_nAR_x = \frac{2 \cdot {}_nN_x}{({}_nN_{x-5} + {}_nN_{x+5})} \cdot 100.$$

Si on suppose que la variation de population est à peu près linéaire entre les groupes d'âge, le rapport doit être proche de 100. Des écarts à 100, en l'absence de facteurs extérieurs plausibles (par exemple la migration ou des calamités passées ayant affecté des groupes d'âge particuliers), indiquent des erreurs dans les données dues au sous-dénombrement ou à des déplacements.

Une aberration dans les effectifs de population d'un groupe d'âge particulier (qu'elle soit réelle ou qu'elle résulte d'erreurs dans les données) risque de perturber les rapports d'âge dans les groupes d'âge de chaque côté. Si un groupe d'âge est particulièrement peu nombreux, le rapport d'âge sera inférieur à l'unité pour ce groupe d'âge, avec des pointes dans les groupes adjacents.

Rapports de masculinité

Une deuxième catégorie de contrôles résulte du calcul des rapports de masculinité dans la population, à la fois globalement et à chaque âge. Le rapport de masculinité global (*sex ratio* SR) est le rapport du nombre d'hommes pour 100 femmes dans la population. Ce rapport peut ensuite être décomposé par âge comme suit :

$${}_nSR_x = \frac{{}_nN_x^m}{{}_nN_x^f} \cdot 100,$$

où ${}_nN_x^i$ représente la population de sexe i ($i = m$ ou f) dénombrée entre les âges x et $x + n$.

Comme la mortalité des femmes est généralement inférieure à celle des hommes dans la plupart des populations,

le rapport de masculinité doit refléter cet écart des mortalités. Dans les pays développés, le rapport de masculinité à la naissance (nombre de naissances de garçons pour 100 naissances de filles) est typiquement de l'ordre de 105, alors qu'il est plus proche de 100 en Afrique subsaharienne (Garenne 2004). Des valeurs du rapport de masculinité à la naissance, tirées par exemple du sexe de la dernière naissance déclarée au recensement ou à l'état civil, qui s'écartent de cet intervalle indiquent l'existence d'avortements en fonction du sexe, d'infanticides ou de problèmes de déclaration.

En l'absence de migration nette importante, le rapport global reflète la mortalité relative des femmes et des hommes. S'il n'y a pas de raisons spécifiques pour que la mortalité des femmes dépasse celle des hommes (par exemple sélection fœtale liée au sexe, infanticide des petites filles, très forte mortalité maternelle, négligence généralisée à l'égard des femmes telle que discutée par Sen (1992)), on doit s'attendre à ce que le rapport de masculinité global soit légèrement inférieur à 100. Étant donné les différences entre la mortalité des hommes et des femmes, en particulier aux âges avancés, le niveau exact du

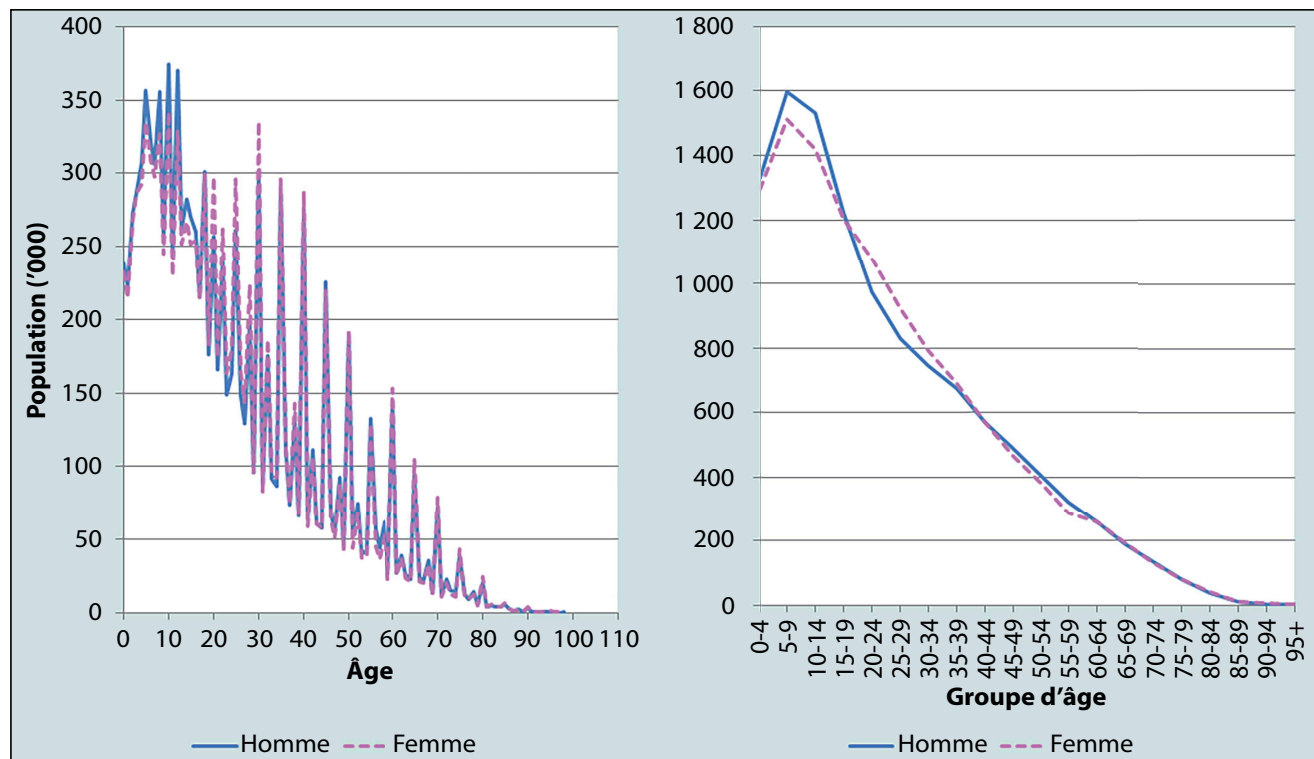
rapport global dépend fortement de la structure par âge de la population : il est plus faible pour les populations âgées et plus élevé pour les populations jeunes.

Entre la naissance et les âges adultes avancés (autour de 45 ans dans les pays en développement, 60 ans ou plus dans les pays développés) le rapport de masculinité recule lentement, sauf en cas de migration nette significative. Par la suite, le rapport de masculinité tend à chuter rapidement à mesure que la mortalité masculine dépasse largement la mortalité féminine. Un écart à ce schéma est couramment visible dans les pays où existe une forte migration de travail des jeunes adultes différenciée selon le sexe. Si de nombreux jeunes hommes vivent à l'étranger au moment du recensement, il en résultera une forte baisse des rapports de masculinité, suivie par une reprise progressive parmi les hommes plus âgés à mesure que les travailleurs migrants rentrent au pays.

Commentaires pour conclure

Une évaluation complète de la qualité des données recueillies dans un recensement et une enquête doit chercher à

FIGURE 1.1 Structure par sexe et âge, par année d'âge et groupe d'âge, Népal recensement 2001



expliquer – avec aussi peu d’hypothèses que possible – les caractéristiques observées des données. L’analyste doit être attentif à des problèmes bien connus, souvent rencontrés dans les données de recensement sur l’âge et le sexe – le sous-dénombrement des jeunes hommes d’âge actif et la tendance à se vieillir, fréquemment observée dans les pays où existent des formes de protection sociale telles que les pensions de vieillesse. Enfin s’il y eu une immigration significative, il peut être utile d’analyser la population née sur place séparément du reste de la population ; il n’y a pas d’exploitation comparable pour l’émigration, à moins que des données existent par âge, sexe et pays de naissance dans les pays de destination les plus importants.

EXEMPLE

La feuille de calcul disponibles sur <http://demographicestimation.iussp.org/fr> présente les données d’un échantillon de 11,35 % du recensement de 2001 au Népal, tel que diffusé par IPUMS (Minnesota Population Center 2010). Les données ont été vraisemblablement corrigées ou apurées,

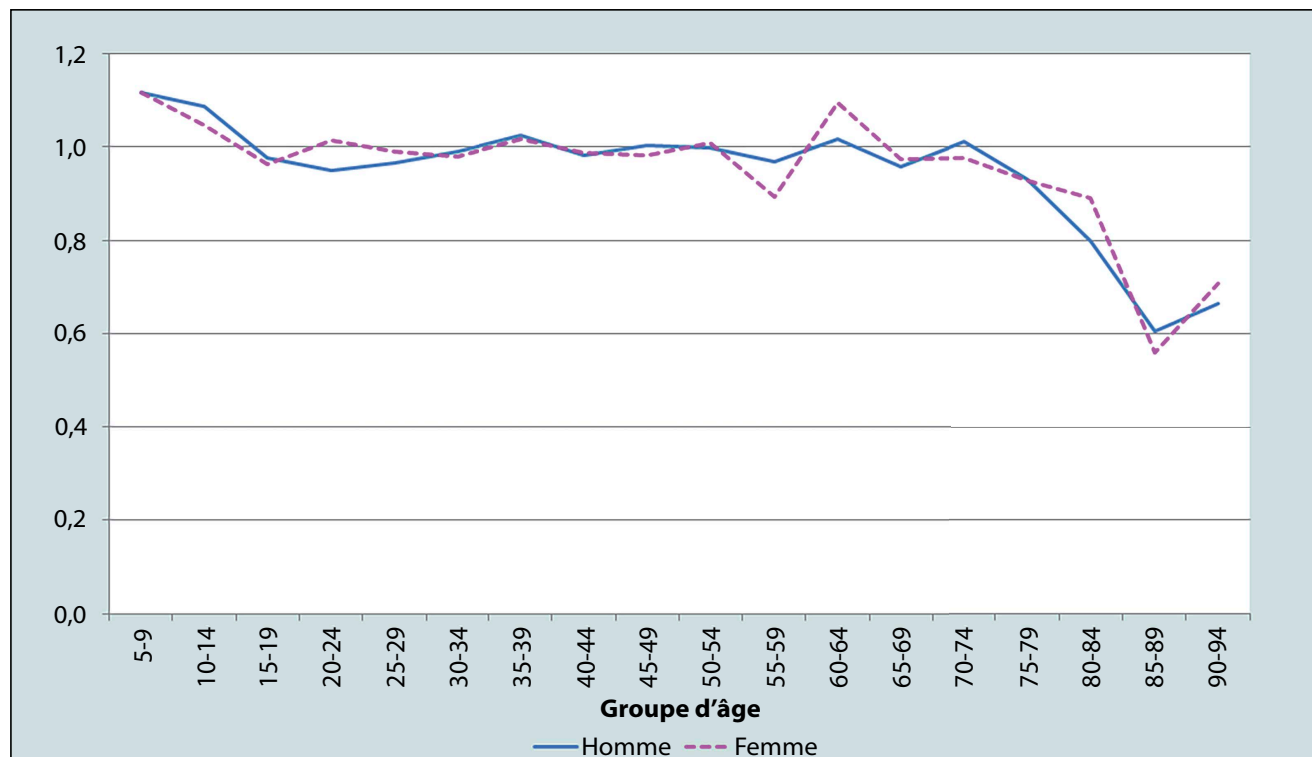
puisqu’il n’y a pas de cas d’âge ou de sexe manquants dans les données. L’analyste doit chercher à déterminer la nature et l’ampleur de ces corrections.

Comme suggéré plus haut, nous commençons par représenter graphiquement la population dénombrée par année d’âge et sexe (Figure 1.1).

Le graphique de gauche fait apparaître une forte attraction par les âges se terminant par 0 et 5 dans ces données. Par exemple, la population des hommes et femmes dénombrés à 30 ans est plus du triple de celle âgée de 29 ou 31 ans. L’attraction est également visible pour les âges se terminant par 2 et 8. Le phénomène est moins marqué pour la population âgée de moins de 30 ans, même si c’est en partie dû à l’attraction exercée par d’autres âges (8, 12 et 18). La déclaration des âges est manifestement mauvaise dans ces données.

Un fort déficit est également perceptible pour la population dénombrée avant 5 ans, la population dénombrée à 1 an étant à peu près égale aux deux tiers de celle âgée de 5 ans. Il est peu vraisemblable que la fécondité ait baissé

FIGURE 1.2 Rapports d’âge par sexe et groupes quinquennaux d’âge, Népal recensement 2001



autant dans une période aussi brève, aussi la première hypothèse est-elle que des jeunes enfants ont été sous-dénombés dans ce recensement. Une mauvaise déclaration des âges des enfants – entraînant une surestimation du nombre d'enfants âgés de 5-9 ans – peut aussi avoir contribué au déficit du nombre de très jeunes enfants.

La représentation de la même distribution par âge, mais avec des groupes quinquennaux pour lisser les données (graphique de droite sur la Figure 1.1), ajoute de nouvelles informations. Le fort déficit de la population âgée de moins de 5 ans apparaît à nouveau, mais la comparaison visuelle de la population âgée de 5-9 ans avec celles de 10-14 ans suggère l'éventualité d'un sous dénombrement des enfants âgés de 5-9 ans aussi. Ceci remet en question l'éventualité d'un large transfert des enfants de 0-4 ans dans le groupe des 5-9 ans. Enfin, les rapports d'âge par groupe quinquennal sont représentés Figure 1.2.

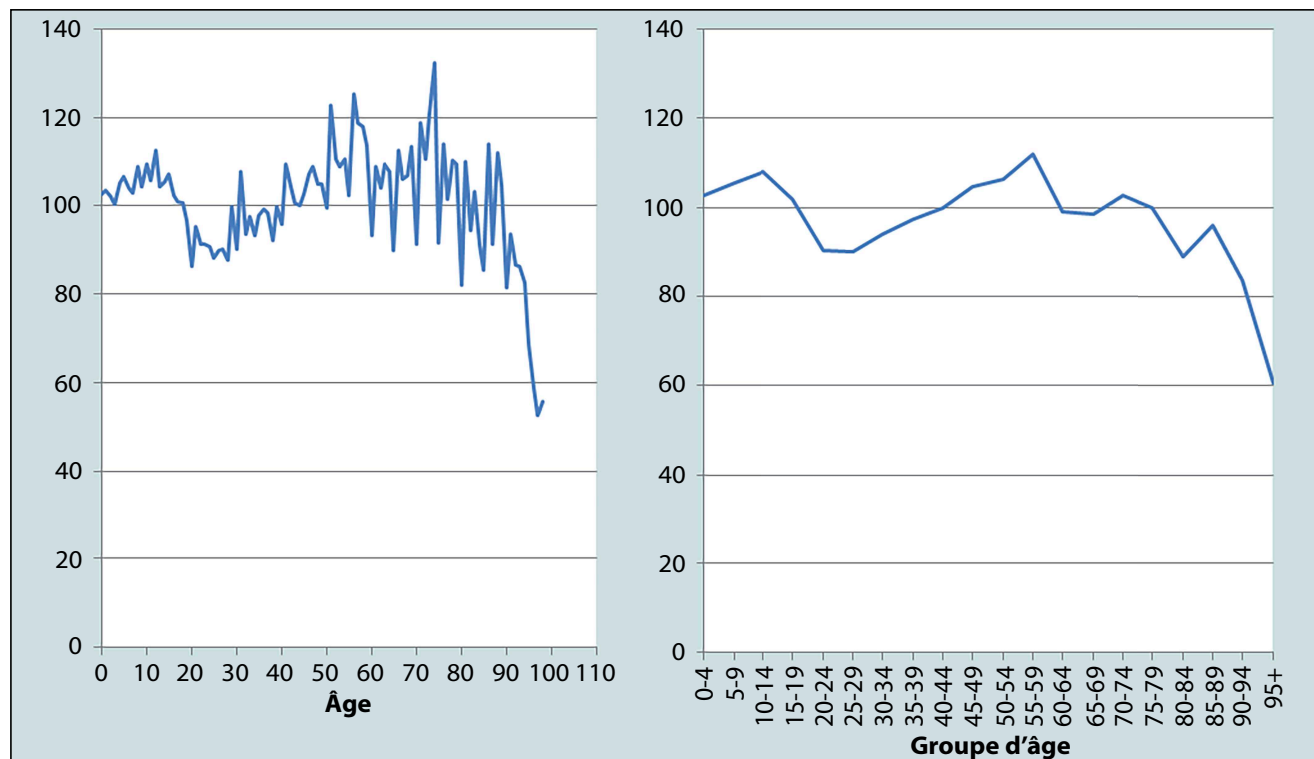
Les rapports d'âge sont généralement proches de l'unité pour les deux sexes, sauf aux plus jeunes âges (indiquant des omissions d'enfants âgés de 0-4 ans et, dans une moindre

mesure, un déplacement de ces enfants dans le groupe 5-9 ans). La chute des rapports d'âge aux âges les plus élevés est conforme aux attentes, étant donné l'accroissement rapide de la mortalité à ces âges.

En l'absence d'autres informations, les distributions par âge et sexe ne peuvent pas être analysées plus avant, mais l'analyste peut envisager de comparer la population aux âges pertinents avec des données administratives indiquant les nombres d'enfants inscrits à l'école, ou comparer avec le nombre de naissances déclarées à l'état civil, 5-9 ans et 10-14 ans avant le recensement. Une comparaison peut aussi être faite avec la population estimée âgée de 5-9 ans calculée en appliquant des estimations des taux de fécondité du milieu des années 1990 à la population féminine estimée à la même période.

Une recherche peut aussi porter sur les populations relatives des hommes et des femmes par groupe d'âge. Dans l'ensemble, le rapport de masculinité de la population dénombrée est de 100,5 hommes pour 100 femmes. Il y a un net excès d'hommes dénombrés jusqu'à 20 ans. Entre 20

FIGURE 1.3 *Rapports de masculinité par année d'âge et groupe d'âge, Népal recensement 2001*



et 40 ans, il y a plus de femmes que d'hommes. Ceci pourrait être la conséquence d'une émigration de travail (masculine) ou d'un sous-dénombrement différentiel des jeunes hommes adultes. L'analyste doit chercher à expliquer ce phénomène. Il est plausible que l'émigration de travail (masculine) rende compte d'une partie du déficit en hommes ; l'excédent d'hommes dénombrés entre 40 et 60 ans coïncide avec les âges auxquels les hommes peuvent rentrer au pays, mais ceci ne saurait expliquer des rapports de masculinité supérieurs à l'unité. Il se pourrait que les phénomènes sociologiques (avortement lié au sexe du fœtus, infanticide des petites filles) décrits par Sen (1992) pour l'Inde valent aussi pour le Népal.

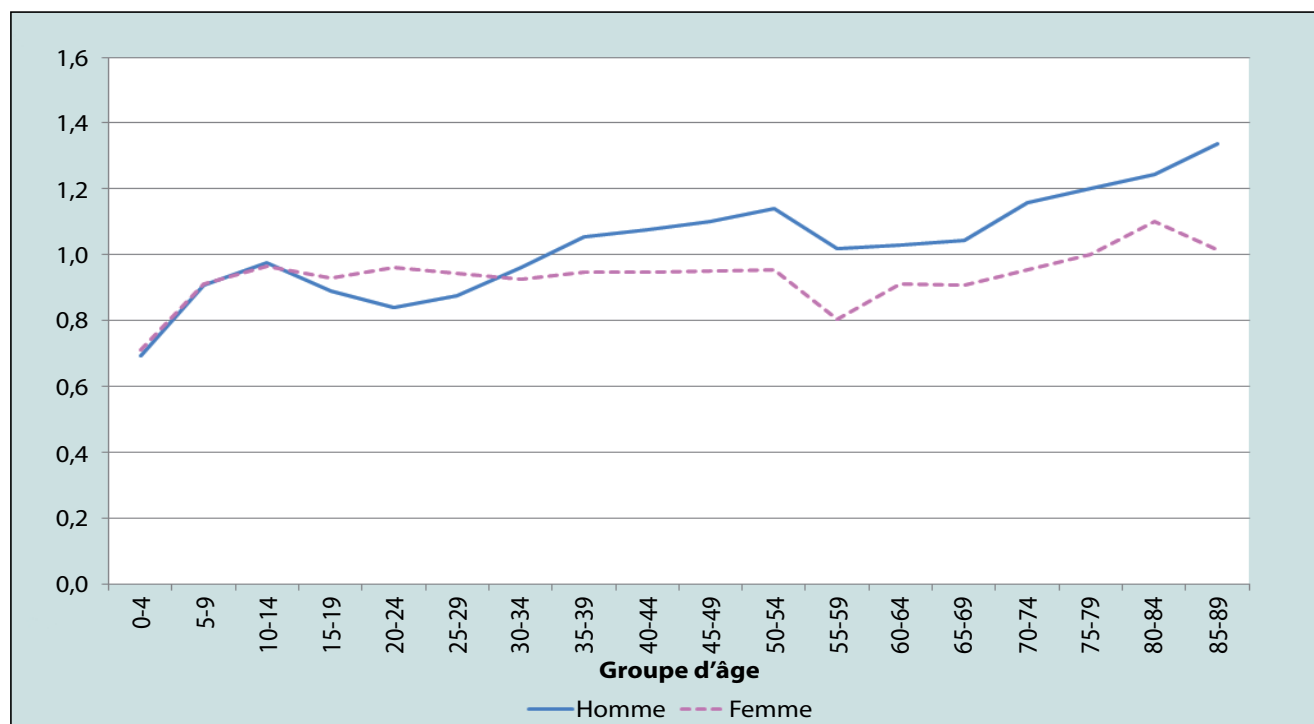
Sur la Figure 1.3, deux caractéristiques des rapports de masculinité par année d'âge (graphique de gauche) apparaissent. Premièrement, ils sont très erratiques, chutant brutalement à partir de 60 ans aux âges se terminant par 0 et 5. Ceci suggère que l'attraction par les âges ronds joue moins sur les hommes que sur les femmes. Deuxièmement, en plus

du déficit en hommes entre 20 et 40 ans déjà décrit, il apparaît aussi un déficit en femmes dans le recensement au-delà de 40 ans, à en juger par le fait que les rapports de masculinité restent supérieurs à (ou proches de) l'unité jusqu'aux âges les plus avancés. Il faudrait une fois de plus s'interroger sur la possibilité d'appliquer l'hypothèse de Sen au Népal.

Les données par groupes quinquennaux (graphique de droite sur la Figure 1.3) sont plus lisses mais elles confirment néanmoins les analyses ci-dessus.

On approfondit ses idées sur la nature et la qualité des données par âge et sexe du recensement de 2001 au Népal en comparant celles-ci aux estimations les plus récentes de la Division de la population des Nations Unies pour le pays en 2001 (UN Population Division 2011). Ces estimations diffèrent largement des données du recensement. La façon la plus efficace d'en rendre compte est de rapporter la population dénombrée (par sexe et âge) à la population estimée par la Division de la population des Nations Unies pour 2001 (Figure 1.4).

FIGURE 1.4 Rapport par âge et sexe de la population dénombrée en 2001 à la population estimée par la Division de la population des Nations Unies pour 2001 [Perspectives de la population mondiale (2011)], Népal



Les rapports ne sont pas représentés au-delà de 90 ans car ils sont encore plus extrêmes à ces âges – atteignant 9,7 (pour les hommes) et 8,6 (pour les femmes) – et leur représentation masquerait les différences aux plus jeunes âges.

Les estimations des Nations Unies reflétant essentiellement les hypothèses qui les sous-tendent, les fortes disparités entre les deux séries d'estimations nécessitent une étude approfondie. Jusqu'à l'âge de 15 ans, les rapports pour les hommes et les femmes suivent à peu près les mêmes tendances. Mais la population des hommes et des femmes dénombrés à 0-4 ans est environ 30 % inférieure à celle estimée par les Nations Unies, alors que celles à 10-14 ans sont très proches à 2 ou 3 % près. Aux âges plus avancés, les schémas des hommes et des femmes divergent fortement, le nombre de femmes entre 15 et 55 ans différant presque constamment de 5 % entre les deux sources.

Par rapport aux projections des Nations Unies, les déclarations des âges apparaissent nettement surestimées aux âges avancés, en particulier chez les hommes.

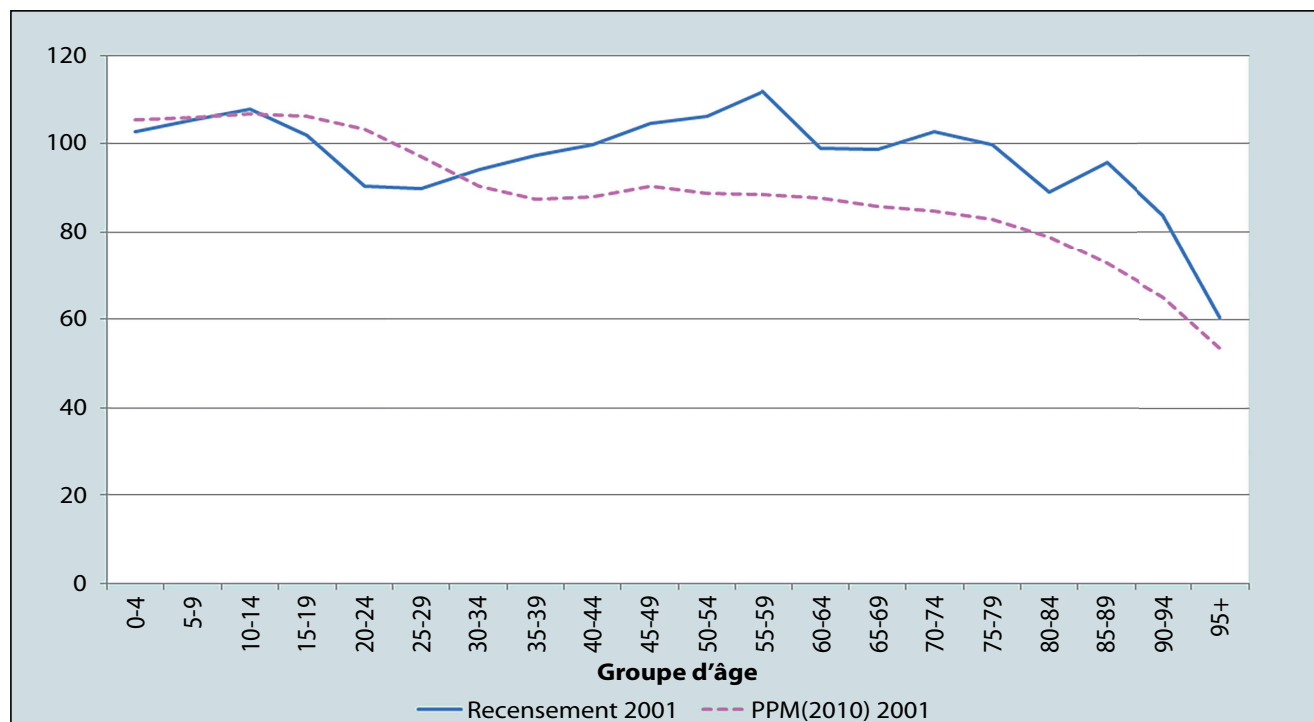
Il y a aussi des différences notables entre les rapports de masculinité par âge calculés d'après les données du recensement de 2001 et ceux estimés pour 2001 par la Division de la population des Nations Unies (Figure 1.5). Des travaux complémentaires seraient nécessaires pour comprendre ce qui peut expliquer les fortes divergences dans la structure démographique du pays.

CONTRÔLES FONDÉS SUR PLUSIEURS RECENSEMENTS

En plus des contrôles décrits précédemment, la disponibilité de séries supplémentaires de données tirées de recensements antérieurs (et de systèmes d'enregistrement à l'état civil) rend d'autres recherches possibles.

Il est souvent difficile de déterminer si des irrégularités révélées lors de l'évaluation de la structure par âge et sexe d'une population dans un seul recensement sont principalement dues à des erreurs dans les données ou à de réelles particularités de la structure de population. Quand les résultats

FIGURE 1.5 Rapports de masculinité par groupe d'âge dans la population dénombrée en 2001 et la population estimée par la Division de la population des Nations Unies pour 2001 [Perspectives de la population mondiale (2011)], Népal



de deux recensements ou plus sont disponibles, il devient possible de lever ces incertitudes, même sans recourir à des techniques plus élaborées que celles décrites à la section précédente. Par exemple, si nous avons les statistiques par âge du recensement de 2008 au Cambodge, la possibilité d'expliquer certaines irrégularités dans les données de 1998 comme conséquences de déficits dans les naissances ou de décès au cours de la période Khmer Rouge à la fin des années 1970 s'en trouverait largement éclairée. Si les résultats de 2008 montraient les mêmes particularités dans des groupes de dix ans plus âgés, et non dans les mêmes groupes d'âge qu'en 1998, nous serions en bonne position pour conclure que ces particularités reflètent une réalité plutôt que des erreurs de dénombrement. Quand les données de deux recensements ou plus sont disponibles à intervalle de quelques années, on obtient une information encore plus décisive concernant les erreurs en utilisant des équations d'équilibre ou des calculs analogues à partir des données de certaines cohortes – par exemple en comparant les nombres déclarés à 10-14 ans lors d'un recensement antérieur avec ceux déclarés à 20-24 ans au recensement dix ans plus tard. Quand des données sont disponibles pour une série de trois recensements ou plus, les déclarations successives peuvent être rattachées les unes aux autres de cette façon sur toute la série. Mais pour expliquer les techniques en jeu, le recours aux données de deux recensements suffit à les illustrer.

Une fois de plus, le principe de base dans la comparaison de deux recensements successifs ou plus est que les variations de population se manifestent normalement de manière régulière. Quand on s'éloigne de cette régularité, les écarts doivent pouvoir être expliqués par des événements connus, comme une réduction de l'immigration, une famine ou un autre événement. Les écarts qui ne peuvent pas être expliqués ainsi avertissent de la possibilité d'erreurs, et la présomption d'erreurs est renforcée si les résultats d'autres tests vont dans la même direction. Dans certains pays, il peut être possible d'appliquer ces tests séparément à différents groupes ethniques, si des tableaux par âge et sexe existent pour ces groupes et si des données sont disponibles sur l'immigration et l'émigration de ces groupes (ou si les groupes en question sont peu affectés par la migration internationale).

Les contrôles recourant spécifiquement à plusieurs recensements sont, pour la plupart, fondés sur des méthodes utilisées pour mesurer la mortalité adulte – en d'autres termes, l'évaluation de la cohérence des données est un sous-produit des méthodes d'estimation de la mortalité adulte ; dans

certains cas, ce sont même ces méthodes qui sont utilisées directement. Cette section décrit certaines de ces méthodes en s'appuyant sur les données le plus couramment disponibles.

Évaluation des taux de croissance intercensitaires

Le taux de croissance, r , est défini comme

$$r = \frac{\ln\left(\frac{N(t_2)}{N(t_1)}\right)}{(t_2 - t_1)},$$

où $N(t_1)$ et $N(t_2)$ désignent les populations totales aux temps t_1 et t_2 respectivement.

Si la variation de la population d'un pays tient seulement à l'accroissement naturel, il est peu vraisemblable que le taux moyen annuel de croissance dépasse 3,5 %. Un taux de ce niveau serait le résultat d'un taux de natalité élevé (disons 45 pour 1 000 ou plus) et d'un taux de mortalité très bas (disons 10 pour 1 000 ou moins). Par ailleurs, en l'absence de forte émigration, la population d'un pays en développement ne décroît que dans des circonstances exceptionnelles. De fait, presque tous les taux d'accroissement naturel observés dans les temps modernes ont été compris dans l'intervalle entre zéro et 3,5 %. Dans quelques pays développés, d'après les Perspectives de la population mondiale 2010, la croissance naturelle est négative. Si dans un pays, le taux de variation de la population approche ou dépasse ces extrêmes sans une immigration ou une émigration massive, la question se pose de savoir si un taux aussi inhabituel a une explication ou si les résultats des recensements sont erronés.

Avec une information, même approximative, sur les conditions de mortalité et de fécondité dans un pays, les limites du taux de croissance vraisemblables peuvent être définies plus précisément.

Si des effectifs de population sont connus grâce à trois recensements ou plus, il devient possible de procéder à une évaluation plus précise en comparant les taux de croissance successifs. Le même principe s'applique à nouveau : on s'attend à ce que la croissance de la population évolue régulièrement, sauf si on peut montrer que des circonstances nouvelles peuvent avoir impliqué un rythme de croissance différent.

Par ailleurs, si les résultats des recensements successifs sous-estiment tous la population dans la même proportion, l'estimation de r est correcte. La connaissance de r peut donc

donner une indication sur le sous dénombrement relatif des recensements successifs.

Rapports de survie des cohortes

Tout groupe d'âge peut être défini comme cohorte : par exemple, les garçons de moins de 5 ans, les femmes de 50-54 ans ou les personnes de 10-19 ans à la date du recensement. Si un deuxième recensement a lieu exactement dix ans plus tard, les survivants de chaque cohorte auront exactement dix ans de plus au moment du deuxième recensement. Leurs nombres seront réduits par les décès mais ils peuvent être augmentés ou diminués du solde des mouvements d'immigration et d'émigration. En général, la mortalité est le facteur principal ; si le solde migratoire est négligeable, la variation des effectifs peut être utilisée pour calculer un taux de survie analogue à celui des tables de mortalité. Calculé pour une seule cohorte, un tel taux de survie apporte souvent peu d'information sur la précision des statistiques. Un résultat ouvertement absurde révèle clairement une erreur. Par exemple, un accroissement des effectifs d'une cohorte d'un recensement à l'autre est clairement impossible, sauf en cas de forte immigration. Semblablement, même quand la mortalité est très élevée, il est invraisemblable qu'une cohorte âgée initialement de n'importe quel âge entre 5 et 60 ans, soit réduite de moitié dans la décennie qui suit.

On peut porter un jugement plus précis si on compare les taux de survie pour des cohortes de chaque sexe à différents âges. Les taux de survie sont fonction des taux de mortalité par âge et, comme eux, ils suivent généralement plus ou moins le même schéma de variation d'âge en âge, que la mortalité soit basse ou élevée. Le taux de survie s'accroît après les toutes premières années d'enfance et il atteint généralement son maximum vers 10 ans ; il recule ensuite, d'abord très progressivement, puis de plus en plus rapidement à mesure que des âges avancés sont atteints. À la plupart des âges, voire à tous, les femmes ont généralement un taux de survie plus élevé que les hommes du même âge. Si la survie hypothétique calculée pour différentes cohortes s'éloigne significativement de ce schéma, et si aucune explication (telle que la migration) ne peut le justifier, on doit suspecter les statistiques d'imprécision.

Dans quelles conditions de telles comparaisons de cohortes à des recensements successifs peuvent-elles les plus significatives ? Une condition est soit l'absence d'une migration nette importante soit une connaissance détaillée de la composition des migrants par âge et sexe. Une deuxième

condition, analogue à la première, est le maintien de frontières immuables. Si les frontières du pays ont changé entre deux recensements, de sorte que des personnes en nombre important ont été ajoutées ou retranchées à la population, la composition par âge et sexe de ces personnes doit être connue pour que l'analyse de cohorte donne une indication précise sur la qualité des statistiques. Une troisième condition est que la population couverte par les deux recensements soit la même. Si, par exemple, l'ensemble de la population est dénombré à un recensement mais que les militaires sont exclus au second, les cohortes d'âge incluant des militaires ne peuvent être comparées sans un ajustement adéquat, à moins que le nombre de militaires soit négligeable. Si les nationaux vivant à l'étranger sont inclus dans un recensement et exclus d'un autre, et si les nombres en jeu sont importants, surtout s'ils sont concentrés sur certains groupes d'âge ou de sexe, ce type d'analyse n'est pas pertinent.

Dans le cas d'un pays où l'immigration est importante, une cohorte peut néanmoins être comparée à deux recensements, même si les données migratoires manquent. Si la population née sur place est réputée n'avoir pas émigré de façon importante, les comparaisons de deux recensements peuvent être limitées à cette population.

Les rapports de survie peuvent être calculés sur n'importe quelle tranche d'âge et n'importe quel intervalle de temps, pourvu qu'on ait des données par année d'âge pour au moins un des deux recensements. Avec le programme décennal de recensements recommandé par les Nations Unies, une tranche de dix ans d'âge est courante.

Méthode

Les rapports de survie des cohortes (CSR) mesurent la proportion des personnes dénombrées aux âges x à $x + n$ au temps t , ${}_nN_x(t)$ au premier recensement, qui sont encore survivantes et dénombrées dans un deuxième recensement a années plus tard lorsqu'elles sont âgées de $x + a$ à $x + n + a$ ans au temps $t + a$, ${}_nN_{x+a}(t + a)$. Ainsi

$${}_nCSR_x(a) = \frac{{}_nN_{x+a}(t + a)}{{}_nN_x(t)}.$$

Lors de la représentation graphique, ces estimations peuvent être placées au milieu de la période intercensitaire (c'est-à-dire en $t + \frac{a}{2}$), et au point médian des âges à cette date, $x + \frac{a + n}{2}$. Une représentation graphique de ces rapports

de survie des cohortes permet de se faire facilement une idée rapide de la qualité des données, même si les mises en garde classiques s'appliquent ; une suite étrange de rapports de survie des cohortes indique un défaut des données, sans indiquer si l'erreur provient du premier recensement, du second ou des deux.

Cependant, quand on dispose des données d'un troisième recensement, il se peut que les taux de survie des cohortes tirés des deux premiers recensements apparaissent vraisemblables, alors que ceux tirés du deuxième et du troisième semblent problématiques. Dans ce cas, on est conduit à supposer que le défaut est dans le troisième recensement, pas dans le deuxième.

Enfin, si on dispose d'une table de mortalité appropriée, on peut en tirer un autre ratio en divisant le rapport de survie des cohortes par le taux de survie équivalent dans la table de mortalité. On obtient ainsi un ratio de ratios à chaque âge

$${}_nR_x = \frac{{}_nN_{x+a}(t+a) / {}_nN_x(t)}{{}_nL_{x+a} / {}_nL_x}.$$

Le ratio est égal à l'unité si trois conditions (fortes) sont respectées : qu'il n'y ait pas d'erreur dans le recensement, que la structure par âge de la population dénombrée soit identique à celle décrite par la table de mortalité et que la mortalité soit bien celle indiquée par la table de mortalité. Les écarts à l'unité indiquent soit des erreurs dans les données, soit le choix d'une table de mortalité inappropriée. En outre, dans ces conditions et en l'absence de migration, des ratios inférieurs à l'unité impliquent un sous-dénombrement relatif du second recensement par rapport au premier et vice-versa.

Exemple

Les recensements du Zimbabwe ont été réalisés à dix ans d'écart en 1982, 1992 et 2002 (la date de référence étant fixée au 18 août). Des tableaux de la population dénombrée par âge et sexe sont disponibles dans les *Annuaire démographiques* disponibles sur le site internet de la Division statistique des Nations Unies (nous avons utilisé le Supplément rétrospectif de 1997 et l'Annuaire 2008). Les données figurent au Tableau 1.1. Les populations âgées de moins d'un an et de 1-4 ans sont présentées séparément ; elles sont maintenues ainsi pour permettre une meilleure compréhension des changements rapides de la mortalité dans les cinq premières années de vie.

Comme nous n'avons pas de raison de penser que les personnes ayant omis de déclarer leur âge sont particulièrement concentrées dans certains groupes d'âge, la première étape consiste à répartir le nombre de ces cas (relativement peu nombreux) proportionnellement aux effectifs de population dans chaque groupe d'âge de 0 à 75 ans+. La proportion de personnes d'âge inconnu était de 0,19 % au recensement de 1982 ; elle a doublé au recensement de 2002, atteignant 0,38 %. Les distributions ajustées figurent au Tableau 1.2.

En respectant les principes soulignés précédemment, nous étudions d'abord les caractéristiques fondamentales de la population par âge et sexe. En l'absence de données par année d'âge, nous ne pouvons pas examiner cet aspect de la qualité des données. Les distributions de la population du Zimbabwe par âge et sexe sont représentées à la Figure 1.6.

Dans les trois recensements il y a un net excédent de femmes entre 15 et (au moins) 35 ans. C'est presque certainement le résultat de la migration de travail des jeunes hommes vers les pays voisins, en particulier l'Afrique du sud. Il semble y avoir eu un sous-dénombrement notable de la population âgée de moins de 5 ans au recensement de 1992 – la population dans ce groupe d'âge est moins nombreuse que celle de 5-9 ans à la même date, à la différence de ce qu'on observe dans les deux recensements adjacents.

Les rapports d'âge et les rapports de masculinité aux trois recensements sont représentés à la Figure 1.7.

Les rapports d'âge dans le groupe 60-64 ans sont particulièrement élevés dans les trois recensements, et l'excédent de population à ces âges contribue à des très faibles rapports d'âge dans les groupes 55-59 et 65-69 ans. Les rapports de masculinité sont proches de 100 au début et diminuent rapidement après 15 ans dans chaque recensement, probablement du fait de la migration des jeunes hommes. Il est plus préoccupant de constater la hausse des rapports de masculinité entre 35 et 55 ans vers des niveaux largement supérieurs à 100 au recensement de 1982. Ceci reflète presque certainement un sous-dénombrement des femmes. Les rapports de masculinité aux âges les plus avancés sont encore très élevés, ce qui reflète probablement une tendance des hommes âgés à se déclarer plus vieux qu'ils ne son en réalité.

L'allure très erratique des rapports d'âge et des rapports de masculinité n'inspire pas une grande confiance dans la qualité des données.

Ensuite, les taux de survie des cohortes sont calculés comme précédemment pour chaque sexe séparément,

TABLEAU 1.1 Population du Zimbabwe
par âge et sexe, recensements de 1982, 1992
et 2002

| Âge | 1982 | | 1992 | | 2002 | |
|--------------|------------------|------------------|------------------|------------------|------------------|------------------|
| | Homme | Femme | Homme | Femme | Homme | Femme |
| 0 | 133 070 | 136 960 | 167 552 | 169 064 | 170 054 | 170 277 |
| 1-4 | 510 260 | 528 390 | 621 411 | 626 664 | 668 008 | 667 730 |
| 5-9 | 612 760 | 619 300 | 821 319 | 832 469 | 764 453 | 769 247 |
| 10-14 | 529 750 | 518 740 | 724 905 | 731 846 | 754 587 | 757 657 |
| 15-19 | 390 160 | 412 610 | 615 728 | 632 510 | 736 686 | 766 890 |
| 20-24 | 290 380 | 364 200 | 466 837 | 523 060 | 564 034 | 658 873 |
| 25-29 | 243 420 | 281 060 | 335 713 | 376 495 | 473 984 | 513 793 |
| 30-34 | 185 400 | 206 760 | 280 066 | 326 299 | 369 836 | 360 291 |
| 35-39 | 147 920 | 170 170 | 229 360 | 259 555 | 235 692 | 268 797 |
| 40-44 | 142 050 | 139 530 | 174 266 | 189 509 | 194 702 | 239 727 |
| 45-49 | 116 490 | 110 390 | 145 437 | 143 441 | 165 437 | 191 168 |
| 50-54 | 111 780 | 90 880 | 133 261 | 147 339 | 128 029 | 173 229 |
| 55-59 | 67 400 | 60 800 | 94 713 | 86 729 | 98 417 | 112 498 |
| 60-64 | 76 850 | 65 260 | 95 510 | 84 213 | 94 447 | 99 420 |
| 65-69 | 38 810 | 38 860 | 51 202 | 50 902 | 64 301 | 67 851 |
| 70-74 | 29 810 | 30 500 | 58 279 | 62 479 | 60 311 | 62 464 |
| 75+ | 39 410 | 46 760 | 52 026 | 68 403 | 71 950 | 92 311 |
| Inconnu | 7 900 | 6 680 | 15 952 | 18 034 | 19 252 | 25 254 |
| TOTAL | 3 673 620 | 3 827 850 | 5 083 537 | 5 329 011 | 5 634 180 | 5 997 477 |

TABLEAU 1.2 Population ajustée
du Zimbabwe par âge et sexe, recensements
de 1982, 1992 et 2002

| Âge | 1982 | | 1992 | | 2002 | |
|--------------|------------------|------------------|------------------|------------------|------------------|------------------|
| | Homme | Femme | Homme | Femme | Homme | Femme |
| 0 | 133 357 | 137 199 | 168 079 | 169 638 | 170 637 | 170 997 |
| 1-4 | 511 360 | 529 314 | 623 367 | 628 792 | 670 298 | 670 554 |
| 5-9 | 614 081 | 620 383 | 823 904 | 835 296 | 767 074 | 772 500 |
| 10-14 | 530 892 | 519 647 | 727 187 | 734 331 | 757 174 | 760 861 |
| 15-19 | 391 001 | 413 331 | 617 666 | 634 658 | 739 212 | 770 133 |
| 20-24 | 291 006 | 364 837 | 468 307 | 524 836 | 565 968 | 661 659 |
| 25-29 | 243 945 | 281 551 | 336 770 | 377 773 | 475 609 | 515 966 |
| 30-34 | 185 800 | 207 121 | 280 948 | 327 407 | 371 104 | 361 815 |
| 35-39 | 148 239 | 170 467 | 230 082 | 260 436 | 236 500 | 269 934 |
| 40-44 | 142 356 | 139 774 | 174 815 | 190 152 | 195 370 | 240 741 |
| 45-49 | 116 741 | 110 583 | 145 895 | 143 928 | 166 004 | 191 976 |
| 50-54 | 112 021 | 91 039 | 133 680 | 147 839 | 128 468 | 173 962 |
| 55-59 | 67 545 | 60 906 | 95 011 | 87 023 | 98 754 | 112 974 |
| 60-64 | 77 016 | 65 374 | 95 811 | 84 499 | 94 771 | 99 840 |
| 65-69 | 38 894 | 38 928 | 51 363 | 51 075 | 64 521 | 68 138 |
| 70-74 | 29 874 | 30 553 | 58 462 | 62 691 | 60 518 | 62 728 |
| 75+ | 39 495 | 46 842 | 52 190 | 68 635 | 72 197 | 92 701 |
| TOTAL | 3 673 620 | 3 827 850 | 5 083 537 | 5 329 011 | 5 634 180 | 5 997 477 |

puisque les schémas et les niveaux de mortalité diffèrent entre hommes et femmes. Comme la population âgée de 0-4 ans en 1992, par exemple, aurait 10-14 ans en 2002, nous supposons que le taux de survie pour cette cohorte s'applique (à peu près) à des personnes âgées de 7½ ans au point médian entre les recensements, en août 1997. Les taux de survie des cohortes ne sont pas estimés pour les très jeunes âges, ni pour l'intervalle ouvert. Les résultats sont représentés graphiquement sur la Figure 1.8.

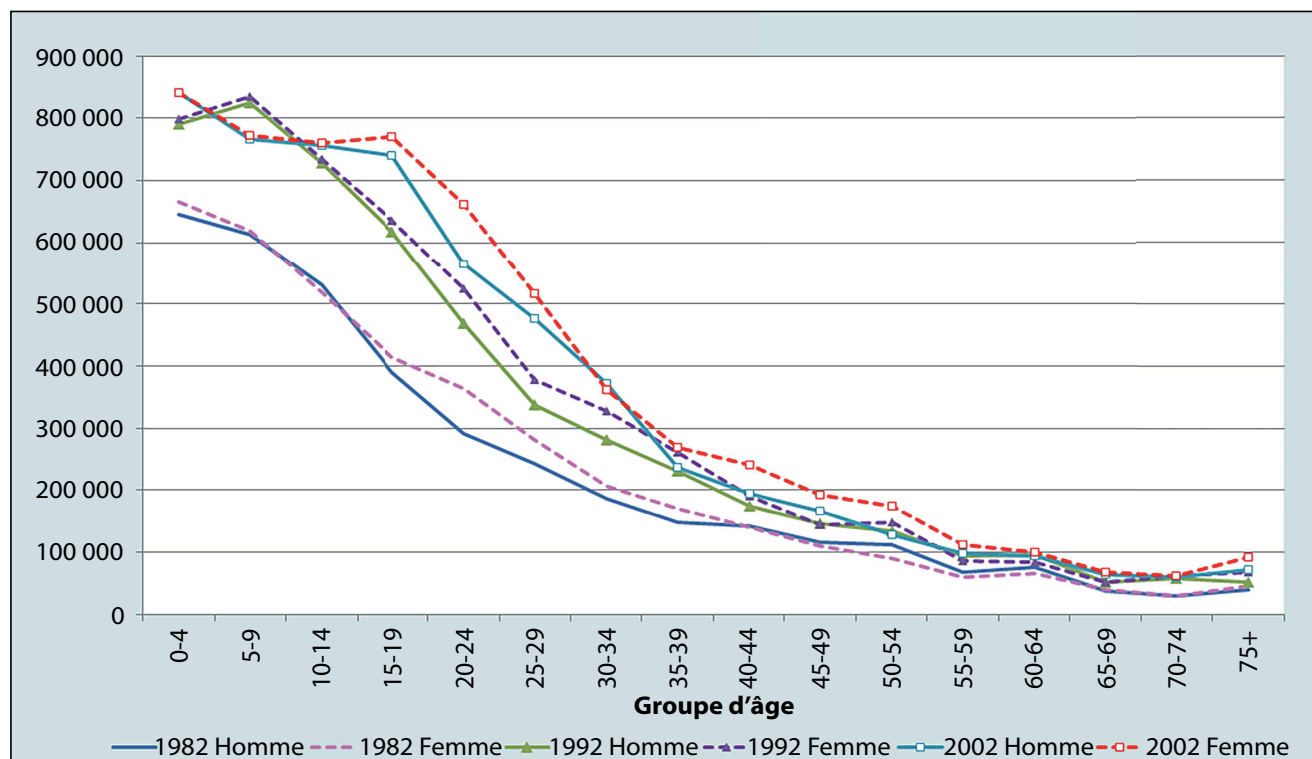
Le graphique en haut à gauche représente les taux de survie des cohortes entre les recensements de 1982 et 1992 par sexe ; celui en bas à gauche représente les données équivalentes des recensements de 1992 et 2002. Le sous-dénombrement est évident pour les enfants des deux sexes ainsi que pour les femmes jusque vers 20 ans au recensement de 1982 (sauf à imaginer une improbable forte immigration d'enfants entre 1982 et 1992), comme l'indiquent des rapports de survie supérieurs à l'unité.

Les deux graphiques de gauche font apparaître (en gros) une décroissance des rapports de survie (une augmenta-

tion de la mortalité) au fil des âges, mais les données sont loin d'être cohérentes soit par sexe soit par âge. Il est peu vraisemblable, par exemple, que les rapports de survie des hommes soient plus élevés que ceux des femmes au même âge. Il y a aussi un curieux pic dans les deux périodes intercensitaires des rapports de survie pour les femmes âgées de 40-44 ans au premier recensement et 50-54 ans au suivant. Ceci mériterait approfondissement.

Les deux graphiques de droite représentent les rapports de survie des cohortes au fil du temps, séparément pour les hommes et les femmes. Ils font apparaître une importante augmentation de la mortalité entre les deux périodes de dix ans. La nature erratique des rapports de survie indique que la qualité relative des données est pauvre, mais l'accroissement de la mortalité est presque certainement dû en large part à l'effet du vih/sida parmi les adultes dans la seconde période, conjointement à la dégradation rapide des conditions socio-économiques qui ont prévalu vers le tournant du siècle, ce qui a presque certainement alimenté une forte émigration de jeunes adultes. L'accroissement apparent de la

FIGURE 1.6 Distribution de la population du Zimbabwe par âge et sexe, recensement de 1982, 1992 et 2002



mortalité chez les enfants et les jeunes adultes, visible sur les deux graphiques de droite, est presque certainement dû au sous-dénombrement de cette population en 1982.

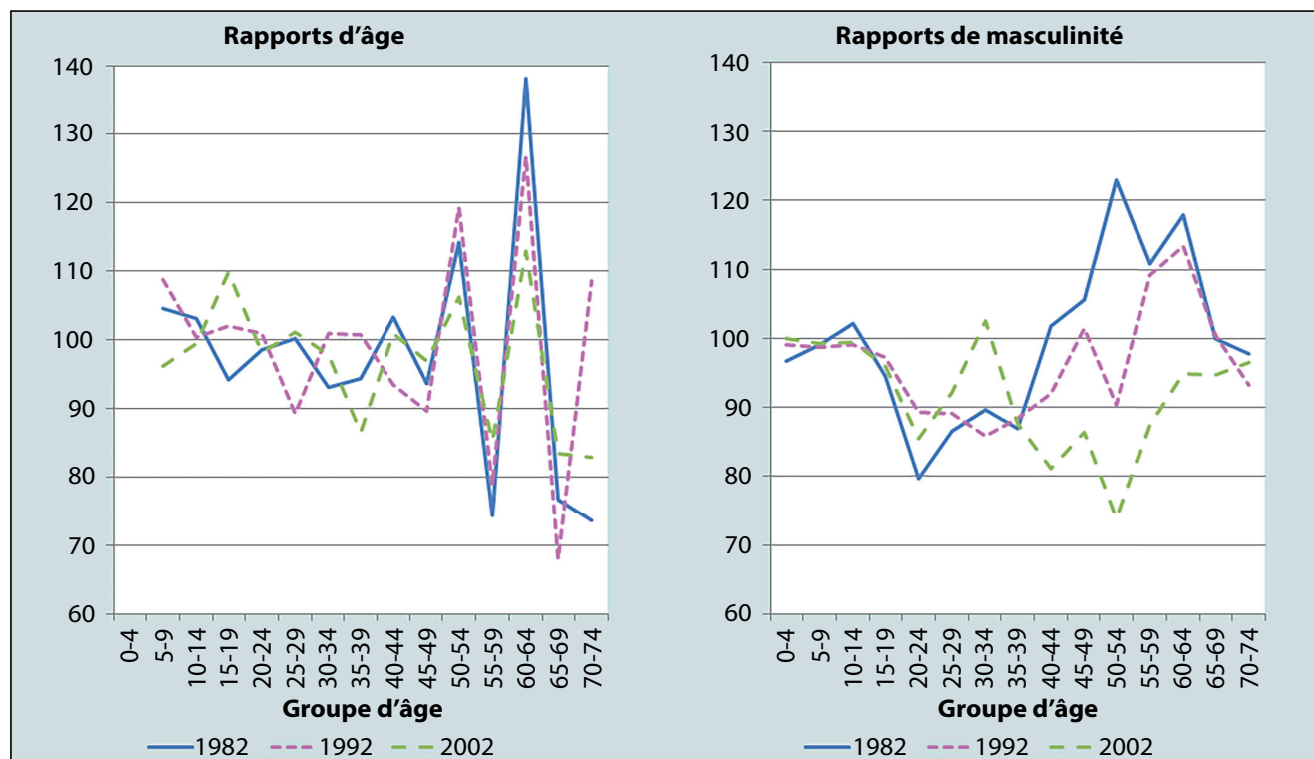
ENQUÊTES POST-CENSITAIRES

Les enquêtes post-censitaires (EPC) utilisent la logique des techniques de capture-recapture pour estimer la proportion de la population qui n'a pas été dénombrée au moment du recensement. La procédure consiste à retourner dans des zones-échantillons pour ré-administrer un second court questionnaire à tous les ménages qui auraient dû être dénombrés dans cette zone, après quoi les ménages et les individus « capturés » lors de l'enquête sont appariés, chaque fois que possible, avec les ménages et les individus dénombrés lors du recensement. Il doit en résulter une estimation concrète de l'ampleur du sous-dénombrement, qui peut être comparée et opposée à celle tirée, par exemple, d'une projection analytique de la population. Les résultats de l'EPC peuvent ensuite servir à ajuster (« pondérer ») les données du recensement pour compenser les effets du sous-dénombrement.

Une EPC peut donc être très utile. Il y a toutefois deux hypothèses essentielles sous-tendant le recours aux techniques de capture-recapture : un, les probabilités qu'une personne soit atteinte par le recensement et par l'enquête sont indépendantes l'une de l'autre, et deux, il est possible d'identifier le même individu sans ambiguïtés dans les deux sources. La première hypothèse est impossible à tenir dans les populations humaines – certains groupes qui évitent d'être décomptés par un recensement (les immigrés illégaux, par exemple) vont vraisemblablement éviter aussi d'être enquêtés. Dans ce sens, l'EPC ne donne d'information que sur ceux dont on sait qu'ils ont été omis par le recensement, pas sur ceux dont on ne sait pas qu'ils ont été omis. La seconde hypothèse est, elle aussi, difficile à tenir, en particulier dans les zones où la population est très mobile ou si l'intervalle est long entre le recensement et l'EPC.

Les principes et les bonnes pratiques associées à la conduite d'une EPC sont présentés dans un manuel de 2010 (UN Statistics Division 2010).

FIGURE 1.7 Rapports d'âge et rapports de masculinité, Zimbabwe 1982



Quand une EPC a eu lieu, il est bon que l'analyste puisse avoir accès au rapport de l'enquête pour en comprendre les éventuelles faiblesses. La capacité d'une EPC à fournir une information fine sur la qualité des données du recensement est directement liée à la taille de l'EPC et elle est inversement proportionnelle au délai écoulé entre le recensement et l'EPC. Contraint par la durée et le coût de la collecte, l'échantillon d'une EPC est inévitablement beaucoup plus petit qu'un recensement complet. En conséquence, les estimations du sous-dénombrement ne sont possibles qu'à un niveau grossier. Par exemple, au recensement de 2001 en Afrique du sud, les estimations ont été faites par grand groupe d'âge, sexe, groupe de population, province et type géographique de secteur de recensement (urbain, rural, formel, informel). La population est donc supposée sous dénombrée dans la même proportion dans chaque groupe défini par ces cinq caractéristiques. À des niveaux plus fins de décomposition, les estimations du sous-dénombrement risquent de ne pas être fiables.

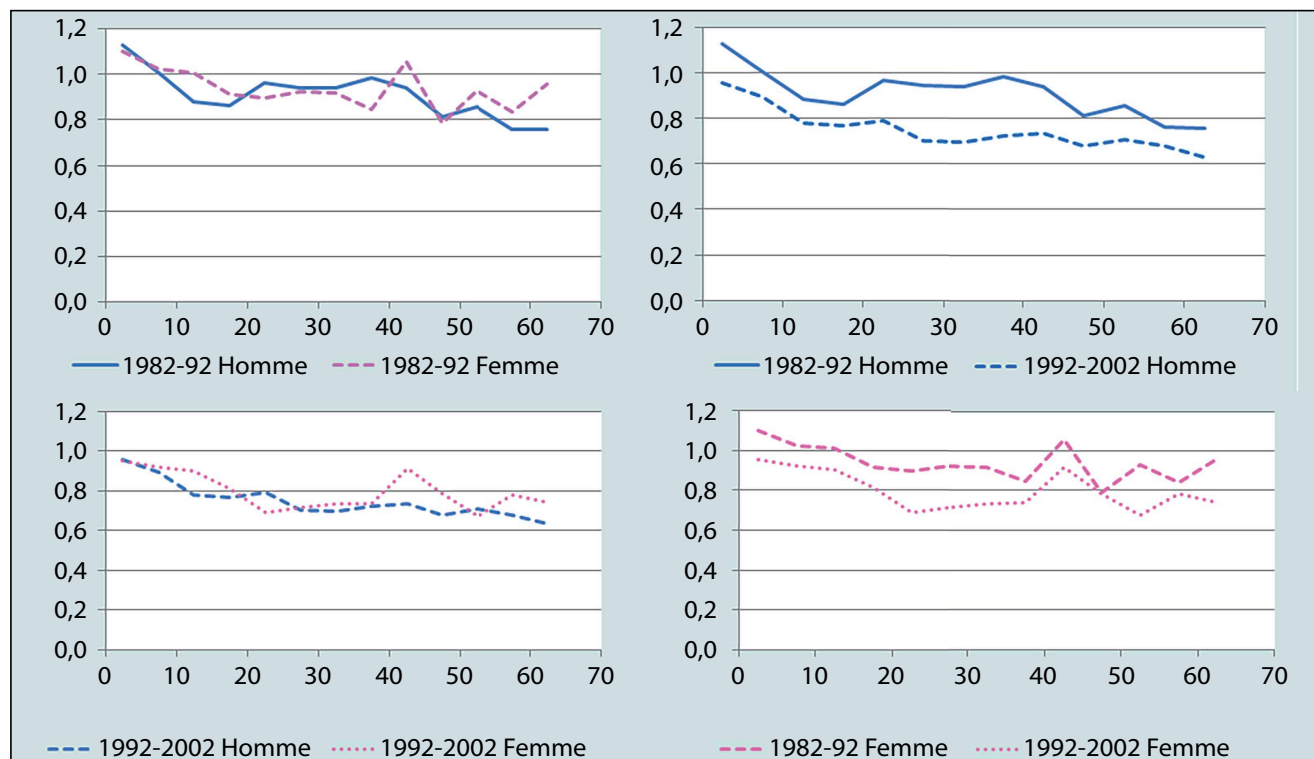
On peut avoir une idée de l'ampleur des ajustements réalisés et de l'importance du sous-enregistrement à partir d'une évaluation des pondérations fournies avec les données. Si les données brutes tirées d'un recensement ne sont pas ajustées à la suite d'une EPC, les pondérations reflètent le taux de sondage : s'il s'agit d'un échantillon de 10 % des données du recensement, chaque enregistrement sera supposé représentatif de 10 personnes et aura donc un poids de 10. Quand une EPC est prise en compte, le supplément de poids qui s'ajoute au taux de sondage révèle le sous-dénombrement. Analytiquement,

$$weight = \frac{sample\ fraction^{-1}}{1 - undercount}$$

$$undercount = 1 - \frac{sample\ fraction^{-1}}{weight}.$$

Si un enregistrement a un poids de 11,8 dans un échantillon de 10 % (c'est-à-dire un taux de sondage de 0,1, c'est qu'un

FIGURE 1.8 Rapports de survie des cohortes par âge et sexe, recensements du Zimbabwe 1992 et 2002



ajustement a été effectué pour tenir compte d'un sous-enregistrement de 15,3 % $(1-(1/0,1)/11,8)$.

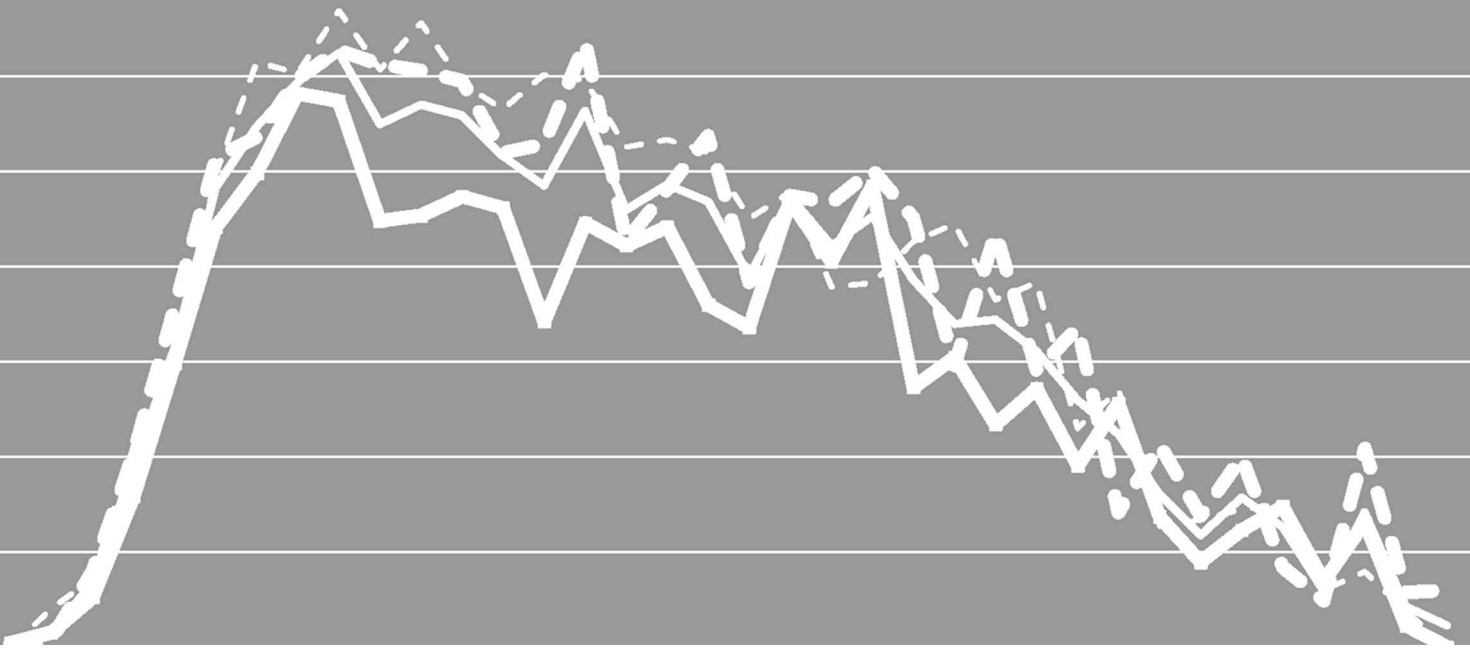
Quand les données ne sont pas accompagnées d'estimations du sous-dénombrement, l'application de cette dernière

formule aux pondérations qui s'attachent aux différents groupes de population permet à l'analyste de retrouver des estimations du sous-dénombrement avec un bon degré de précision.

RÉFÉRENCES

- Division statistique des Nations Unies 2008. *Principes et recommandations concernant les recensements de la population et des logements. Deuxième révision*. New York : Département des Affaires Économiques et Sociales, ST/ESA/STAT/SER.M/67/Rev.2. http://unstats.un.org/unsd/publication/seriesM/seriesm_67Rev2f.pdf
- Division statistique des Nations Unies 2010. *Manuel de vérification des recensements de la population et de l'habitation. Révision*. New York : Département des Affaires Économiques et Sociales, ST/ESA/STAT/SER.F/82/Rev.1. http://unstats.un.org/unsd/publication/SeriesF/SeriesF_82Rev1f.pdf
- Feeney G. 2003. "Data assessment," in Demeny, P and G McNicoll (eds). *Encyclopaedia of Population*. Vol. 1. New York: Macmillan Reference USA, pp. 190-193.
- Garenne M. 2004. "Sex ratios at birth in populations of Eastern and Southern Africa", *Southern African Journal of Demography* **9**(1):91-96. http://www.commerce.uct.ac.za/Organisations/Demography/SA_Journal_of_Demography/SAJD/Volume%209/SAJDem_2004_9_1_Garenne_b.pdf
- Minnesota Population Center. 2010. Integrated Public Use Microdata Series, International. Version 6.0 [Machine-readable database]. Minneapolis:University of Minnesota.
- National Research Council. 2004. *The 2000 Census: Counting under Adversity*. Panel to Review the 2000 Census. Citro, Constance F., Daniel L. Cork and Janet L. Norwood (eds), Committee on National Statistics, Division of Behavioural and Social Sciences and Education. Washington DC: National Academies Press.
- Sen A. 1992. "Missing women", *British Medical Journal* **304** (6827):587-588. doi: <http://dx.doi.org/10.1136/bmj.304.6827.587>
- Shryock HS and JS Siegel. 1976. *The Methods and Materials of Demography (Condensed Edition)*. San Diego: Academic Press.
- UN Population Branch. 1955. *Manual II: Methods of Appraisal of Quality of Basic Data for Population Estimates*. New York: United Nations, Department of Economic and Social Affairs, ST/EOA/Series A/23. <http://www.un.org/esa/population/techcoop/DemEst/manual2/manual2.html>
- UN Population Division. 2011. *World Population Prospects: The 2010 Revision, Volume I: Comprehensive Tables*. New York: United Nations, Department of Economic and Social Affairs, ST/ESA/SER.A/313. http://esa.un.org/unpd/wpp/Documentation/pdf/WPP2010_Volume-I_Comprehensive-Tables.pdf
- UN Statistics Division. 2010. *Post-Enumeration Surveys: Operational Guidelines - Technical Report*. New York: United Nations, Department of Economic and Social Affairs. http://unstats.un.org/unsd/demographic/standmeth/handbooks/Manual_PESen.pdf
- US Bureau of the Census. 1985. *Evaluating Censuses of Population and Housing*. Statistical Training Document ISP-TR-5. Washington, DC: US Bureau of the Census. <http://www.census.gov/srd/papers/pdf/rr85-24.pdf>
- US Census Bureau. 1997. *Population Analysis Spreadsheets for Excel*. Washington, DC: US Bureau of the Census. <http://www.census.gov/population/international/software/pas/>

Fécondité



Chapitre 2 INTRODUCTION À L'ANALYSE DE LA FÉCONDITÉ

Tom A Moultrie

INTRODUCTION

Dans la plupart des groupes humains, la fécondité est, à long terme, le déterminant le plus important de la dynamique et de la croissance des populations. Cette section donne une vue d'ensemble des types de méthodes disponibles pour la mesure de la fécondité. Les méthodes elles-mêmes seront décrites dans d'autres sections.

SOURCES DE DONNÉES

D'une manière générale, les données pour la mesure de la fécondité proviennent de trois sources.

La première est l'information recueillie en continu par un système d'enregistrement à l'état civil. Le plus souvent, l'enregistrement des naissances est incomplet dans les pays en développement. Les parents manquent fréquemment d'incitations à faire enregistrer leurs naissances ; les nouveau-nés qui décèdent peu après leur naissance risquent de ne pas être enregistrés, soit comme naissance, soit comme décès ; et l'enregistrement tardif des naissances (par exemple quand l'enfant atteint l'âge d'aller à l'école) peut impliquer un délai de plusieurs années avant que soient enregistrés tous les survivants de la cohorte née une année donnée.

La deuxième source est constituée des réponses aux questions sur la fécondité recueillies lors d'un recensement. Ces questions sont généralement posées à toutes les femmes d'âge fécond (12 ou 15 ans et plus, souvent avec une limite supérieure à 49 ans). Compte tenu de la complexité de l'exercice censitaire, et de la nécessité de chercher à dénombrer chaque individu, il n'est pas possible de poser des questions détaillées sur la fécondité. En pratique, ces questions visent une information synthétique sur la descendance (le nombre total d'enfants déjà nés, et le nombre d'enfants toujours vivants) et sur la fécondité au cours d'une période étroitement définie précédant le

recensement. Cette dernière information doit permettre aux démographes d'estimer les taux de fécondité actuels. La nature concise des questions posées limite les possibilités d'une validation interne et d'un recoupement des réponses.

En outre, comme on le verra plus en détail dans la section sur l'évaluation de la qualité des données de fécondité, les données sur la fécondité recueillies dans un recensement souffrent très souvent de deux erreurs. Premièrement, les données sur la fécondité cumulée au long de la vie tendent à se détériorer avec l'âge des mères. Deuxièmement, les informations sur la fécondité récente tendent à être systématiquement sous-déclarées par toutes les femmes (de même que les très jeunes enfants sont fréquemment sous-dénombrés dans l'effectif des ménages). Un sur-dénombrement des naissances récentes est également possible, lorsque les bornes de la période de référence sont mal comprises ou lorsque les naissances les plus récentes sont abusivement intégrées dans la période de référence. Les méthodes utilisées pour estimer la fécondité à partir des données de recensement visent explicitement à prendre en compte ces erreurs.

Troisièmement, en plus de ces questions synthétiques sur la fécondité, des enquêtes recueillent auprès des mères leur histoire génésique détaillée. De telles histoires incluent la date de naissance de chaque enfant, son statut vital (l'enfant est-il toujours vivant ?) et – si l'enfant est décédé – sa date de décès. Les données ainsi obtenues permettent de faire des estimations détaillées de la fécondité. Toutefois, en contrepartie de l'effort consacré au recueil de ces données détaillées, les échantillons de ces enquêtes sont souvent d'une taille limitée. En conséquence, les taux estimés sont soumis à une importante variabilité et ne permettent pas l'analyse de la fécondité à un niveau géographique fin et d'autres études différentielles.

TYPES DE MÉTHODES D'ESTIMATION DE LA FÉCONDITÉ

Les méthodes d'estimation de la fécondité sont étroitement associées aux types de données disponibles.

Estimation directe

En fonction des données disponibles, trois démarches sont possibles pour mesurer la fécondité directement.

La première recourt aux données du système d'enregistrement à l'état civil conjointement avec des estimations de population par âge et sexe (par exemple à partir d'un registre de population ou d'estimations de population en milieu d'année). Si les données du numérateur et du dénominateur sont à la fois complètes et non biaisées et que le dénominateur reflète bien la population exposée au risque de donner naissance par âge, le calcul des taux de fécondité est simple.

La deuxième démarche utilise les données de l'histoire génésique recueillies dans une enquête et les divers contrôles de validité qui peuvent être incorporés dans l'enquête. On dispose ainsi d'une information détaillée sur la naissance de chaque enfant, ainsi que sur l'âge de la mère. En conséquence, il est possible de déterminer exactement l'âge de la mère à la naissance de chaque enfant, et les naissances et l'exposition au risque peuvent être affectées aux différentes années de calendrier ou autres période de temps. Cette démarche est décrite en détail dans la section consacrée à l'estimation directe de la fécondité à partir des données d'enquête (chapitre 12).

La troisième approche s'appuie sur les mesures synthétiques de fécondité recueillies couramment dans les recensements pour estimer la fécondité récente. Les méthodes sont décrites dans la section consacrée à l'évaluation de la qualité des données de fécondité tirées des recensements, chapitre 5. Si les données souffrent de sous-déclaration des naissances récentes fréquente dans les recensements, les estimations qui en résulteront seront évidemment trop basses.

Estimation indirecte

L'estimation indirecte de la fécondité s'appuie sur les estimations directes tirées des informations synthétiques sur les naissances récentes, mais – sachant que les naissances récentes sont souvent mal déclarées dans les recensements – elle utilise aussi l'information sur leur descendance telle que donnée par les jeunes femmes au même recensement, afin d'ajuster les estimations directes. La plus ancienne de ces méthodes est la méthode P/F de Brass, exposée pour la première fois par Brass (1964). Elle est exposée dans le *Manuel X* (Divi-

sion de la population des NU 1984), en même temps que diverses variantes qui étendent ses possibilités d'application en fonction des données précisément disponibles. La section sur les méthodes d'estimation de la fécondité fondées sur le quotient P/F (chapitre 6) donne une brève description de la méthode P/F . Le *Manuel X* mentionnait en passant un perfectionnement de la méthode P/F – le modèle relationnel de Gompertz mais l'essentiel du développement de ce modèle est survenu après la publication du *Manuel X*. Nous présentons ici plusieurs extensions à partir du modèle relationnel de Gompertz (dont la section sur le modèle relationnel de Gompertz, chapitre 7, donne la version de base), en nous inspirant des extensions à la méthode P/F présentées dans le *Manuel X*. Ces extensions s'appliquent à des situations où :

- on dispose de données sur la descendance et sur la fécondité récente à partir de plusieurs recensements, et on cherche à estimer la fécondité pour la période intercensitaire – le modèle relationnel synthétique de Gompertz (chapitre 10).
- On ne dispose que de données sur la descendance – à partir de deux recensements ou enquêtes conduits à cinq ou dix ans d'écart – ce qui rend nécessaire une estimation de la fécondité à partir des accroissements des parités. C'est la méthode d'accroissement des parités présentée dans la section sur les estimations de fécondité tirées des accroissements des nombres d'enfants nés par cohorte, chapitre 11.
- On dispose de données sur les descendance à partir de deux recensements, ainsi que d'informations sur les naissances de la période intercensitaire, par exemple à partir d'un système d'enregistrement à l'état civil. Cette démarche est présentée dans la section sur la comparaison du nombre moyen de naissances enregistrées par une cohorte de femmes avec la parité moyenne déclarée de la même cohorte, chapitre 14. Elle permet d'évaluer la complétude des données sur les naissances enregistrées.

Dans le présent manuel, nous remanions toutes les méthodes d'estimation de la fécondité recourant à la méthode du quotient P/F en nous appuyant sur le modèle relationnel de Gompertz.

Autres analyses de la fécondité

Enfin, plusieurs autres méthodes peuvent éclairer les tendances et la dynamique de la fécondité.

À partir des données de recensement, on peut calculer à la fois des probabilités d'agrandissement des familles (chapitre 8) conventionnelles et projetées. Ces mesures

indiquent la propension des femmes d'une population à avoir d'autres enfants, en fonction du nombre de naissances qu'elles ont déjà eues. Les probabilités d'agrandissement projetées indiquent ce que pourrait être l'évolution future des probabilités d'agrandissement des jeunes femmes, en tenant compte de la fécondité actuelle et de l'histoire génésique des femmes jusqu'à la date du recensement.

Par ailleurs, des méthodes existent (la méthode du quotient P/F avec des données d'enquête, (chapitre 13) qui permettent le calcul de taux de fécondité par cohorte et période

à partir de données d'histoires génésiques détaillées. Ces taux fournissent une information sur les tendances de la fécondité, mais ils permettent aussi d'évaluer la qualité des données d'histoires génésiques.

Troisièmement, une méthode permet d'estimer des mesures de la fécondité à partir de la survie inversée (chapitre 9) de la population dénombrée des enfants et des adultes.

Toutes les méthodes citées plus haut sont décrites et discutées dans ce manuel.

RÉFÉRENCES ET LECTURES COMPLÉMENTAIRES

La meilleure description des méthodes d'estimation indirecte de la fécondité est dans le chapitre 2 du *Manuel X* (Division de la Population des NU 1984). Le manuel écrit par le Centre for Population Studies pour l'Institut Statistique des NU pour l'Asie et le Pacifique (Sloggett, Brass, Elridge *et al.* 1994) offre un exposé simple de l'évaluation de la qualité des données de fécondité, de l'utilisation du modèle relationnel de Gompertz pour l'estimation de la fécondité, et du calcul des probabilités d'agrandissement projetés. L'article de Brass (1964) a cédé la place à des développements plus récents, mais il demeure instructif par son exposé de la démarche qui a façonné presque toutes les méthodes ultérieures pour la mesure de la fécondité à partir de données limitées et déficientes.

Brass W. 1964. Utilisation des données des recensements ou des enquêtes pour l'estimation des taux de natalité et de mortalité. Préparé pour le Cycle d'études sur les statistiques d'état civil, Addis Abéba 14-19 décembre 1964. Document No. E/CN.14/CAS.4/V57. New York : Nations Unies. <http://repository.uneca.org/bitstream/handle/10855/9306/Bib-49868.pdf?sequence=1>

Division de la Population des Nations Unies. 1984. *Manuel X. Techniques indirectes d'estimation démographique*. New York : Nations Unies, Département des affaires économiques et sociales internationales, ST/ESA/SER.A/81. http://unstats.un.org/unsd/demographic/standmeth/handbooks/Manuel_X-fr.pdf

Sloggett A, W Brass, SM Eldridge, IM Timæus, P Ward and B Zaba (eds). 1994. *Estimation of Demographic Parameters from Census Data*. Tokyo: Statistical Institute for Asia and the Pacific.

Évaluation et correction des données de fécondité

INTRODUCTION

Tom A Moultrie

Les questions sur la fécondité généralement posées dans les recensements fournissent des données à la fois sur les naissances récentes et sur la fécondité cumulée au long de la vie. Dans les années 1960 Brass et ses collègues ont noté que ces données sont très souvent affectées par les mêmes erreurs. En particulier les mères de tous âges tendent à déclarer de façon erronée leurs naissances récentes suite à une combinaison d'erreurs sur la période de référence (les répondantes interprétant mal l'intervalle de temps sur lequel elles sont interrogées) et d'omission des décès néonataux.

Les femmes tendent également à sous déclarer leur nombre d'enfants déjà nés. On peut penser que ce défaut augmente avec l'âge. Les raisons avancées pour expliquer ce biais incluent la surestimation de leur âge par les mères adolescentes l'omission des enfants décédés et l'omission

des enfants âgés qui ont quitté le domicile parental. Une sur-déclaration des enfants déjà nés peut se produire quand les mères incluent abusivement des enfants accueillis parmi leurs « propres » enfants et des mort-nés parmi leurs naissances vivantes.

Dans cette section nous décrivons les recherches qui doivent être conduites pour évaluer la qualité des données sur la fécondité recueillies dans un recensement. Nous décrivons successivement

- l'évaluation de la qualité des données sur la parité y compris la correction d'El-Badry pour les nombres d'enfants considérées à tort comme « non déclarés » ; et
- l'évaluation de la qualité des données sur la fécondité récente . Cette section décrit aussi le processus d'estimation directe de la fécondité à partir des données de recensement.

Chapter 3 ÉVALUATION DE LA QUALITÉ DES DONNÉES SUR LES DESCENDANCES

Tom A Moultrie

INTRODUCTION

Le premier type de question sur la fécondité posée dans les recensements concerne la descendance des femmes. Celles-ci sont interrogées sur leur nombre total de naissances vivantes. En vue de réduire la sous-déclaration des enfants décédés ou absents (qui représente généralement une proportion plus élevée des enfants nés des femmes âgées que des femmes jeunes) et d'éviter la sous déclaration des filles, l'interrogation prend souvent la forme d'une série de six questions sur le nombre de fils et de filles :

- nés vivants et vivant actuellement avec la mère ;
- nés vivants mais vivant actuellement ailleurs ; et
- nés vivants mais aujourd'hui décédés.

TOTAL DES ENFANTS NÉS ET SURVIVANTS

La somme des réponses aux questions sur les enfants vivants, présents et absents, donne le nombre total d'enfants nés et survivants. Quand on y ajoute les nombres déclarés d'enfants décédés, on obtient le nombre total d'enfants déjà nés de la femme. En faisant l'addition de ces réponses, on doit prendre garde à ne pas assimiler les codes des valeurs erronées ou manquantes à des réponses numériques. Par exemple, si une valeur manquante est codée « 9 », la procédure pour établir le total des enfants déjà nés, survivants et décédés doit expressément exclure ces codes.

Les tabulations des nombres d'enfants déclarés en réponse à ces six questions sont souvent tronquées au-delà d'un nombre d'enfants relativement élevé (par exemple 9 ou +). Quand c'est le cas, la seule hypothèse plausible est que les femmes dans cette catégorie ont eu un nombre d'enfants égal à la limite inférieure de l'intervalle ouvert. Les erreurs qui en résultent sont généralement faibles, même quand la fécondité est extrêmement forte, sauf si la troncature s'ap-

plique au nombre total d'enfants déjà nés, plutôt qu'aux catégories séparées de garçons et filles présents, absents et décédés.

Parités peu vraisemblables

En évaluant la qualité des données sur la descendance, l'analyste doit être attentif aux parités peu probables et peu vraisemblables compte tenu de l'âge de la mère. En particulier aux jeunes âges, un petit nombre de jeunes femmes déclarant des nombres excessivement élevés d'enfants déjà nés peut avoir un impact sur le nombre moyen estimé d'enfants déjà nés. Il peut s'agir de déclarations erronées ou d'erreurs de saisie des données, que celle-ci soit manuelle ou automatique. En pratique, il est recommandé de limiter le nombre maximum de naissances vivantes qu'une femme peut avoir eu à une naissance tous les 18 mois depuis l'âge de 12 ans, en arrondissant vers le bas au nombre entier le plus proche. Selon ce principe, au 20^e anniversaire (une fois passé le groupe 15-19 ans), une femme ne peut avoir eu plus de 5 enfants, au 25^e (à l'issue des 20-24 ans), 8. Si la descendance déclarée dépasse ce maximum, elle doit être recodée en « valeur manquante ».

Évaluation des erreurs de l'agent recenseur

Une autre erreur courante dans l'enregistrement de la descendance survient lorsque, en l'absence d'enfants, l'agent recenseur laisse en blanc l'espace de réponse sur le questionnaire de recensement au lieu d'y inscrire un « zéro ». On ne peut alors savoir avec certitude si le blanc signifie que l'agent recenseur a omis de poser la question ou de reporter la réponse ou s'il s'agit d'une absence d'enfants. Cette erreur est généralement plus courante dans les données des jeunes femmes, qui ont davantage de chances

d'être sans enfants ou de répondre zéro aux six questions ci-dessus. L'erreur survient parfois parce que l'agent recenseur présume que la question ne s'applique pas aux jeunes femmes, ou parce qu'il est mal à l'aise pour la poser. On recommande souvent dans ce cas un ajustement spécifique des données, la correction d'El-Badry. Toutefois, si dans chaque groupe d'âge le nombre de femmes qui n'ont pas déclaré de nombre d'enfants est faible (disons moins de 2 % du total), cette erreur de déclaration aura peu d'impact sur le nombre moyen et pourra être ignorée lors des calculs ultérieurs. Ceci équivaut à faire l'hypothèse explicite que les femmes n'ayant pas déclaré de parité ont le même nombre moyen d'enfants que celles qui, dans le même groupe d'âge, l'ont déclarée.

Proportions de femmes sans enfants

Les proportions de femmes sans enfants doivent être calculées par groupe d'âge de la mère. Les proportions doivent reculer fortement quand l'âge augmente. Dans la plupart des cas, environ 3 à 10 % des femmes restent sans enfants dans les groupes d'âge les plus élevés, du fait de la stérilité primaire et de l'infécondité volontaire. Dans les pays de faible fécondité, la proportion de femmes sans enfants à 45-49 ans peut même être plus élevée. Quand les proportions de femmes sans enfants dépassent 10 % aux âges élevés des recherches complémentaires sont nécessaires, car cela peut indiquer d'importantes erreurs dans les données.

TABLEAU 3.1 Nombre total d'enfants déjà nés par groupe d'âge de la mère, Cambodge, recensement de 2008

| Parité | Groupe d'âge de la mère | | | | | | | Total |
|--------------|-------------------------|----------------|----------------|----------------|----------------|----------------|----------------|------------------|
| | 15-19 | 20-24 | 25-29 | 30-34 | 35-39 | 40-44 | 45-49 | |
| 0 | 743 190 | 426 760 | 191 720 | 58 530 | 46 650 | 36 050 | 28 780 | 1 531 680 |
| 1 | 29 560 | 167 810 | 142 720 | 44 310 | 34 530 | 25 790 | 21 740 | 466 460 |
| 2 | 4 240 | 78 410 | 171 450 | 90 990 | 79 080 | 51 980 | 36 680 | 512 830 |
| 3 | 1 200 | 16 940 | 82 960 | 84 220 | 98 640 | 67 690 | 48 190 | 399 840 |
| 4 | 830 | 4 020 | 26 870 | 48 510 | 79 480 | 70 400 | 56 190 | 286 300 |
| 5 | 430 | 1 340 | 6 910 | 21 010 | 49 250 | 56 980 | 51 500 | 187 420 |
| 6 | 270 | 630 | 2 150 | 8 710 | 26 020 | 37 070 | 41 420 | 116 270 |
| 7 | 120 | 380 | 630 | 3 410 | 12 530 | 23 730 | 29 680 | 70 480 |
| 8 | 80 | 200 | 400 | 1 000 | 5 450 | 12 180 | 18 320 | 37 630 |
| 9 | 60 | 100 | 120 | 350 | 2 410 | 6 030 | 10 040 | 19 110 |
| 10 | 40 | 120 | 140 | 190 | 1 090 | 3 120 | 5 660 | 10 360 |
| 11 | 50 | 0 | 70 | 70 | 360 | 1 420 | 2 010 | 3 980 |
| 12 | 20 | 50 | 20 | 30 | 170 | 670 | 1 350 | 2 310 |
| 13 | 10 | 10 | 0 | 10 | 60 | 270 | 410 | 770 |
| 14 | 0 | 10 | 10 | 0 | 10 | 60 | 190 | 280 |
| 15 | 0 | 0 | 10 | 0 | 20 | 90 | 150 | 270 |
| 16 | 0 | 0 | 0 | 0 | 0 | 10 | 30 | 40 |
| 17 | 0 | 0 | 0 | 0 | 0 | 10 | 30 | 40 |
| 18 | 0 | 0 | 0 | 0 | 0 | 0 | 20 | 20 |
| 19 | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 10 |
| 20 | 0 | 0 | 0 | 20 | 0 | 0 | 0 | 20 |
| Inconnue | 220 | 380 | 250 | 290 | 130 | 210 | 120 | 1 600 |
| TOTAL | 780 320 | 697 160 | 626 430 | 361 650 | 435 880 | 393 760 | 352 520 | 3 647 720 |

PARITÉS MOYENNES

Une distribution vraisemblable des nombres moyens d'enfants vivants et décédés par groupe d'âge de la mère est une indication essentielle de la cohérence des données sur la descendance. En général, on s'attend à ce que les parités moyennes (le nombre moyen total d'enfants nés, présents, absents et décédés) s'accroissent continuellement avec l'âge des femmes. La forme de la distribution par âge doit être sigmoïde, avec des sections un peu plus plates au début et à la fin, du fait d'une moindre fécondité aux âges les plus jeunes et les plus avancés auxquels les femmes ont des enfants. D'importantes augmentations du nombre d'enfants

à ces âges – c'est-à-dire de forts accroissements des parités moyennes entre groupes d'âge successifs – sont peu vraisemblables.

On peut aussi s'attendre à ce que les nombres moyens d'enfants vivants et d'enfants décédés, ainsi que la proportion d'enfants décédés augmentent avec l'âge.

Un deuxième contrôle consiste à comparer les parités observées aux résultats obtenus dans des Enquêtes Démographiques et de Santé (EDS), ou lors de recensements antérieurs ou dans d'autres enquêtes. On peut alors comparer les parités moyennes dans des générations féminines réelles. Ainsi, si deux recensements sont réali-

TABLEAU 3.2 Nombre total d'enfants déjà nés par groupe d'âge de la mère après correction pour les parités peu vraisemblables, Cambodge, recensement de 2008

| Parité | Groupe d'âge de la mère | | | | | | | Total |
|-----------------------|-------------------------|----------------|----------------|----------------|----------------|----------------|----------------|------------------|
| | 15-19 | 20-24 | 25-29 | 30-34 | 35-39 | 40-44 | 45-49 | |
| 0 | 743 190 | 426 760 | 191 720 | 58 530 | 46 650 | 36 050 | 28 780 | 1 531 680 |
| 1 | 29 560 | 167 810 | 142 720 | 44 310 | 34 530 | 25 790 | 21 740 | 466 460 |
| 2 | 4 240 | 78 410 | 171 450 | 90 990 | 79 080 | 51 980 | 36 680 | 512 830 |
| 3 | 1 200 | 16 940 | 82 960 | 84 220 | 98 640 | 67 690 | 48 190 | 399 840 |
| 4 | 830 | 4 020 | 26 870 | 48 510 | 79 480 | 70 400 | 56 190 | 286 300 |
| 5 | 430 | 1 340 | 6 910 | 21 010 | 49 250 | 56 980 | 51 500 | 187 420 |
| 6 | 0 | 630 | 2 150 | 8 710 | 26 020 | 37 070 | 41 420 | 116 000 |
| 7 | 0 | 380 | 630 | 3 410 | 12 530 | 23 730 | 29 680 | 70 360 |
| 8 | 0 | 200 | 400 | 1 000 | 5 450 | 12 180 | 18 320 | 37 550 |
| 9 | 0 | 0 | 120 | 350 | 2 410 | 6 030 | 10 040 | 18 950 |
| 10 | 0 | 0 | 140 | 190 | 1 090 | 3 120 | 5 660 | 10 200 |
| 11 | 0 | 0 | 70 | 70 | 360 | 1 420 | 2 010 | 3 930 |
| 12 | 0 | 0 | 20 | 30 | 170 | 670 | 1 350 | 2 240 |
| 13 | 0 | 0 | 0 | 10 | 60 | 270 | 410 | 750 |
| 14 | 0 | 0 | 0 | 0 | 10 | 60 | 190 | 260 |
| 15 | 0 | 0 | 0 | 0 | 20 | 90 | 150 | 260 |
| 16 | 0 | 0 | 0 | 0 | 0 | 10 | 30 | 40 |
| 17 | 0 | 0 | 0 | 0 | 0 | 10 | 30 | 40 |
| 18 | 0 | 0 | 0 | 0 | 0 | 0 | 20 | 20 |
| 19 | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 10 |
| 20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Inconnue | 870 | 670 | 270 | 310 | 130 | 210 | 120 | 2 580 |
| TOTAL | 780 320 | 697 160 | 626 430 | 361 650 | 435 880 | 393 760 | 352 520 | 3 647 720 |
| Proportion manquante | 0,111 % | 0,096 % | 0,043 % | 0,086 % | 0,030 % | 0,053 % | 0,034 % | |
| Proportion ss enfants | 95,24 % | 61,21 % | 30,61 % | 16,18 % | 10,70 % | 9,16 % | 8,16 % | |
| Parités moyennes | 0,0604 | 0,5833 | 1,4382 | 2,4035 | 3,1670 | 3,8126 | 4,3184 | |

sés à 10 ans d'intervalle, les parités moyennes des femmes âgées de x à $x + 4$ au premier recensement peuvent être comparées à celles des femmes âgées de $x + 10$ à $x + 14$ au second. Non seulement les nombres moyens d'enfants devraient s'accroître continuellement avec l'âge à chaque recensement, mais il devrait aussi y avoir un accroissement raisonnable du nombre d'enfants au sein des cohortes entre deux recensements.

Si on dispose de données pour des femmes âgées de 50 ans et plus, on peut comparer directement la cohérence des nombres moyens des femmes qui ont terminé leur période féconde – par exemple en comparant les parités moyennes des femmes âgées de 45-49 ans dans un recensement avec celles des femmes âgées de 55-59 ans au second recensement dix ans plus tard. Lors de telles comparaisons, en particulier lorsqu'elles impliquent des femmes âgées, l'analyste doit être attentif à d'éventuelles différences de mortalité entre les femmes ayant eu des nombres d'enfants différents soit du fait d'un lien direct, soit parce que forte fécondité et statut socio-économique peuvent être corrélés. Ceci peut empêcher de tirer des conclusions définitives sur les tendances de la descendance finale.

Feeney (1991) a suggéré un perfectionnement de l'analyse lorsque on dispose d'informations sur la parité moyenne de femmes ayant terminé leur vie féconde : il s'agit de situer ces moyennes approximativement dans le temps et de les représenter graphiquement. La localisation dans le temps consiste à situer les parités moyennes en un point du temps défini en retranchant le point médian de chaque groupe d'âge à la date du recensement et en supposant que toutes les naissances dans chaque génération se sont produites à l'âge moyen à la maternité, m . En supposant par exemple $m = 27,5$ ans, si un recensement a eu lieu en 1960, les nombres moyens des femmes âgées de 50-54 ans se rapportent (à peu près) à $1960 - 52,5 + 27,5$, soit 1935.

La parité moyenne des femmes d'un âge donné x , P_x , est calculée en divisant le nombre total d'enfants déjà nés des femmes âgées de x à la date du recensement par le nombre de femmes âgées de x au recensement :

$$P_x = \frac{\sum_{j=0}^{\omega} j \cdot N_{x,j}}{\sum_{j=0}^{\omega} N_{x,j}}$$

où $N_{x,j}$ est le nombre de femmes âgées de x et ayant j enfants dans la population, et oméga (ω) est la limite supérieure du nombre d'enfants déclaré dans la population, après exclusion des valeurs numériques considérées comme des codes erronés dans les données. Avec des groupes d'âges quinquennaux, le nombre moyen d'enfants des femmes dans chaque groupe d'âge est égal à

$${}_5P_x = \frac{\sum_{j=0}^{\omega} j \cdot {}_5N_{x,j}}{\sum_{j=0}^{\omega} {}_5N_{x,j}}$$

pour $x = 15, 20, \dots, 45$.

Pour simplifier l'exposé de nombreuses méthodes, les nombres moyens d'enfants par groupes d'âges quinquennaux, 15-19, 20-24, ... sont souvent désignés par $P(i)$, $i = 1, 2, \dots$, où $P(1)$ se réfère au groupe 15-19 ans $P(2)$ au groupe 20-24 ans, etc.

Comparaison avec d'autres estimations des parités moyennes

Si d'autres données sur la fécondité sont disponibles pour le même pays à peu près à la même date, il est bon de comparer les estimations. Si elles divergent significativement, l'analyste doit s'efforcer de comprendre pourquoi, mais il sera souvent impossible de conclure définitivement laquelle des deux sources est déficiente.

Comparaison avec les indices synthétiques de fécondité

Un dernier contrôle résulte de la comparaison de la parité moyenne des femmes de 45-49 ans avec une estimation de l'indice synthétique de fécondité tirée des données sur la fécondité récente. Si la fécondité est constante depuis longtemps et que les données sont déclarées précisément, les deux mesures doivent être proches l'une de l'autre puisque la fécondité du moment et celle des cohortes doivent être égales dans ces conditions. Si la fécondité a baissé, la parité moyenne des femmes âgées doit être supérieure à l'indice synthétique de fécondité. Comme la sous-déclaration de la fécondité récente réduit artificiellement l'indice synthétique, alors que l'omission des naissances par les femmes âgées réduit la parité moyenne dans ce groupe, il est important de vérifier que chacune des deux mesures est plausible. Une des méthodes pour le faire s'appuie sur le modèle relationnel de Gompertz pour étudier les distributions des fécondités et des parités et leurs relations.

EXEMPLE : ÉVALUATION DE LA QUALITÉ DES DONNÉES SUR LES DESCENDANCES

L'exemple ci-dessous repose sur les données tirées du recensement de 2008 au Cambodge diffusées par IPUMS. Les données (pondérées, pour tenir compte du fait que les données d'IPUMS ne sont qu'un échantillon de l'ensemble des données) figurent au tableau 3.1.

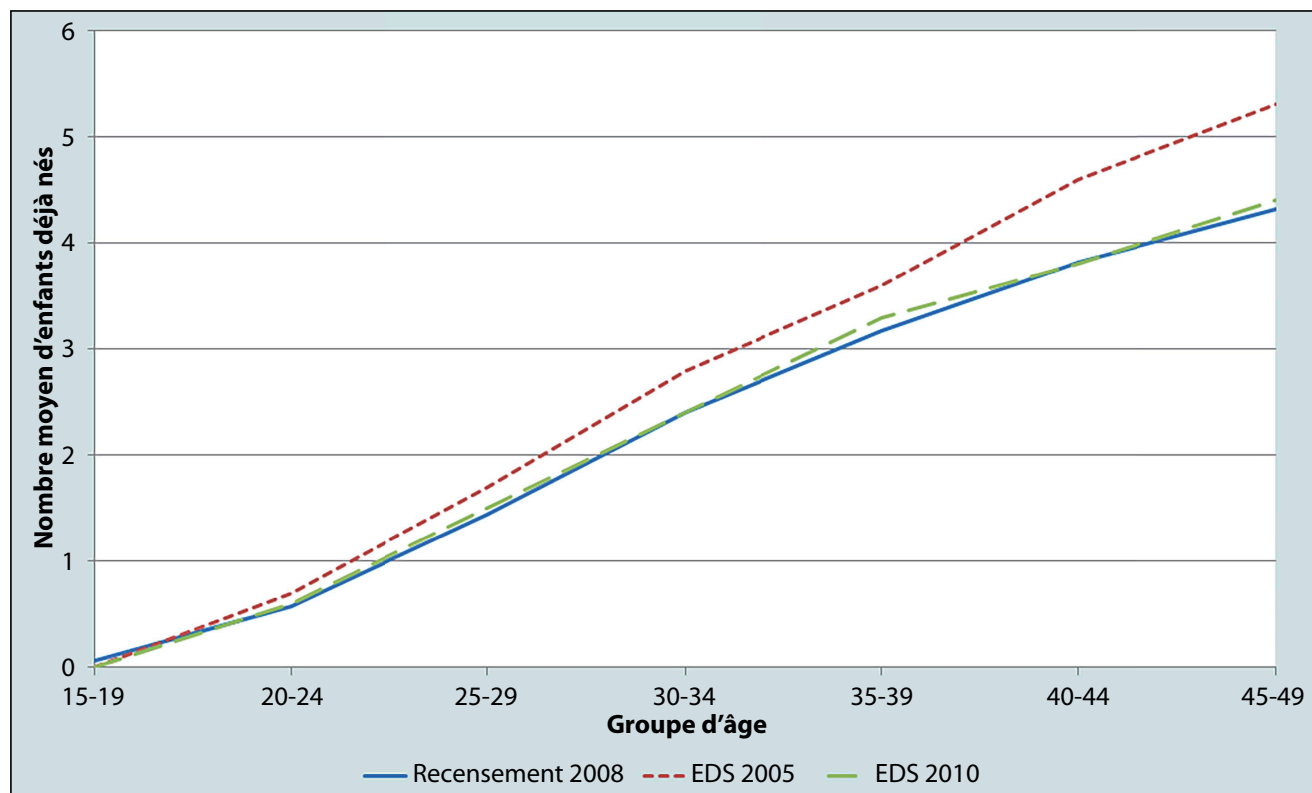
Les nombres en *italiques* correspondent à des parités peu vraisemblables selon la règle pratique exposée plus haut. Les valeurs dans ces cellules sont ajoutées aux effectifs de femmes de chaque groupe d'âge dont la parité est inconnue. Les valeurs originelles sont ensuite mises à zéro, le résultat figurant au tableau 3.2.

La proportion de femmes dont la parité est inconnue après cet ajustement figure sur l'antépénultième ligne du tableau 3.2. Dans chaque groupe d'âge, cette proportion est insignifiante. Elle est un peu plus élevée chez les

jeunes femmes que chez les plus âgées, mais même chez les 15-19 ans seules 0,11 % des femmes ont une parité inconnue ou peu vraisemblable. Il n'est donc pas nécessaire de faire une correction d'El-Badry et les cas inconnus peuvent être exclus du calcul des parités moyennes, en supposant donc implicitement que les femmes dont les données sont manquantes ou peu vraisemblables ont les mêmes parités moyennes que les autres femmes du même âge. (Les données présentées ici ont été choisies parce qu'une correction d'El-Badry n'était pas nécessaire. La section du manuel traitant de la correction d'El-Badry (chapitre 4) présente le cas d'un autre pays dont les données de parité ne sont pas d'aussi bonne qualité.)

La proportion de femmes sans enfants, à l'avant dernière ligne du tableau 3.2 recule rapidement avec l'âge ; à 40 ans, moins de 10 % des femmes n'ont pas eu d'enfants. Comme attendu, cette proportion ne diminue que

FIGURE 3.1 Parités moyennes par groupe d'âge, Cambodge, recensement de 2008, EDS 2005 et EDS 2010



lentement ensuite entre les deux derniers groupes d'âge : peu de femme débutent leur vie féconde après 40 ans. La proportion de femmes sans enfants à 45-49 ans (8,2 %) est relativement élevée. Les parités moyennes laissent penser que la fécondité des adolescentes est très faible, la descendance finale atteignant par ailleurs 4,3 enfants par femme dans le groupe 45-49 ans. La représentation graphique des parités moyennes a une forme sigmoïde, les hausses de descendance les plus fortes se produisant entre 20 et 35 ans, aux âges où la fécondité est généralement la plus élevée. (Figure 3.1).

Sur la figure 3.1, on a également représenté les parités moyennes par groupe d'âge d'après les Enquêtes Démographiques et de Santé du Cambodge en 2005 et 2010 (disponibles sur le site internet des EDS www.statcompiler.com). Les parités moyennes au recensement et dans l'EDS de 2010

sont très voisines. Deux caractéristiques suggèrent cependant qu'il faut être prudent avant de conclure que ces données sont de bonne qualité. Premièrement, étant donné les dates des trois sources, les données du recensement devraient se situer à peu près à mi-chemin entre les estimations des deux EDS, ce qui n'est pas le cas. Deuxièmement, la descendance des femmes âgées de 40-44 ans à l'enquête EDS de 2005 est un peu plus élevée (de 0,2 enfant) que celle des femmes âgées de 45-49 ans à l'enquête EDS de 2010. La fécondité est faible chez les femmes à l'approche de la cinquantaine et des erreurs aléatoires ne sont pas à exclure, mais dans tous les cas ces résultats poussent à un certain scepticisme à l'égard des données. Toutefois dans l'ensemble, les parités moyennes tirées des deux EDS ne contredisent pas fondamentalement celles tirées des recensements.

RÉFÉRENCES

Feeney G. 1991. "Child survivorship estimation: Methods and data analysis", *Asian and Pacific Population Forum* 5(2-3):51-55, 76-87. <http://hdl.handle.net/10125/3600>.

Chapitre 4 LA CORRECTION D'EL-BADRY

Tom A Moultrie

DESCRIPTION DE LA MÉTHODE

La correction d'El-Badry consiste à corriger les erreurs dans les données sur les enfants déjà nés dues au fait que les agents recenseurs laissent vide la réponse à la question sur la descendance, quand il faudrait indiquer « zéro ». Quand ceci survient, la réponse est codée « manquante » ou « inconnue » lors de l'exploitation des données, même s'il était évident pour l'agent recenseur au moment de la collecte des données que la bonne réponse était « zéro ». La méthode répartit le nombre de femmes dont la parité est enregistrée comme « manquante » entre celles dont la parité est considérée comme réellement inconnue, et celles qui auraient dû être enregistrées comme sans enfants mais dont les réponses ont été laissées vides. Cette répartition est faite au niveau agrégé pas au niveau individuel.

DONNÉES NÉCESSAIRES ET HYPOTHÈSES

La méthode s'appuie sur le nombre d'enfants déjà nés, classé par groupe d'âge de la mère, y compris les femmes dont les données sont manquantes (c'est-à-dire dont la réponse a été laissée vide ou a été codée hors de l'intervalle des possibles ou a été codée « non réponse » ou « refus »).

La méthode suppose qu'une proportion constante de femmes à chaque âge n'a effectivement pas répondu à la question sur la parité lors de la collecte des données. Les autres femmes sans réponse sont supposées être à tort considérées comme non répondantes, alors qu'elles sont en fait sans enfants.

PRÉCAUTIONS ET MISES EN GARDE

La méthode repose sur l'existence d'une relation linéaire entre les proportions de femmes dont la parité n'est pas déclarée et celles des femmes déclarées sans enfants. Si on observe une telle relation linéaire, le dénominateur ajusté utilisé dans le calcul des parités moyennes doit *exclure* les femmes dont la parité (après correction) continue d'être considérée

comme inconnue. Ceci reflète l'hypothèse implicite que la distribution de ces femmes par parité ne diffère pas de celle des femmes du même âge dont la parité est connue.

Lorsque les données indiquent la nécessité d'une correction parce que la proportion d'informations manquantes sur la parité est importante, mais que la méthode ne peut pas être appliquée (par exemple parce qu'on n'a pas de données par âge ou parce que l'hypothèse de linéarité n'est pas respectée), les femmes de parité inconnue doivent être incluses dans le dénominateur utilisé pour calculer les parités moyennes. L'hypothèse implicite est alors que la parité de ces femmes est nulle (c'est-à-dire que toutes les femmes de parité inconnue sont sans enfants). Il en résulte évidemment une sous-estimation des parités moyennes, puisque en fait une partie des femmes de parité inconnue ne sont pas sans enfants.

APPLICATION DE LA MÉTHODE

Nous définissons $N_i = {}_5N_a$ comme étant le nombre de femmes dans le groupe d'âge i dans la population, pour $a = 15, 20, \dots, 45$, et $i = a/5 - 2$. Ainsi, N_1 représente le nombre de femmes âgées de 15-19 ans dans la population. Nous notons N_{ij} le nombre de femmes du groupe d'âge i de parité j et $N_{i,u}$ le nombre de femmes du groupe d'âge i dont la parité est inconnue.

Étape 1 : Déterminer la proportion de femmes dans chaque groupe d'âge dont la parité est a) non déclarée et b) déclarée nulle

Extraire des données du recensement un tableau du nombre d'enfants déjà nés (j) par groupe d'âge des femmes (i) pour obtenir $N_{i,j}$. Pour chaque groupe d'âge, les données où la parité manque (c'est-à-dire les réponses laissées vides et les codes non valables) doivent être combinées aux codes de parité non déclarée pour établir $N_{i,u}$. La proportion de femmes du groupe d'âge i dont la parité est inconnue est alors

$$U_i = \frac{N_{i,u}}{N_i}$$

La proportion de femmes du groupe d'âge i qui se sont déclarées sans enfants (c'est-à-dire de parité zéro) est donnée par

$$Z_i = \frac{N_{i,0}}{N_i}$$

Si les U_i sont faibles (moins de 2 % dans chaque groupe d'âge), la correction n'en vaut pas la peine. Dans ce cas, les parités moyennes doivent être calculées en supposant que la distribution par parité des femmes dont la parité est inconnue est la même que celle des femmes dont la parité est connue, en omettant les femmes dont la parité est inconnue du dénominateur lors du calcul. Ainsi, si P_i est la parité moyenne des femmes du groupe d'âge i ,

$$P_i = \frac{\sum_{j=0}^{\omega} j \cdot N_{i,j}}{\sum_{j=0}^{\omega} N_{i,j}}$$

Si les proportions de femmes dont la parité n'est pas établie dépassent 2 %, il est utile de voir s'il est possible d'appliquer la correction.

Étape 2 : Représenter graphiquement les points (Z_i , U_i) et évaluer la qualité des données

Pour que la méthode fonctionne correctement, la série de points (Z_i , U_i) doit se situer sur une droite, ou très proche d'une droite. Dans certains cas, une courbure peut être observée aux points correspondant soit aux jeunes âges soit aux âges avancés. Si la courbure affecte seulement les âges avancés, même si elle est forte, on peut exclure de l'ajustement le groupe d'âge le plus élevé, ou les deux plus élevés, et ajuster une droite sur les points restants puisque la méthode a son impact absolu le plus fort sur les proportions non déclarées aux plus jeunes âges. Si la courbure est surtout prononcée chez les jeunes femmes, le recours à la méthode est impossible car une régression excluant les points relatifs aux femmes âgées de 15-24 ans conduirait à une extrapolation hors de l'échantillon, qui pourrait suggérer des ajustements illogiques dans ces groupes d'âge.

Si on ne peut pas identifier clairement une relation, même après avoir exclu un ou deux points relatifs aux femmes âgées, la méthode ne peut pas être appliquée. Dans un tel cas, il est préférable de supposer que toutes les femmes sans parité établie sont sans enfants, et de les inclure dans le dénominateur du calcul de la parité moyenne

$$P_i = \frac{\sum_{j=0}^{\omega} j \cdot N_{i,j}}{N_i} \quad (1)$$

Le rapport d'analyse doit indiquer qu'il a été procédé ainsi, et que par conséquent les valeurs des parités moyennes sont susceptibles d'être sous-estimées.

Étape 3 : Déterminer la pente et l'ordonnée à l'origine de la meilleure droite d'ajustement aux données

La pente (γ) et l'ordonnée à l'origine (β) de la droite d'ajustement sont déterminées au moyen d'une régression linéaire de U_i sur Z_i appliquée aux points retenus, c'est-à-dire $U_i = \beta + \gamma Z_i$.

L'ordonnée à l'origine (β), qui est indépendante de l'âge (i), est l'estimation de la proportion des femmes dans chaque groupe d'âge de parité inconnue dont la parité est considérée comme réellement inconnue, et non mal déclarée.

Étape 4 : Estimation révisée des nombres de femmes sans enfants et de femmes dont la parité n'est pas déclarée

La proportion ajustée des femmes du groupe d'âge i qu'on a estimées être vraiment sans enfants est donnée par $Z_i^* = Z_i + U_i - \beta$.

La proportion révisée de femmes de parité zéro est donc égale à la proportion effectivement déclarée de parité zéro, à laquelle s'ajoute la proportion de femmes de ce groupe d'âge de parité non déclarée et se retranche la proportion estimée de femmes dont la parité est considérée comme vraiment inconnue. L'estimation révisée du nombre de femmes sans enfants dans le groupe d'âge i est donnée par $N_{i,0}^* = N_i \times Z_i^*$. L'estimation de la proportion réelle de femmes de chaque groupe d'âge dont la parité est inconnue est donc donnée par $N_{i,u}^* = N_i \times \beta$. Les $N_{i,j}^*$ pour les autres parités ($j > 0$) restent inchangées.

Étape 5 : Calcul des parités moyennes

Si une correction d'El-Badry a été appliquée aux données, les parités moyennes sont données par :

$$P_i = \frac{\sum_{j=0}^{\omega} j \cdot N_{i,j}^*}{(1 - \beta)N_i} \quad (2)$$

reflétant l'hypothèse selon laquelle les femmes dans le groupe d'âge i qui restent de parité inconnue, βN_i – qui sont omises au dénominateur – ont la même parité moyenne que les femmes du groupe d'âge i dont la parité est connue.

INTERPRÉTATION ET CONTRÔLES

La valeur de β mesure la proportion estimée de femmes dont la parité n'est réellement pas déclarée. Plus β est élevé, moins bonne est la qualité des données.

La méthode peut parfois avoir un effet contraire, en donnant à penser que le nombre de femmes de parité non déclarée est sous-estimé et que le nombre de femmes ayant déclaré une parité zéro doit être réduit. C'est le cas lorsque $\beta > U_i$. La correction ne doit alors pas être appliquée dans ce groupe d'âge.

EXEMPLE

La feuille de calcul disponible sur internet à <http://demographicestimation.iussp.org/fr> met la méthode en application sur les données du recensement du Kenya en 1989, diffusées par IPUMS. Les données originelles sont au tableau 4.1.

L'examen des données montre qu'elles ont été corrigées pour éliminer les déclarations de parités élevées par les femmes de moins de 35 ans. La règle de correction appliquée lors de la phase préparatoire a été plus stricte que celle que nous avons suggérée à la section sur l'évaluation de la qualité des données sur les descendance. Ainsi les déclarations de femmes âgées de 20-24 ans ont été limitées aux parités 6 ou moins (plutôt que 8), les déclarations des femmes âgées de 25-29 ans ont été tronquées à la parité 7 (au lieu de 12) et celles des femmes âgées de 30-34 ans à la parité 8 (et non 15). En revanche, des parités peu vraisemblables ont été conservées à partir de 35 ans. C'est pourquoi nous pouvons corriger légèrement les données en italiques du tableau 4.1 en plaçant dans la catégorie Non déclarée les déclarations de parité supérieures à 18 pour le groupe 35-39 ans, supérieures à 22 dans le groupe 40-44 ans, et supérieures à 25 dans le dernier groupe, 45-49 ans.

Vous pouvez sélectionner une option dans le tableau **Introduction** de la feuille de calcul pour transférer les parités peu vraisemblables dans la catégorie « Non déclarée » avant l'application de la méthode.

Étape 1 : Déterminer la proportion de femmes dans chaque groupe d'âge dont la parité est a) non déclarée et b) déclarée nulle

Le tableau 4.2 présente les données révisées, en même temps que le calcul de la proportion de femmes de parité zéro, et de parité nulle dans chaque groupe d'âge.

Les données incluent des proportions élevées de femmes de parité non déclarée à 15-19 ans $\left(\frac{402\,780}{1\,192\,840} = 0,338 \right)$, à 20-24 ans (0,147) et, dans une moindre mesure, dans les groupes d'âge plus avancé. La proportion de femmes se déclarant sans enfants (Z_i) recule rapidement, de 50 % environ dans le premier groupe d'âge à 3 % environ à la fin de la période féconde. Sur cette base, il est utile de voir si une correction d'El-Badry peut être appliquée aux données.

Étape 2 : Représenter graphiquement les points (Z_i , U_i) et évaluer la qualité des données

Les points correspondant à chaque paire de Z_i et U_i sont portés sur la figure 4.1 (sous forme de losanges bleus). La ligne droite ajustée aux points est représentée en rouge. Si un point est exclu de l'ajustement, il est représenté sur la feuille de calcul par un losange sans couleur.

Étape 3 : Déterminer la pente et l'ordonnée à l'origine de la meilleure droite d'ajustement aux données

La régression linéaire de U_i sur Z_i pour les points sélectionnés donne comme valeur de l'ordonnée à l'origine (béta) 0,02745. Ce résultat conduit à penser qu'environ 2,7 % des données sur les parités des femmes peuvent être considérées comme réellement manquantes.

Étape 4 : Estimation révisée des nombres de femmes sans enfants, et femmes dont la parité n'est pas déclarée

Le nombre révisé de femmes de parité zéro est donné par $N_{i,0}^* = N_i (Z_i + U_i - \beta)$, alors que les nombres révisés des femmes de parité inconnue sont calculés en multipliant le nombre total de femmes dans chaque groupe d'âge par β , comme le montre le tableau 4.3. Par exemple, le nombre de femmes âgées de 20-24 ans qu'on estime vraiment de parité inconnue est donné par $0,02745 \times 1\,005\,500 = 27\,603$. L'estimation corrigée du nombre de femmes sans enfants âgées de 15-19 ans est donnée par :

$$1\,192\,840 \times (0,501 + 0,338 - 0,027) = 967\,594.$$

Étape 5 : Calcul des parités moyennes

Puisqu'on a appliqué une correction d'El-Badry, l'équation 2 permet d'obtenir les parités moyennes corrigées figurant au tableau 4.4.

Par rapport aux parités moyennes obtenues sans application de la correction (en supposant donc que les femmes sans parité déclarée sont de parité zéro), la

correction accroît les parités dans chaque groupe d'âge par une constante, $1/(1-\beta)$.

DESCRIPTION DÉTAILLÉE DE LA MÉTHODE

La méthode est décrite entièrement dans *El-Badry (1961)*. L'idée fondamentale d'El-Badry est que, si on peut supposer que :

TABLEAU 4.1 Enfants déjà nés, par groupe d'âge de la mère à la date de recensement, Kenya, recensement de 1989

| Parité | Groupe d'âge (i) | | | | | | |
|--------------|------------------|-----------|-----------|-----------|-----------|-----------|-----------|
| | 15-19 (1) | 20-24 (2) | 25-29 (3) | 30-34 (4) | 35-39 (5) | 40-44 (6) | 45-49 (7) |
| 0 | 597 560 | 198 600 | 59 400 | 23 120 | 14 580 | 11 040 | 9 560 |
| 1 | 134 700 | 224 660 | 83 140 | 26 140 | 13 620 | 9 460 | 7 740 |
| 2 | 38 120 | 202 300 | 120 940 | 38 340 | 19 180 | 13 240 | 9 280 |
| 3 | 11 120 | 126 500 | 150 500 | 53 880 | 28 020 | 17 000 | 12 440 |
| 4 | 6 820 | 59 700 | 146 500 | 73 280 | 37 340 | 21 400 | 14 800 |
| 5 | 1 740 | 33 720 | 102 300 | 87 720 | 48 140 | 28 980 | 18 560 |
| 6 | 0 | 12 480 | 58 980 | 83 580 | 56 520 | 35 260 | 26 280 |
| 7 | 0 | 0 | 57 180 | 91 800 | 56 240 | 41 260 | 28 640 |
| 8 | 0 | 0 | 0 | 64 740 | 56 560 | 42 700 | 32 920 |
| 9 | 0 | 0 | 0 | 0 | 40 780 | 39 480 | 33 000 |
| 10 | 0 | 0 | 0 | 0 | 26 840 | 32 240 | 27 920 |
| 11 | 0 | 0 | 0 | 0 | 14 920 | 22 840 | 21 920 |
| 12 | 0 | 0 | 0 | 0 | 8 280 | 14 660 | 14 720 |
| 13 | 0 | 0 | 0 | 0 | 3 740 | 7 900 | 8 920 |
| 14 | 0 | 0 | 0 | 0 | 2 180 | 4 080 | 4 900 |
| 15 | 0 | 0 | 0 | 0 | 1 260 | 2 100 | 2 860 |
| 16 | 0 | 0 | 0 | 0 | 960 | 1 200 | 1 540 |
| 17 | 0 | 0 | 0 | 0 | 520 | 680 | 1 000 |
| 18 | 0 | 0 | 0 | 0 | 420 | 520 | 620 |
| 19 | 0 | 0 | 0 | 0 | 140 | 340 | 380 |
| 20 | 0 | 0 | 0 | 0 | 160 | 300 | 280 |
| 21 | 0 | 0 | 0 | 0 | 240 | 160 | 280 |
| 22 | 0 | 0 | 0 | 0 | 40 | 100 | 60 |
| 23 | 0 | 0 | 0 | 0 | 20 | 20 | 80 |
| 24 | 0 | 0 | 0 | 0 | 60 | 20 | 80 |
| 25 | 0 | 0 | 0 | 0 | 60 | 40 | 0 |
| 26 | 0 | 0 | 0 | 0 | 60 | 40 | 80 |
| 27 | 0 | 0 | 0 | 0 | 80 | 40 | 60 |
| 28 | 0 | 0 | 0 | 0 | 20 | 40 | 40 |
| 29 | 0 | 0 | 0 | 0 | 20 | 0 | 40 |
| 30 | 0 | 0 | 0 | 0 | 340 | 440 | 360 |
| Non déclarée | 402 780 | 147 540 | 61 920 | 31 580 | 20 240 | 15 420 | 12 960 |
| TOTAL | 1 192 840 | 1 005 500 | 840 860 | 574 180 | 451 580 | 363 000 | 292 320 |

- 1) Il existe une relation linéaire entre les proportions de femmes sans enfants d'un âge donné dans la population et la proportion de femmes dont la parité n'est pas déclarée, et
- 2) La proportion réelle, inconnue, de femmes dont la parité n'est pas connue est une constante indépendante de l'âge, alors

$$U_i = \alpha Z_i^* + \beta \quad (3)$$

où αZ_i^* est la proportion de femmes réellement sans enfants déclarées de parité inconnue, et β est la proportion réelle, constante, de femmes sans parité déclarée.

Donc, si αZ_i^* ont été classées par erreur comme non déclarées, alors qu'elles sont vraiment sans enfants, alors $Z_i = Z_i^* - \alpha Z_i^* = (1 - \alpha) Z_i^*$ et donc

$$Z_i^* = \frac{Z_i}{(1 - \alpha)} \quad (4)$$

TABLEAU 4.2 Correction des données de parité, et calcul de la proportion de femmes de parité zéro et de parité non déclarée, Kenya, recensement de 1989

| Parité | Groupe d'âge (i) | | | | | | |
|----------------|------------------|-----------|-----------|-----------|-----------|-----------|-----------|
| | 15-19 (1) | 20-24 (2) | 25-29 (3) | 30-34 (4) | 35-39 (5) | 40-44 (6) | 45-49 (7) |
| 0 | 597 560 | 198 600 | 59 400 | 23 120 | 14 580 | 11 040 | 9 560 |
| 1 | 134 700 | 224 660 | 83 140 | 26 140 | 13 620 | 9 460 | 7 740 |
| 2 | 38 120 | 202 300 | 120 940 | 38 340 | 19 180 | 13 240 | 9 280 |
| 3 | 11 120 | 126 500 | 150 500 | 53 880 | 28 020 | 17 000 | 12 440 |
| 4 | 6 820 | 59 700 | 146 500 | 73 280 | 37 340 | 21 400 | 14 800 |
| 5 | 1 740 | 33 720 | 102 300 | 87 720 | 48 140 | 28 980 | 18 560 |
| 6 | 0 | 12 480 | 58 980 | 83 580 | 56 520 | 35 260 | 26 280 |
| 7 | 0 | 0 | 57 180 | 91 800 | 56 240 | 41 260 | 28 640 |
| 8 | 0 | 0 | 0 | 64 740 | 56 560 | 42 700 | 32 920 |
| 9 | 0 | 0 | 0 | 0 | 40 780 | 39 480 | 33 000 |
| 10 | 0 | 0 | 0 | 0 | 26 840 | 32 240 | 27 920 |
| 11 | 0 | 0 | 0 | 0 | 14 920 | 22 840 | 21 920 |
| 12 | 0 | 0 | 0 | 0 | 8 280 | 14 660 | 14 720 |
| 13 | 0 | 0 | 0 | 0 | 3 740 | 7 900 | 8 920 |
| 14 | 0 | 0 | 0 | 0 | 2 180 | 4 080 | 4 900 |
| 15 | 0 | 0 | 0 | 0 | 1 260 | 2 100 | 2 860 |
| 16 | 0 | 0 | 0 | 0 | 960 | 1 200 | 1 540 |
| 17 | 0 | 0 | 0 | 0 | 520 | 680 | 1 000 |
| 18 | 0 | 0 | 0 | 0 | 420 | 520 | 620 |
| 19 | 0 | 0 | 0 | 0 | 0 | 340 | 380 |
| 20 | 0 | 0 | 0 | 0 | 0 | 300 | 280 |
| 21 | 0 | 0 | 0 | 0 | 0 | 160 | 280 |
| 22 | 0 | 0 | 0 | 0 | 0 | 100 | 60 |
| 23 | 0 | 0 | 0 | 0 | 0 | 0 | 80 |
| 24 | 0 | 0 | 0 | 0 | 0 | 0 | 80 |
| 25 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| U | 402 780 | 147 540 | 61 920 | 31 580 | 21 480 | 16 060 | 13 540 |
| TOTAL | 1 192 840 | 1 005 500 | 840 860 | 574 180 | 451 580 | 363 000 | 292 320 |
| U _i | 0,338 | 0,147 | 0,074 | 0,055 | 0,048 | 0,044 | 0,046 |
| Z _i | 0,501 | 0,198 | 0,071 | 0,040 | 0,032 | 0,030 | 0,033 |

TABLEAU 4.3 Estimations révisées des nombres de femmes de parité non déclarée et de femmes sans enfants par âge, Kenya, recensement de 1989

| Groupe d'âge (i) | Parité non déclarée révisée | Parité zéro révisée |
|------------------|-----------------------------|---------------------|
| 15-19 (1) | 32 746 | 967 594 |
| 20-24 (2) | 27 603 | 318 537 |
| 25-29 (3) | 23 084 | 98 236 |
| 30-34 (4) | 15 763 | 38 937 |
| 35-39 (5) | 12 397 | 23 663 |
| 40-44 (6) | 9 965 | 17 135 |
| 45-49 (7) | 8 025 | 15 075 |

TABLEAU 4.4 Parités moyennes corrigées par groupe d'âge, Kenya, recensement de 1989

| Groupe d'âge (i) | Parité moyenne |
|------------------|----------------|
| 15-19 (1) | 0,242 |
| 20-24 (2) | 1,525 |
| 25-29 (3) | 3,214 |
| 30-34 (4) | 4,760 |
| 35-39 (5) | 6,239 |
| 40-44 (6) | 7,120 |
| 45-49 (7) | 7,510 |

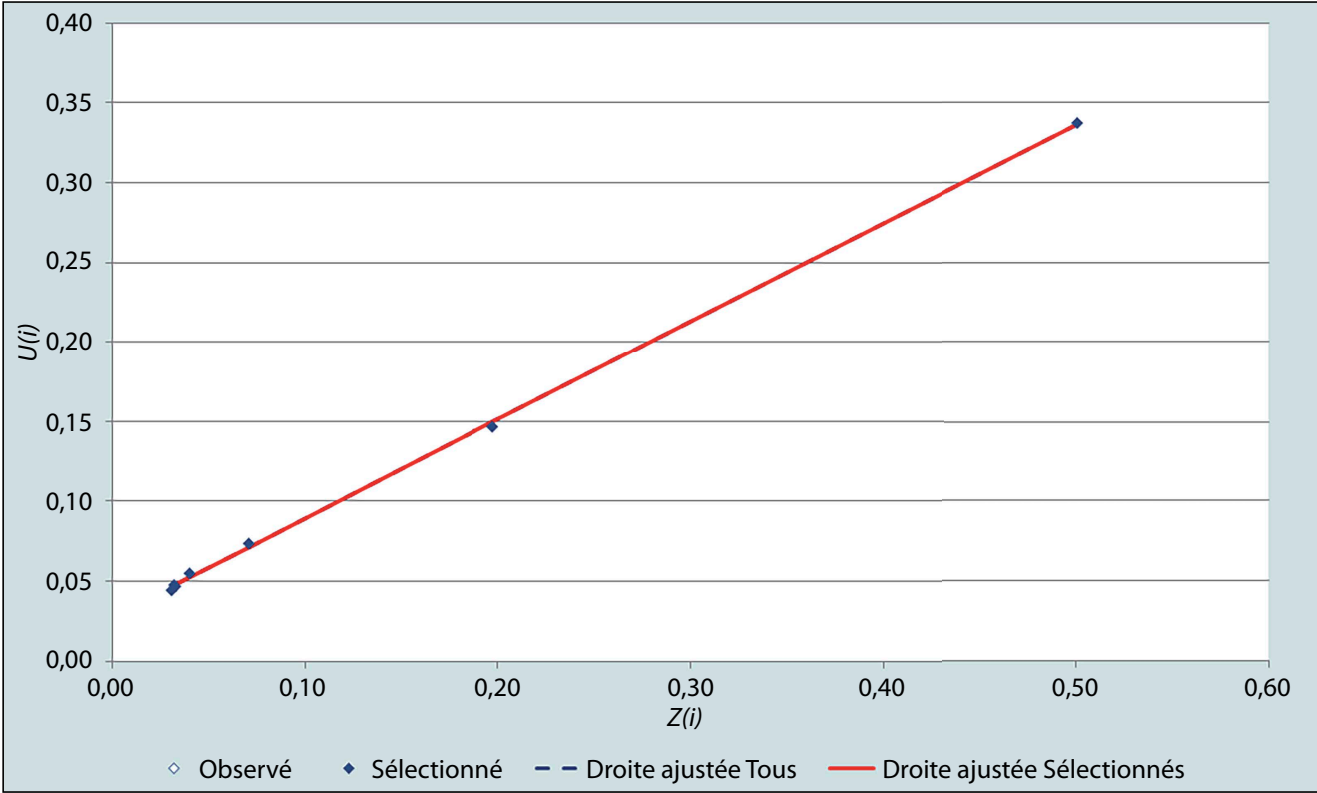
et en remplaçant dans l'équation 3,

$$U_i = \frac{\alpha}{1 - \alpha} Z_i + \beta = \gamma Z_i + \beta$$

où gamma peut être considéré comme la probabilité qu'une femme sans enfants soit classée comme étant de parité inconnue.

Donc une régression de Z_i sur U_i donnera des estimations de β (ainsi que de γ et α).

FIGURE 4.1 Ajustement de la correction d'el Badry, Kenya recensement de 1989



De l'équation 3, nous tirons alors :

$$U_i - \beta = \alpha Z_i^* = Z_i^* - Z_i$$

et donc $Z_i^* = N_{i,0}^* = U_i - \beta + Z_i$ et $U_i^* = \beta N_i$.

Notons que bien que nous ayons deux identités incluant Z_i elles vont donner une seule et même réponse si l'ajustement est exact. Par convention nous préférons utiliser l'équation 3 plutôt que l'équation 4 parce qu'elle repose sur les valeurs ajustées de β (la proportion estimée des parités vraiment non déclarées), plutôt que sur la valeur de α , qui est difficile à interpréter intuitivement.

Une fois obtenues les valeurs corrigées de Z_i^* et U_i^* les parités moyennes peuvent être calculées à partir de l'équation 2.

En appliquant la correction, on doit s'assurer que, dans chaque groupe d'âge, le nombre ajusté de femmes sans enfants (c'est-à-dire de parité zéro) est inférieur au nombre de femmes ne déclarant pas d'enfants au cours de la période de référence, en réponse à la question sur la fécondité récente. Le Z_i^* révisé peut donc être utilisé pour déterminer le nombre minimum de femmes qui ne pourraient pas avoir eu de naissance dans la période de référence précédant le recensement.

L'annexe II du *Manuel X (Division de la population des NU 1984)* donne une version de la correction d'El-Badry adaptée aux situations (désormais rares) où les questions sur les nombres d'enfants déjà nés ne sont posées qu'aux femmes mariées.

RÉFÉRENCES

Division de la Population des Nations Unies. 1984. *Manuel X. Techniques indirectes d'estimation démographique*. New York : Nations Unies, Département des affaires économiques et sociales internationales, ST/ESA/SER.A/81. http://unstats.un.org/unsd/demographic/standmeth/handbooks/Manuel_X-fr.pdf

El-Badry MA. 1961. "Failure of enumerators to make entries of zero: errors in recording childless cases in population censuses", *Journal of the American Statistical Association* **56**(296):909–924. doi: <http://dx.doi.org/10.1080/01621459.1961.10482134>

Chapitre 5 ÉVALUATION DE LA QUALITÉ DES DONNÉES SUR LA FÉCONDITÉ RÉCENTE TIRÉES DES RECENSEMENTS

Tom A Moultrie

INTRODUCTION

Avant d'évaluer la qualité des données sur la fécondité récente recueillies dans un recensement, il importe d'examiner le libellé précis des questions utilisées pour saisir l'information sur les naissances récentes en consultant le questionnaire. Des questions très différentes ont été utilisées au gré des pays et des vagues successives de recensements. La formulation peut affecter la validité des estimations et le sens et l'ampleur des biais ou des erreurs dans les données.

Les formes génériques des questions sur la fécondité récente dans les recensements appartiennent à trois grandes catégories :

- *Avez-vous donné naissance à un enfant au cours de la dernière année (ou une autre période de référence) ?*

Cette question produit une réponse binaire simple. Les naissances multiples au cours de la même période de référence ne sont pas saisies. Celles-ci peuvent résulter soit de la naissance de jumeaux ou de triplés lors d'une seule grossesse soit d'un très court intervalle séparant deux grossesses différentes. Aucune de ces éventualités n'est susceptible d'affecter substantiellement les taux fécondité car les intervalles entre naissances inférieurs à un an sont rares et la probabilité qu'une grossesse donne lieu à des naissances multiples est inférieure à 2 % dans la plupart des populations (à l'exception peut-être de l'Afrique sub-saharienne). Face à des données collectées sous cette forme, il est recommandé de faire l'hypothèse simplificatrice que toutes les naissances ont eu lieu au milieu de la période de référence, et que chaque grossesse ne donne lieu qu'à une naissance.

- *À combien d'enfants avez-vous donné naissance au cours de l'année écoulée (ou de toute autre période de référence) ?*

Cette question est plus raffinée que la précédente. Elle ne donne pas d'information sur le calendrier de la naissance au

cours de la période de référence, mais elle saisit l'information sur les naissances multiples qu'aurait eues une femme, sans distinguer toutefois entre les jumeaux et les intervalles courts entre naissances. Cette fois encore, il est raisonnable pour les calculs de supposer que les naissances se sont produites à la mi-temps de la période de référence.

- *Quelle a été la date de votre dernière naissance vivante ?*

Cette question cherche à déterminer la date du dernier accouchement avec un plus grand degré de précision, même si, en général, seuls le mois et l'année de la dernière naissance sont enregistrés. S'il y a en outre des questions sur le nombre de naissances qui se sont produites à cette date, l'ensemble donne une information plus précise sur le nombre de naissances récentes.

D'autres questions sont parfois posées (par exemple, sur la survie du dernier né, son sexe ou la date de naissance de l'enfant précédent). On peut en tirer des estimations directes sur les taux de mortalité infantile par sexe ou le rapport de masculinité à la naissance, par exemple.

L'évaluation de la qualité des données sur la fécondité récente peut inclure les contrôles suivants :

1. Comparaison du nombre total de naissances avec le nombre attendu (par exemple, avec les nombres tirés du système d'enregistrement à l'état civil ou de l'application d'une série précise de taux de fécondité par âge à la population recensée des femmes – bien que, dans ce cas, le sous-dénombrement systématique des femmes puisse aussi causer une sous-estimation des taux) ;
2. Évaluation de la vraisemblance de la distribution des taux de fécondité par âge tirés directement des données. Des distributions plausibles de la fécondité par âge sont presque invariablement uni modales, concaves, légèrement asymétriques à droite, et proches de zéro aux deux âges extrêmes de la période féconde. La distribution

doit également présenter une évolution raisonnablement régulière des taux de fécondité d'un âge à l'autre ;

3. Contrôle de la vraisemblance des nombres déclarés de naissances dans la période de référence. Dans certains recensements (par exemple l'Afrique du sud 1996), une proportion significative de répondantes confondent les questions sur le nombre total d'enfants déjà nés et celles sur la fécondité récente et donnent la même réponse aux deux questions. Cette erreur se manifeste par une forte diagonale dans les tableaux croisés des nombres d'enfants déjà nés par les nombres d'enfants nés dans l'année écoulée par âge de la mère (Moultrie and Timæus 2002) ;
4. Si les données ont été recueillies sur le sexe du dernier né, le rapport de masculinité à la naissance doit faire l'objet d'un contrôle. Le rapport de masculinité à la naissance est habituellement de l'ordre de 1,05, mais peut n'être que de 0,95 dans les populations africaines, et atteindre 1,1 dans certaines populations d'Asie. Des valeurs hors de l'intervalle 0,99-1,06 doivent être examinées soigneusement.

Dans tous les cas, il est indispensable d'identifier précisément l'univers des femmes à qui sont adressées les questions (en particulier les âges et l'état matrimonial des répondantes éligibles), ainsi que les règles d'enregistrement et de codage des non réponses et des données incorrectes.

Évaluation de la qualité des données sur la fécondité actuelle

Avant d'analyser les taux de fécondité par âge, il est recommandé d'évaluer dans quelle mesure les données sur les naissances récentes sont manquantes ou peu vraisemblables. L'absence de toute donnée manquante indique presque certainement que les données ont été corrigées. Si c'est le cas, d'autres recherches doivent porter sur l'ampleur des corrections et/ou des imputations dans la mesure du possible, par exemple, en examinant la distribution de valeurs imputées lorsque des étiquettes d'imputation ont été ajoutées aux données.

La proportion de données manquantes doit aussi être vérifiée. Si elle dépasse 5 % du nombre total des enregistrements permettant d'établir les données de fécondité actuelle, des recherches plus approfondies sont nécessaires. Il faut en particulier examiner la distribution par âge des cas manquants. Si ceux-ci sont concentrés aux jeunes âges ou chez les quadragénaires, on peut penser que les cas manquants correspondent à des femmes qui n'ont pas eu de naissance au cours de la période de référence, l'agent recenseur n'ayant pas enregistré de réponse au lieu

d'indiquer zéro. C'est une erreur de même type que celle conduisant à la correction d'El-Badry.

Quand les données donnent lieu à des tableaux par nombre de naissances dans la période de référence (plutôt que d'indiquer simplement si, oui ou non, une naissance a eu lieu au cours de la période de référence), on doit s'attacher aux nombres respectifs de naissances simples et multiples. En général, moins de 2 % des grossesses donnent naissance à plusieurs enfants. Les triplés et plus sont extrêmement rares (moins de 0,5 % des accouchements). Si la proportion de naissances multiples au cours de la période de référence semble trop élevée, il est recommandé d'établir des tableaux croisant le nombre total d'enfants déjà nés et les naissances au cours de l'année écoulée dans chaque groupe d'âge de femmes. Si le total d'enfants déjà nés et les naissances de l'année écoulée sont identiques dans une large proportion de cas, même pour des parités de deux ou plus, on peut penser que les personnes recensées ou les agents recenseurs n'ont pas compris la distinction entre les questions sur la descendance et sur la fécondité récente. Il se peut toutefois qu'une large proportion de jeunes femmes n'ayant qu'un enfant déjà né ait eu la naissance correspondante au cours de la période de référence, et une coïncidence fréquente entre des descendes d'un seul enfant et des déclarations d'une naissance récente par les jeunes femmes peut très bien ne pas indiquer des erreurs de déclaration.

MESURE DIRECTE DE LA FÉCONDITÉ À PARTIR DES DONNÉES DE RECENSEMENT

Quand la qualité des données est suffisante, on peut estimer directement les taux de fécondité par âge. Quand la qualité des données est médiocre, les taux de fécondité par âge obtenus par calcul direct sont incorporés dans diverses méthodes visant à produire des estimations plus fiables du niveau de fécondité par recours à des techniques indirectes.

La forme exacte des taux de fécondité par âge qu'on peut calculer dépend de la nature des données recueillies. Un taux de fécondité par âge à un âge (ou dans un groupe d'âge) donné est le rapport du nombre de naissances des femmes de cet âge (dans ce groupe d'âge) dans une période définie au nombre de femmes-années vécues par les femmes du même âge (dans le même groupe d'âge) dans la même période. Pour que le calcul soit exact, il faut connaître de façon fiable les dates de naissance des mères (pour déterminer l'âge de la mère) et de leurs enfants. On peut alors calculer précisément l'âge de la mère à la naissance de son enfant, et allouer son

exposition au risque aux âges ou groupes d'âge appropriés au long de la période couverte par la recherche.

Les données nécessaires pour ces calculs précis ne sont généralement pas disponibles dans les micro-données de recensement, soit parce que les dates complètes ne sont pas recueillies sur le terrain, soit parce qu'il y aurait un risque de violation de la confidentialité si les dates de naissance complètes étaient accessibles aux utilisateurs finaux des données. En outre, les données de recensement sont rarement de qualité suffisante pour justifier un tel surcroît de précision. L'attraction exercée par certains mois de naissance (par exemple janvier) ainsi que certaines années de naissance (par exemple celles finissant par 0 ou 5) pose fréquemment problème. Des périodes de dénombrement étendues peuvent créer des problèmes quand il s'agit de passer de la période de référence (par exemple les douze mois écoulés jusqu'à la date d'interview) à une période calendaire (par exemple 2008). En outre, les questions rétrospectives sur les naissances récentes posées dans un recensement excluent évidemment l'information relative aux naissances des mères qui sont décédées ou ont quitté le pays depuis leur accouchement.

Il y a quatre combinaisons possibles d'informations essentielles relatives à la mère et aux naissances récentes (Tableau 5.1).

Même dans le cas (4) du tableau 5.1, qui repose sur les informations les plus détaillées, le souci de calculer précisément l'exposition au risque ne donne pas forcément des résultats garantis, du fait que l'attraction exercée par certains mois lors de la déclaration des dates et autres problèmes de qualité des données peuvent sévèrement altérer les estimations. Le recours à de simples approximations pour le calcul des taux de fécondité à partir des données de recensement est généralement suffisant. Dans la section sur la mesure directe de la fécondité à partir des données d'enquête, nous décrivons

le calcul plus précis de l'exposition au risque et l'estimation des taux de fécondité à partir de données de bonne qualité.

Cas 1 et 2 : Estimation des taux de fécondité par âge directement à partir de données ne fournissant pas d'information sur la date de naissance de l'enfant

Dans les cas (1) et (2) du tableau 5.1, ce qui est connu sur la fécondité récente de la mère se limite au fait qu'elle a donné naissance (ou non) à au moins un enfant au cours de la période précédant le recensement. Dans des variantes plus informatives de la question sur la fécondité récente, la mère peut être interrogée sur le nombre de naissances vivantes qu'elle a eues dans la période précédant le recensement. Une telle question permet d'identifier à la fois les naissances multiples issues d'une même grossesse (jumeaux, triplés, etc.), et les cas où plus d'une grossesse arrive à son terme au cours de la période de référence.

Comme l'âge de la mère à l'accouchement n'est pas connu, l'approximation couramment utilisée consiste à classer les taux de fécondité selon l'âge déclaré par la mère à la date du recensement. On fait alors l'hypothèse supplémentaire que les naissances ont toutes eu lieu au milieu de l'intervalle en question. Ceci signifie qu'au moment du recensement les mères sont, en moyenne, plus âgées de la moitié de la durée de l'intervalle. Il en résulte que les âges auxquels les taux de fécondité se réfèrent réellement sont plus jeunes que les âges des femmes déclarés au recensement. La plupart des méthodes standards d'estimation indirecte de la fécondité corrigent les âges de ce décalage.

L'information supplémentaire (sur le mois et l'année de naissance de la mère) dont on dispose dans le deuxième cas ne permet guère d'affiner les estimations de la fécondité car on continue d'avoir besoin d'hypothèses supplémentaires d'uniformité de la distribution des dates de naissance des

TABLEAU 5.1 Taxonomie des données relatives à la mère et aux enfants pour l'estimation de la fécondité récente

| Information sur les enfants nés au cours de la période précédant le recensement (*) | Informations sur la mère | |
|---|---------------------------|--|
| | Âge révolu au recensement | Date de naissance (au moins mois, année) |
| Nombre d'enfants nés (ou simplement, oui/non) | (1) | (2) |
| Date de naissance du dernier né (au moins mois, année) | (3) | (4) |

(*) En général la période précédant le recensement est de 12 mois, mais l'analyste doit être attentif à la possibilité de périodes non-standard, fondée par exemple sur le temps écoulé depuis un événement national important.

enfants. Donc, dans le cas (2) comme dans le cas (1) du tableau 5.1, les taux de fécondité sont estimés en divisant le nombre de naissances déclarées au cours de la période de référence (classées par âge de la mère à la date du recensement) par le nombre de femmes de cet âge. Le nombre total de naissances au cours de la période de référence déclaré par les femmes âgées x à la date du recensement, B_x , est donné par

$$B_x = \sum_{k=0}^{\omega} k \cdot N_{x,k}$$

où k est le nombre déclaré de naissances au cours de la période de référence, ω est la valeur maximale de k dans les données et $N_{x,k}$ est le nombre de femmes âgées de x au recensement déclarant avoir eu k naissances au cours de la dernière année. Si ω est exprimé comme un intervalle ouvert, par exemple 3 naissances ou plus au cours de la période de référence, les femmes dans cette catégorie sont supposées avoir eu le nombre de naissances qui ouvre cet intervalle (soit 3 naissances ici). L'erreur qui en résulte est de nouveau faible.

Le nombre de femmes âgées de x est donné par

$$N_x = \sum_{k=0}^{\omega} N_{x,k}$$

Les femmes dont les naissances récentes ne sont pas connues ou n'ont pas été déclarées doivent être exclues à la fois du numérateur et du dénominateur, avec l'hypothèse implicite que leur fécondité ne diffère pas de celle des femmes dont la fécondité récente est connue. Les taux de fécondité par âge x sont égaux à

$$f_x = B_x / N_x$$

Dans l'intervalle d'âge conventionnel (de 15 ans à 49 ans, inclus), l'indice synthétique de fécondité (TF) calculé à partir des taux par année d'âge est égal à

$$TF = \sum_{a=15}^{49} f_a$$

L'indice synthétique de fécondité désigne le nombre d'enfants qu'une femme aurait si elle survivait jusqu'à 50 ans (considéré comme la fin de la période féconde) en ayant, au long de sa vie reproductive, les taux de fécondité par âge observés au cours de la période précédant immédiatement le recensement.

Il convient de calculer et de représenter graphiquement les taux de fécondité par année d'âge afin de vérifier la cohérence interne des données. Les taux de fécondité par

âge sont en général moins erratiques que les numérateurs ou les dénominateurs eux-mêmes et ils peuvent indiquer des niveaux et des distributions vraisemblables de la fécondité. Une série de taux par âge fortement erratique, s'écartant fortement de la forme prévisible en n , laisse fortement supposer que les données de fécondité récente posent problème et invite à pousser les recherches plus avant.

Enfin on peut calculer des taux de fécondité par groupes d'âges quinquennaux, ${}_5f_x$, où $x = 15, 20, \dots, 45$:

$$f_i = {}_5f_x = \frac{\sum_{x=5i+10}^{5i+14} B_x}{\sum_{x=5i+10}^{5i+14} N_x}$$

où l'indice i est égal à $i = (x/5) - 2$. La mesure de l'indice synthétique de fécondité est donc

$$TF = 5 \cdot \sum_{i=1}^7 f_i$$

Alors que l'indice synthétique est une mesure de la fécondité standardisée par l'âge (la distribution par âge de la population féminine d'âge fécond étant supposée implicitement uniforme par groupe d'âge), le taux de fécondité de chaque groupe d'âge n'est pas standardisé au sein du groupe. Il peut en résulter un léger écart entre les indices synthétiques résultant respectivement d'un calcul par groupe d'âge et par année d'âge, mais cet écart ne porte généralement que sur la deuxième ou la troisième décimale.

L'indice synthétique de fécondité doit être comparé aux estimations tirées d'autres sources de données pour le même pays (par exemple des EDS). Il faut toutefois se souvenir que les taux de fécondité par âge et les indices synthétiques obtenus par cette méthode ne prennent pas en compte la vraie exposition au risque dans la détermination du dénominateur. En outre, le numérateur inclut des événements ayant eu lieu pendant la période de référence classés selon l'âge de la mère à la fin de celle-ci, et non par son âge au moment où l'événement a eu lieu. La plupart des méthodes d'estimation indirecte de la fécondité corrigent les taux de fécondité calculés pour tenir compte de ce décalage. Dans des comparaisons de base visant à évaluer les forme et le niveau des distributions de fécondité, les différences dans les classements par âge ont peu d'importance. La variante *F-seulement* du modèle relationnel de Gompertz offre toutefois une méthode pour ne pas décaler les taux de fécondité lors de leur lissage, si on le souhaite.

EXEMPLE :**CALCUL DIRECT DE LA FÉCONDITÉ**

Au recensement de 2008 au Cambodge, les femmes ont été interrogées sur le nombre d'enfants qu'elles ont eu l'année

précédente. L'âge des mères est l'âge à la date du recensement. Les données figurent au tableau 5.2.

La colonne « Manquant » indique que seulement 1 600 femmes sur près de 3,65 millions âgées entre 15 et 49 ans n'ont

*TABLEAU 5.2 Fécondité récente par âge de la mère
à la date du recensement, Cambodge, recensement de 2008*

| Naissances au cours de l'année précédente | | | | | | | | | |
|---|------------------|----------------|--------------|------------|------------|--------------|----------------|------------------|---------------|
| Âge | 0 | 1 | 2 | 3 | 4 | Manquant | Naissances | Femmes | Taux* |
| 15 | 160 980 | 120 | 0 | 0 | 0 | 80 | 120 | 161 180 | 0,0007 |
| 16 | 152 710 | 500 | 0 | 0 | 0 | 50 | 500 | 153 260 | 0,0033 |
| 17 | 144 970 | 1 250 | 10 | 10 | 0 | 20 | 1 300 | 146 260 | 0,0089 |
| 18 | 182 500 | 3 540 | 20 | 0 | 0 | 40 | 3 580 | 186 100 | 0,0192 |
| 19 | 127 840 | 5 640 | 10 | 0 | 0 | 30 | 5 660 | 133 520 | 0,0424 |
| 20 | 147 990 | 8 840 | 80 | 0 | 0 | 90 | 9 000 | 157 000 | 0,0574 |
| 21 | 123 960 | 9 500 | 30 | 0 | 0 | 70 | 9 560 | 133 560 | 0,0716 |
| 22 | 126 030 | 11 600 | 80 | 0 | 0 | 30 | 11 760 | 137 740 | 0,0854 |
| 23 | 123 750 | 11 830 | 70 | 10 | 0 | 110 | 12 000 | 135 770 | 0,0885 |
| 24 | 121 820 | 11 010 | 150 | 10 | 20 | 80 | 11 420 | 133 090 | 0,0859 |
| 25 | 137 460 | 12 420 | 100 | 0 | 0 | 60 | 12 620 | 150 040 | 0,0841 |
| 26 | 115 370 | 11 320 | 110 | 0 | 0 | 80 | 11 540 | 126 880 | 0,0910 |
| 27 | 117 840 | 11 580 | 190 | 0 | 0 | 40 | 11 960 | 129 650 | 0,0923 |
| 28 | 118 270 | 10 690 | 110 | 0 | 10 | 30 | 10 950 | 129 110 | 0,0848 |
| 29 | 82 990 | 7 600 | 120 | 0 | 0 | 40 | 7 840 | 90 750 | 0,0864 |
| 30 | 77 690 | 5 950 | 40 | 10 | 0 | 30 | 6 060 | 83 720 | 0,0724 |
| 31 | 58 800 | 4 820 | 50 | 20 | 0 | 30 | 4 980 | 63 720 | 0,0782 |
| 32 | 67 110 | 4 480 | 150 | 20 | 0 | 110 | 4 840 | 71 870 | 0,0674 |
| 33 | 67 080 | 4 240 | 40 | 0 | 0 | 50 | 4 320 | 71 410 | 0,0605 |
| 34 | 67 010 | 3 800 | 30 | 10 | 10 | 70 | 3 930 | 70 930 | 0,0555 |
| 35 | 90 720 | 4 570 | 60 | 20 | 0 | 30 | 4 750 | 95 400 | 0,0498 |
| 36 | 77 950 | 3 800 | 10 | 10 | 0 | 30 | 3 850 | 81 800 | 0,0471 |
| 37 | 81 320 | 4 070 | 50 | 10 | 10 | 10 | 4 240 | 85 470 | 0,0496 |
| 38 | 92 290 | 3 780 | 30 | 20 | 30 | 30 | 4 020 | 96 180 | 0,0418 |
| 39 | 74 030 | 2 920 | 50 | 0 | 0 | 30 | 3 020 | 77 030 | 0,0392 |
| 40 | 88 940 | 2 720 | 70 | 10 | 10 | 50 | 2 930 | 91 800 | 0,0319 |
| 41 | 71 250 | 2 140 | 0 | 0 | 0 | 20 | 2 140 | 73 410 | 0,0292 |
| 42 | 81 560 | 2 010 | 30 | 0 | 0 | 60 | 2 070 | 83 660 | 0,0248 |
| 43 | 72 930 | 1 270 | 10 | 0 | 0 | 30 | 1 290 | 74 240 | 0,0174 |
| 44 | 69 660 | 930 | 10 | 0 | 0 | 50 | 950 | 70 650 | 0,0135 |
| 45 | 84 290 | 760 | 30 | 10 | 10 | 30 | 890 | 85 130 | 0,0105 |
| 46 | 67 330 | 510 | 0 | 50 | 30 | 40 | 780 | 67 960 | 0,0115 |
| 47 | 66 220 | 270 | 10 | 0 | 10 | 0 | 330 | 66 510 | 0,0050 |
| 48 | 74 790 | 310 | 10 | 10 | 0 | 30 | 360 | 75 150 | 0,0048 |
| 49 | 57 600 | 120 | 0 | 20 | 10 | 20 | 220 | 57 770 | 0,0038 |
| TOTAL | 3 473 050 | 170 910 | 1 760 | 250 | 150 | 1 600 | 175 780 | 3 647 720 | 1,6157 |

* Taux de fécondité par âge

pas eu leur fécondité récente enregistrée. Ceci représente 0,04 % de l'ensemble des femmes et n'aura pas d'impact sur la fécondité estimée des femmes au Cambodge. Un contrôle supplémentaire portant sur la distribution par âge de ces cas ne fait pas apparaître de schéma particulier d'omission. Le nombre de naissances est égal à la somme pondérée des femmes déclarant 1, 2, 3 ou 4 accouchements, sur la dernière ligne. 173 070 femmes ($170\,910 + 1\,760 + 250 + 150$) ont eu un total de 175 780 naissances ($1 \times 170\,910 + 2 \times 1\,760 + 3 \times 250 + 4 \times 150$) au cours de l'année précédant le recensement. Parmi ces femmes, 98,8 % ($170\,910/173\,070$) ont eu une seule naissance, 1,0 % ont eu des jumeaux et 0,2 % des triplés ou plus. L'éventualité de quintuplés (ou de cinq naissances en deux accouchements au cours de la période) est si faible qu'elle peut être ignorée. Si le recensement n'avait pas comptabilisé les naissances multiples séparément, le taux brut de natalité aurait été sous-estimé par un facteur de $173\,070/175\,780 = 0,984$. La sous-estimation est donc de 1,6 %.

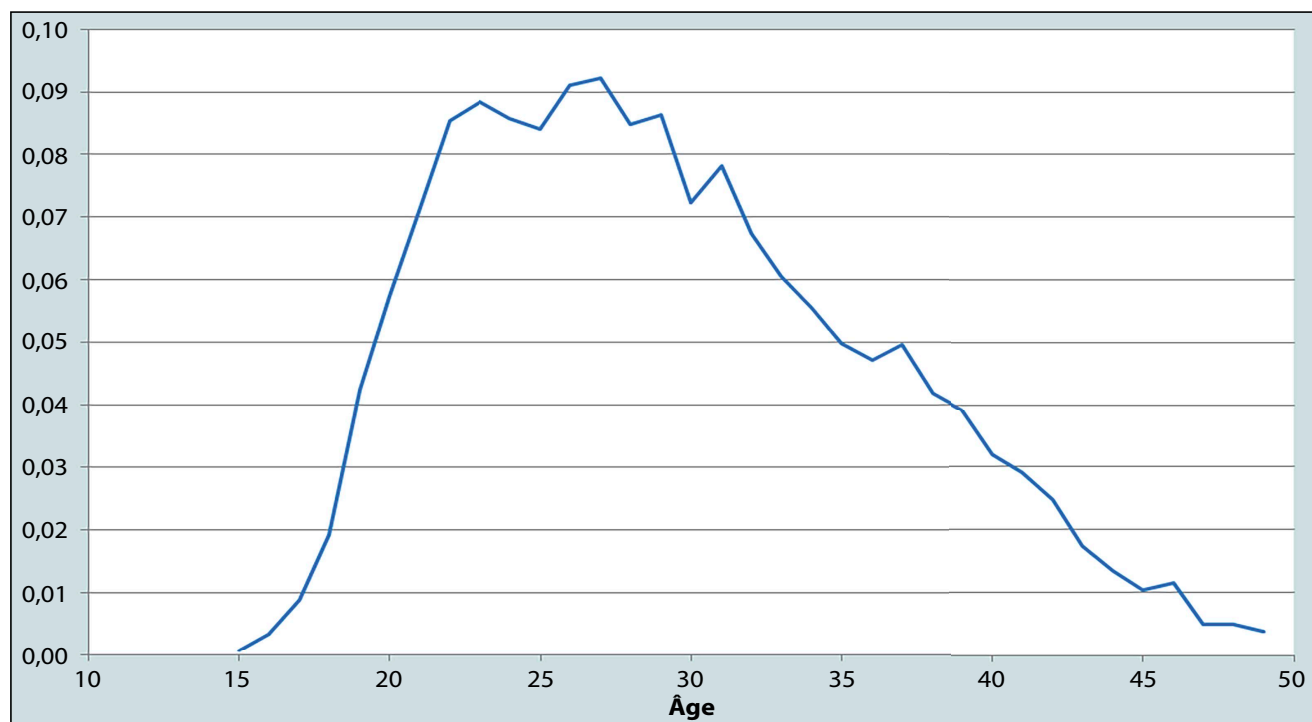
À partir des données ci-dessus, la série des taux de fécondité par année d'âge est obtenue en divisant le nombre total de naissances issues des femmes de chaque âge par le nombre de femmes ayant déclaré leur fécondité récente, c'est-à-dire en excluant les femmes qui n'ont pas déclaré

combien de naissances elles ont eu au cours de l'année passée. Les taux sont représentés sur la figure 5.1. Alors que le nombre de femmes dénombrées à chaque âge est erratique, les taux de fécondité par année d'âge sont relativement lisses, avec un schéma de fécondité clairement défini et un pic caractéristique aux environs de 25 ans.

Selon ces données, l'indice synthétique de fécondité est de 1,61 enfants par femme. Le calcul conduit par groupes d'âges quinquennaux donne le même résultat (tableau 5.3), même si les mesures diffèrent à la troisième décimale, comme on l'a évoqué plus haut.

Même en l'absence de contrôle externe, les résultats tirés des données du recensement de 2008 font apparaître des niveaux de fécondité invraisemblablement bas au Cambodge. Les données sont également incohérentes avec les parités moyennes calculées plus haut à la section sur l'évaluation de la qualité des données sur les descendance. Ceci suggère que les données sur la fécondité récente recueillies dans ce recensement sont fortement déficientes. Des contrôles externes le confirment, sous forme d'estimations de fécondité tirées de deux EDS réalisées avant et après le recensement. Selon les données des deux dernières colonnes du tableau 3, l'indice synthétique de fécondité estimé à partir de l'EDS 2010

FIGURE 5.1 Taux de fécondité par âge, Cambodge recensement de 2008



(fondé sur les naissances des trois ans précédant l'enquête) est de 3,1 enfants par femme. L'estimation à partir de l'EDS de 2005 était de 3,4 enfants par femme. Seulement la moitié environ des naissances survenues dans l'année précédant le recensement ont été déclarées aux agents recenseurs.

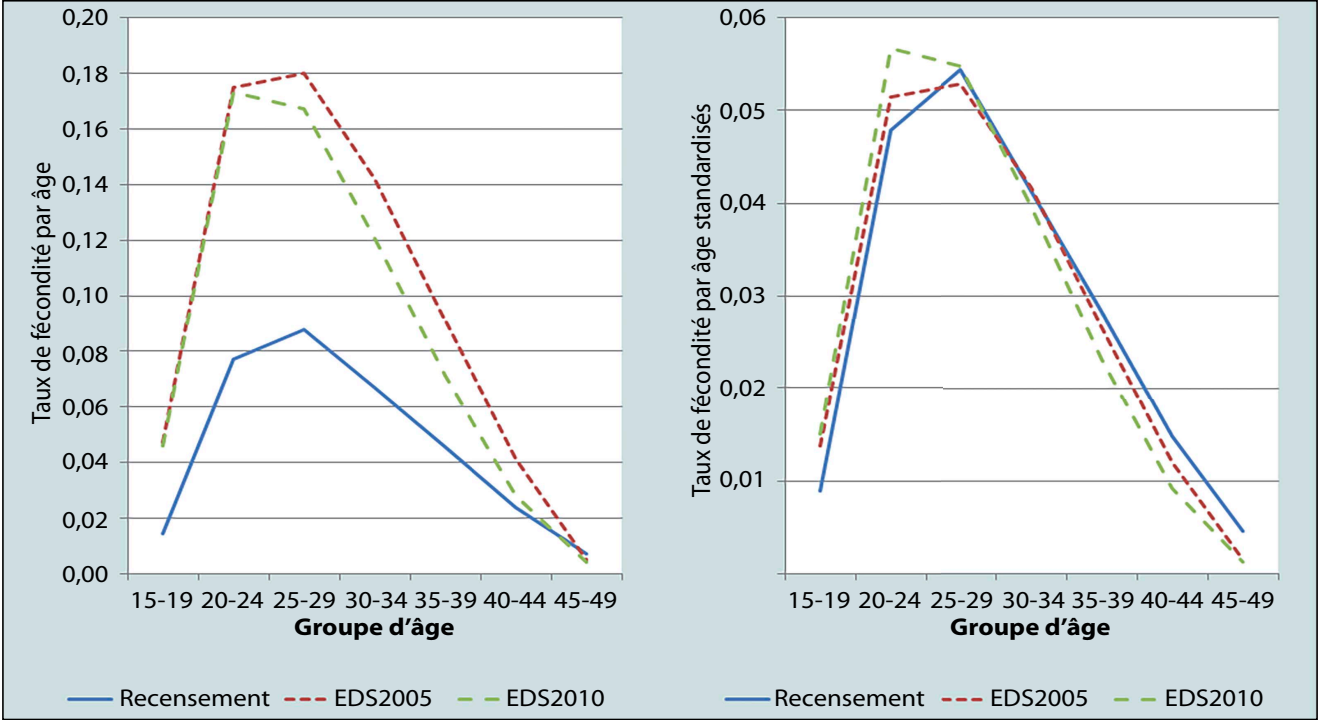
Le graphique de gauche sur la figure 5.2 représente les taux de fécondité par âge calculés à partir du recensement de 2008 et des deux EDS. Les taux tirés du recensement diffèrent clairement de ceux tirés des EDS. Ces derniers, de leur côté, présentent un schéma plutôt étrange d'évolution

TABLEAU 5.3 Taux de fécondité par groupes d'âge quinquennaux, Cambodge, recensement de 2008 et Enquêtes Démographiques et de Santé 2005 et 2010

| Groupe d'âge | Femmes | Manquant | Naissances | Taux | EDS2005 | EDS2010 |
|--------------|---------|----------|------------|-------|---------|---------|
| 15-19 | 780 320 | 220 | 11 160 | 0,014 | 0,047 | 0,046 |
| 20-24 | 697 160 | 380 | 53 740 | 0,077 | 0,175 | 0,173 |
| 25-29 | 626 430 | 250 | 54 910 | 0,088 | 0,180 | 0,167 |
| 30-34 | 361 650 | 290 | 24 130 | 0,067 | 0,142 | 0,121 |
| 35-39 | 435 880 | 130 | 19 880 | 0,046 | 0,091 | 0,071 |
| 40-44 | 393 760 | 210 | 9 380 | 0,024 | 0,041 | 0,028 |
| 45-49 | 352 520 | 120 | 2 580 | 0,007 | 0,005 | 0,004 |
| TF | | | | 1,61 | 3,41 | 3,05 |

Source : Recensement, nos calculs ; EDS, StatCompiler (www.statcompiler.com)

FIGURE 5.2 Taux de fécondité par âge et taux de fécondité par âge standardisés, Cambodge, recensement de 2008, EDS 2005 et EDS 2010



de la fécondité en cinq ans, marquée par des réductions à peu près constante entre 25 et 44 ans. Le graphique de droite de la figure 5.2 représente les mêmes taux mais ramenés à un même indice synthétique de fécondité de un enfant par femme. Malgré des différences fortes dans le niveau global de la fécondité, la forme des trois distributions par âge est très semblable, la seule réelle différence entre elles étant à 20-24 ans. C'est pourquoi il semble peu vraisemblable qu'il y ait d'importantes différences dans la qualité de la déclaration de la fécondité récente selon l'âge des femmes dans le recensement de 2008 au Cambodge.

Même si le niveau de la fécondité tiré des données du recensement de 2008 est sérieusement biaisé, la forme de la distribution par âge de la fécondité est raisonnablement précise. Ceci est une condition préalable à l'application de nombreuses méthodes indirectes d'estimation de la fécondité.

Cas 3 et 4 : Estimation des taux de fécondité par âge quand la date de naissance de l'enfant est connue

Si les données sont classées par date de naissance du dernier né des femmes, il convient de choisir une période appropriée pour les analyses sur la fécondité. En général, il est conseillé de ne pas prendre des périodes beaucoup plus longues qu'un an car des périodes plus longues accroissent la probabilité que les femmes aient eu plus d'une grossesse au cours de cette période. Les naissances ayant eu lieu plus anciennement sont donc omises (la condition requise étant de déclarer la date de naissance du dernier né, pas de tous les enfants nés au cours de la période), ce qui signifie que les estimations de la fécondité excluront systématiquement des naissances dans le passé le plus ancien. En outre, si la fécondité a évolué rapidement, allonger la période de recherche au-delà d'un an signifie que les estimations représenteront une sorte de moyenne de la fécondité au long de la période. Si le recensement a eu lieu assez tôt ou assez tard dans l'année, il peut être intéressant de calculer les taux sur les naissances depuis le début de l'année précédente ou de l'année courante car ce choix ne nécessite pas que les femmes se souviennent précisément du mois de naissance de leur enfant. Le nombre de naissances déclaré au cours de la période de référence peut alors être ramené à une estimation des naissances annuelles par un calcul au prorata. On peut calculer les taux à la fois de cette façon et en s'appuyant sur une période de référence de 12 mois puis comparer les résultats.

Le scénario (3) au tableau 5.1 ne permet pas d'obtenir une mesure tout à fait précise de la fécondité puisque l'âge de la mère à la naissance de l'enfant ne peut pas être établi précisément. La connaissance de la date de naissance de l'enfant permet bien, toutefois, de déterminer plus soigneusement le dénominateur des taux de fécondité par âge.

Dans la situation courante où la question posée porte sur le mois et l'année de naissance du dernier enfant, une démarche plus soignée peut permettre de déterminer le nombre de naissances au cours de l'année passée. En général la date du recensement est fixée conventionnellement. Les questions dans le questionnaire de recensement se réfèrent généralement à un jour particulier, même si le processus effectif de dénombrement prend plusieurs semaines. Une liste des dates de recensement pour les trois dernières vagues de recensement a été établie par les Nations Unies : <http://unstats.un.org/unsd/demographic/sources/census/censusdates.htm> ; une liste des dates de recensement pour les données diffusées par IPUMS est disponible à <https://international.ipums.org/international/samples.shtml>.

Lors de la détermination du numérateur, toutes les naissances déclarées dans le mois du recensement, et une proportion au prorata des naissances déclarées s'être produites le mois équivalent un an plus tôt doivent être incluses. Pour tirer cette information des données de recensement, en fonction de la capacité de traitement des dates des logiciels statistiques utilisés, on peut recourir à la procédure de numérotation des mois à partir du début du siècle des EDS.

Lors du recensement de 1999 au Kenya, la date de référence officielle était la nuit du 24 au 25 août 1999. Pour estimer les naissances survenues dans l'année précédant le recensement, toutes les naissances déclarées entre septembre 1998 et août 1999 doivent être incluses, ainsi que 1-24/31 (= 7/31) des naissances déclarées en août 1998. Ceci suppose que les naissances sont uniformément réparties au long du mois (Tableau 5.4). Par exemple, le nombre estimé de naissances au cours de l'année précédant le recensement dans le groupe 30-34 ans est donné par

$$\frac{7}{31}(13\ 940) + 9\ 560 + 9\ 600 + \dots + 9\ 280 = 146\ 807,74$$

En l'absence d'information supplémentaire sur la date de naissance de la mère, les données ci-dessus sont classées selon le groupe d'âge de la mère à la date du recensement. Comme

TABLEAU 5.4 Naissances déclarées chaque mois par âge de la mère à la date de recensement (24-25 août 1999), Kenya, recensement de 1999

| Mois | Âge de la mère au recensement | | | | | | |
|------------------------------------|-------------------------------|------------|------------|------------|-----------|-----------|----------|
| | 15-19 | 20-24 | 25-29 | 30-34 | 35-39 | 40-44 | 45-49 |
| Août 1998 | 13 240 | 31 300 | 23 120 | 13 940 | 8 940 | 3 220 | 560 |
| Septembre 1998 | 9 800 | 22 900 | 17 260 | 9 560 | 6 180 | 2 080 | 680 |
| Octobre 1998 | 9 240 | 21 580 | 15 520 | 9 600 | 5 880 | 1 880 | 500 |
| Novembre 1998 | 9 040 | 21 940 | 16 060 | 9 880 | 5 280 | 1 660 | 540 |
| Décembre 1998 | 10 200 | 23 700 | 18 000 | 10 580 | 5 940 | 2 080 | 480 |
| Janvier 1999 | 14 720 | 28 620 | 20 620 | 12 260 | 7 300 | 2 180 | 660 |
| Février 1999 | 20 740 | 42 140 | 30 860 | 17 400 | 11 220 | 4 560 | 2 060 |
| Mars 1999 | 15 620 | 31 480 | 21 320 | 12 520 | 7 340 | 2 820 | 520 |
| Avril 1999 | 18 660 | 33 160 | 24 260 | 12 240 | 7 820 | 2 860 | 720 |
| Mai 1999 | 19 660 | 33 880 | 22 860 | 13 960 | 7 440 | 2 480 | 760 |
| Juin 1999 | 20 100 | 32 140 | 23 380 | 12 580 | 7 300 | 2 720 | 560 |
| Juillet 1999 | 21 600 | 32 360 | 23 860 | 13 800 | 7 060 | 2 640 | 520 |
| Août 1999 | 15 900 | 25 020 | 16 720 | 9 280 | 5 840 | 1 620 | 360 |
| Naissances estimées l'année passée | 188 269,68 | 355 987,74 | 255 940,65 | 146 807,74 | 86 618,71 | 30 307,10 | 8 486,45 |
| Nombre de femmes | 1 700 060 | 1 495 180 | 1 205 060 | 849 620 | 725 780 | 519 740 | 417 500 |
| Taux de fécondité par âge | 0,1107 | 0,2381 | 0,2124 | 0,1728 | 0,1193 | 0,0583 | 0,0203 |

on l'a dit précédemment, les taux ainsi calculés doivent faire l'objet d'un décalage d'une demi-année.

Les taux de fécondité par âge sont obtenus en divisant ces naissances par le nombre de femmes dans chaque groupe d'âge. Les estimations de l'indice synthétique de fécondité qui en résultent sont sensiblement éloignées des autres estimations de la fécondité dans le pays vers cette époque. Comme au Cambodge, ceci donne à penser que les naissances au cours de l'année précédant le recensement sont fortement sous-déclarées.

On ne peut prendre une mesure précise de la fécondité que dans le cas (4), quand on dispose d'une information détaillée sur la date de naissance aussi bien de la mère que de l'enfant. Mais si certaines dates de naissance (par exemple

le 1^{er} janvier) exercent une forte attraction, le recours à des mesures affinées ne sert à rien car elles seront biaisées par l'attraction. Comme il y a peu de chances que la qualité et la cohérence interne des données recueillies lors d'un recensement soient aussi bonnes que dans une EDS, il est inutile d'appliquer le même calcul précis que celui qu'on utilise dans une EDS. Il y a toutefois des situations où l'attraction lors de la déclaration des dates de naissance et les autres erreurs dans les données jouent un rôle suffisamment limité pour qu'il vaille la peine de faire un calcul direct des estimations de fécondité. Dans ces cas, on doit appliquer les principes décrits pour le calcul des estimations de la fécondité récente à partir des données d'enquête.

RÉFÉRENCES

Moultrie, Tom A. and Ian M. Timæus. 2002. *Trends in South African Fertility between 1970 and 1998: An Analysis of the 1996 Census and the 1998 Demographic and Health Survey*. Cape Town: Medical Research Council. <http://www.mrc.ac.za/bod/trends.pdf>. Accessed 1 May 2011.

Méthodes fondées
sur un seul recensement

Chapter 6 VUE D'ENSEMBLE SUR LES MÉTHODES D'ESTIMATION DE LA FÉCONDITÉ FONDÉES SUR LE QUOTIENT P/F

Tom A Moultrie

INTRODUCTION

Presque toutes les méthodes d'estimation indirecte de la fécondité ont leur origine dans la méthode du quotient P/F proposée initialement par Brass (1964). En outre, l'interprétation des résultats d'autres méthodes (par exemple les taux de fécondité par cohorte et période) et certains des outils diagnostiques utilisés pour évaluer la qualité des données lors de l'estimation de la mortalité infantile reposent aussi sur la logique intrinsèque de la démarche du quotient P/F . Même si la méthode dans sa forme originelle et dans ses formes modifiées a été aujourd'hui remplacée par le modèle relationnel de Gompertz et ses variantes, il reste utile d'en présenter ici la logique essentielle. Les lecteurs intéressés peuvent se reporter au *Manuel X* (Division de la Population des NU 1984) pour un exposé complet de la démarche.

LA MÉTHODE DU QUOTIENT P/F DE BRASS

La méthode repose sur le fait que les mesures de fécondité par cohorte et par période sont identiques lorsque la fécondité reste durablement constante. En d'autres termes, dans ces circonstances, la fécondité cumulée d'une cohorte de femmes jusqu'à un âge donné est égale à la fécondité cumulée jusqu'au même âge au cours de n'importe quelle période.

Si nous supposons qu'il n'y a pas de différences de mortalité importantes en fonction de la fécondité des mères, de sorte que les femmes survivantes n'ont pas une descendance substantiellement différente de celle des femmes décédées, la fécondité cumulée d'une cohorte de femmes jusqu'à un âge donné est égale à la parité moyenne dans cette cohorte. (Cette hypothèse n'est pas très importante, car même s'il y a des différences dans la fécondité des femmes vivantes et décédées, dans la plupart des populations la mortalité aux

âges reproductifs est très faible et l'effet de la survie différentielle est, en conséquence, très réduit.)

Brass définit P comme la parité moyenne (la fécondité cumulée au long de la vie) d'une cohorte de femmes jusqu'à un âge donné, alors que F est étroitement lié à la fécondité actuelle (du moment) cumulée jusqu'au même âge. La méthode du quotient P/F exprime ces deux quantités en relation l'une à l'autre sous la forme d'un rapport pour chaque groupe d'âge.

Il est un peu plus compliqué d'obtenir F qu'on ne l'a suggéré plus haut pour deux raisons. D'abord, toute comparaison entre fécondité par cohorte et par période doit tenir compte du probable décalage des données sur la fécondité actuelle résultant du fait que la question s'appuie sur l'âge de la mère au moment de l'interview plutôt que sur celui qu'elle avait lors de sa naissance la plus récente. Ensuite, alors que la fécondité d'une période cumulée jusqu'à un âge donné reflète la fécondité de l'ensemble des femmes jusqu'à cet âge, les parités moyennes sont généralement celles des femmes d'un groupe quinquennal d'âge et reflètent donc à peu près la parité moyenne des femmes à l'âge médian du groupe quinquennal. La méthode formulée par Brass prend en compte ces deux éléments.

Si la fécondité est constante depuis longtemps dans une population, et si les données ne comportent pas d'erreurs, le quotient P/F doit être égal à 1 dans chaque groupe d'âge. Mais si la fécondité recule, la fécondité cumulée au long de la vie doit être supérieure à la fécondité actuelle cumulée. Dans ce cas (en l'absence d'erreurs dans les données), le quotient P/F doit s'écarter systématiquement de l'unité à mesure que s'élève l'âge des mères.

Le corollaire de cette observation est que le quotient P/F devrait être assez proche de l'unité aux jeunes âges,

car même vers 25 ans la fécondité cumulée par période ne devrait pas s'écarter sensiblement de celle cumulée au sein d'une cohorte, la plupart des naissances des femmes de cette cohorte ayant eu lieu récemment. D'où on déduit que le quotient P/F calculé pour les femmes âgées de 20-24 ans au moment d'une enquête est sans doute l'indicateur le plus fiable de la qualité des données recueillies sur la fécondité. On peut aussi penser raisonnablement que les parités moyennes des jeunes femmes sont généralement bien déclarées, au moins par rapport à celles des femmes plus âgées.

Ce schéma caractéristique d'écart par rapport à l'unité lorsque l'âge des femmes augmente est à la base de nombreuses analyses sur la nature et la qualité des données tirées des questions sur la fécondité récente et sur la descendance.

DIAGNOSTICS FONDÉS SUR LE QUOTIENT P/F

En réalité, les données ne sont jamais exemptes d'erreurs et celles-ci rendent confus et obscur le schéma hypothétique d'écart à l'unité du quotient P/F .

Comme on l'a vu dans les sections sur l'évaluation de la qualité des données sur la fécondité récente (chapitre 3) et sur l'évaluation de la qualité des données sur la parité (chapitre 4), deux erreurs affectent souvent ces données. La première est que les déclarations de descendance – c'est-à-dire de fécondité cumulée par cohorte – deviennent de plus en plus inexactes à mesure qu'augmente l'âge des répondantes, les femmes âgées tendant à sous-déclarer leur descendance. Les erreurs de ce type tendent donc à faire baisser le numérateur du quotient P/F , en particulier aux âges avancés. Lorsque de telles erreurs affectent les données, le quotient se rapproche de l'unité davantage qu'il ne ferait dans la réalité.

Le second type d'erreur fréquemment rencontré tient à ce que les femmes tendent à sous-déclarer leurs naissances récentes, quel que soit leur âge. Il en résulte un niveau de fécondité récente inférieur à ce qu'on attendrait et un quotient P/F surestimé.

RÉFÉRENCES

Brass W. 1964. Utilisation des données des recensements ou des enquêtes pour l'estimation des taux de natalité et de mortalité. Préparé pour le Cycle d'études sur les statistiques d'état civil, Addis Abéba 14-19 décembre 1964. Document No. E/CN.14/CAS.4/V57. New York : Nations Unies. <http://repository.uneca.org/bitstream/handle/10855/9306/Bib-49868.pdf?sequence=1>

La méthode du quotient P/F s'efforce de corriger le second défaut en appliquant le quotient P/F calculé pour les jeunes femmes (pour les raisons exposées ci-dessus) au schéma de fécondité observé directement comme un facteur de proportionnalité.

RÉSUMÉ DES MÉTHODES FONDÉES SUR LE QUOTIENT P/F

Un certain nombre des méthodes décrites ici ont été originellement présentées dans le *Manuel X* comme des extensions de la méthode du quotient P/F . Le modèle relationnel de Gompertz peut être considéré comme une version améliorée et plus souple de la méthode du quotient P/F de Brass. Le modèle s'appuie sur les mêmes données (et fait les mêmes hypothèses sur les erreurs affectant les données de fécondité) que son précurseur. Mais il est important de noter que la méthode ne nécessite pas d'hypothèse sur la constance passée de la fécondité. La comparaison des descendances et de la fécondité par période reste néanmoins au cœur de la méthode.

La plupart des extensions de la méthode du quotient P/F de Brass présentées dans le *Manuel X* ont été remaniées en extensions du modèle relationnel de Gompertz. Ces extensions incluent les méthodes qui recourent aux données sur les accroissements de parité entre deux recensements pour estimer la fécondité ; les méthodes qui utilisent conjointement les accroissements de parité et un schéma-type de taux de fécondité intercensitaires (le modèle relationnel de Gompertz synthétique ; et les méthodes indirectes qui s'appuient sur des données du système d'enregistrement à l'état civil. Les taux de fécondité par cohorte et période tirés de données d'enquête reposent aussi sur la logique de la méthode du quotient P/F pour éclairer les tendances et la dynamique à long terme de la fécondité.

Division de la Population des Nations Unies. 1984. *Manuel X. Techniques indirectes d'estimation démographique*. New York : Nations Unies, Département des affaires économiques et sociales internationales, ST/ESA/SER.A/81. http://unstats.un.org/unsd/demographic/standmeth/handbooks/Manuel_X-fr.pdf

Chapitre 7 LE MODÈLE RELATIONNEL DE GOMPERTZ

Tom A Moultrie

DESCRIPTION DE LA MÉTHODE

La méthode relationnelle de Gompertz est un perfectionnement de la méthode du quotient P/F de Brass qui cherche à estimer les taux de fécondité par âge et l'indice synthétique de fécondité en déterminant la forme de la courbe de fécondité à partir des données sur les naissances récentes recueillies dans des recensements ou des enquêtes, son niveau étant fixé d'après les parités moyennes déclarées par les jeunes femmes.

En faisant ces estimations, la méthode cherche à corriger les erreurs couramment rencontrées dans les données de fécondité, lorsque trop ou trop peu de naissances sont déclarées au cours de la période de référence et que les descendance sont sous-déclarées et qu'il y a des erreurs de déclaration des âges chez les femmes âgées. Ces erreurs sont décrites en détail dans la section sur l'évaluation de la qualité des données de fécondité.

La méthode repose sur une propriété utile d'une distribution (cumulée) de Gompertz, $G(x) = \exp(a \cdot \exp(bx))$, qui est sigmoïdale (c'est-à-dire en forme de S), mais qui a aussi une fonction associée de risque, asymétrique à droite et qui saisit donc assez bien à la fois la distribution des parités moyennes par âge des femmes et leur fécondité cumulée. La forme de $G(x)$ implique qu'une double transformation logarithmique négative des fécondités cumulées proportionnelles ou des parités moyennes s'approche d'une ligne droite sur l'essentiel de l'intervalle des âges. La double transformation logarithmique, $Y(x) = -\ln(-\ln(G(x)))$ est intitulée *gompit* et présente une étroite analogie avec la transformation *logit* fréquemment utilisée dans l'analyse de la mortalité. Mais Brass a montré qu'on pouvait obtenir un ajustement beaucoup plus proche de la linéarité en utilisant un modèle relationnel qui exprime les gompits d'une série observée de données de fécondité comme une fonction linéaire des gompits d'une distribution de fécondité standard définie. En d'autres termes, $Y(x) = \alpha + \beta Y^s(x)$, où $Y^s(x)$ est le *gompit* d'une distribution standard de fécondité. Évidemment, si $\alpha = 0$ and

$\beta = 1$, la distribution de fécondité sera identique au schéma standard. Alpha (α) indique la mesure dans laquelle l'âge à la maternité dans la population diffère de l'âge dans le standard (des valeurs négatives impliquent des âges plus élevés que dans le standard), alors que bêta (β) est une mesure de la dispersion de la fécondité (des valeurs supérieures à 1 correspondent à une distribution plus resserrée).

La méthode nécessite de connaître les parités moyennes dans chaque groupe d'âge, ${}_5P_x$ pour $x = 15, 20, \dots, 45$, et les taux de fécondité dans chaque groupe d'âge, ${}_5f_x$. Pour simplifier l'exposé, et pour distinguer clairement les données de descendance et de fécondité récente, ${}_5P_{15}$ est intitulé $P(1)$, ${}_5P_{20}$ $P(2)$ et ainsi de suite. La section du manuel traitant de l'évaluation de la qualité des données de fécondité explique comment ces éléments sont tirés des données de recensement. Comme avec d'autres méthodes, les parités moyennes doivent être ajustées pour la correction d'El-Badry si nécessaire.

La fécondité du moment cumulée jusqu'au terme de chaque groupe d'âge est égale à $F(x+5) = 5 \cdot \sum_{a=15,5}^x {}_5f_a$. La méthode originellement proposée par Brass (1978) utilisait la série des gompits du rapport entre fécondité cumulée au terme de chaque groupe d'âge et taux de fécondité cumulés jusqu'à 50 ans (c'est-à-dire indice synthétique de fécondité, TF), courbe de forme sigmoïdale avec un minimum de 0 et un maximum de 1 (au dernier groupe d'âge). Les gompits des parités moyennes sont obtenus de la même façon.

Il y a deux faiblesses inhérentes à cette démarche. Premièrement, elle nécessite la connaissance de l'indice synthétique de fécondité et les estimations de l'indice synthétique de fécondité tirées des taux de fécondité par âge peuvent être biaisées. En fait, l'indice synthétique de fécondité est souvent le paramètre que l'analyste cherche en priorité à estimer. La seconde faiblesse est l'hypothèse implicite de fécondité constante au fil du temps qui résulte du traitement des gompits de parité. Néanmoins, la formulation de Brass

est à l'origine du développement par Booth (1980, 1984) de la distribution de fécondité standard, qui est utilisé dans le modèle encore aujourd'hui.

La reformulation de la méthode par Zaba (1981) traite largement le problème de ces deux limites, en évitant la circularité de la méthode initiale et en abandonnant l'hypothèse de fécondité constante. Un travail ultérieur non publié de Zaba a généralisé la démarche pour incorporer des variantes du modèle (dont certaines sont décrites ici). Nous faisons plus loin un exposé complet de la reformulation de Zaba. Mais en résumé elle montre que le modèle peut s'exprimer comme

$$z(x) - e(x) = \alpha + \beta g(x) + \frac{c}{2}(\beta - 1)^2, \quad (1)$$

où $e(x)$, $g(x)$ et c sont des fonctions du standard retenu et $z(x)$ est le gompit des rapports entre deux mesures successives de la fécondité cumulée par période, c'est-à-dire $F(x)/F(x+5)$, au lieu de $F(x)/F50$ comme Brass le suggérerait à l'origine.

En d'autres termes,

$$z(x) = -\ln \left(-\ln \left(\frac{F(x)}{F(x+5)} \right) \right).$$

Pour les données de parité, le modèle est ajusté sur les rapports de parités moyennes successives, $P(i)/P(i+1)$. Ceci signifie que le modèle peut être utilisé sans qu'il soit besoin d'estimer l'indice synthétique de fécondité avant d'ajuster les paramètres de forme. Il découle de l'équation 1 qu'une représentation graphique de $z(x) - e(x)$ en fonction de $g(x)$ doit être une droite de pente β et d'ordonnée à l'origine

$$\alpha + \frac{1}{2}c(\beta - 1)^2 \quad (\text{En notant que } \beta \text{ devrait être proche de}$$

l'unité, les premières formulations de la procédure tendaient à négliger le second terme de l'ordonnée à l'origine, réduisant celle-ci à α . Avec les capacités de calcul aujourd'hui disponibles, cette perte de précision dans le calcul de l'ordonnée à l'origine n'a plus lieu d'être. La nécessité que β soit proche de l'unité demeure néanmoins).

Le même raisonnement s'applique à l'évaluation des données de parité. En utilisant $P(i)/P(i+1)$, le rapport des parités moyennes dans deux groupes d'âge consécutifs, l'équation linéaire rattachant $z(i) - e(i)$ à $g(i)$ donne

$$z(i) - e(i) = \alpha + \beta g(i) + \frac{c}{2}(\beta - 1)^2 \quad (2)$$

Par convention, on appelle P -points les points tirés des données de parité et F -points ceux tirés des taux de fécondité. La procédure d'ajustement a pour but de trouver une combinaison cohérente des P - et des F -points, les deux séries de points définissant à peu près les mêmes droites, et de déterminer ensuite grâce à ceux-ci conjointement les paramètres α et β des équations 1 et 2 ci-dessus. Les valeurs de α et β permettent d'établir les gompits relationnels, $Y(x) = \alpha + \beta Y^s(x)$, et de même pour $Y(i)$.

L'obtention d'une distribution de fécondité ajustée par la méthode relationnelle de Gompertz nécessite des tableaux de parités moyennes et de taux de fécondité par âge. Les taux de fécondité sont cumulés et les rapports des valeurs cumulées successives sont calculés. Les rapports des parités moyennes successives sont également calculés. Les gompits de ces rapports sont calculés et utilisés pour établir le graphique des deux paires de points, $z(x) - e(x)$ en fonction de $g(x)$, et $z(i) - e(i)$ en fonction de $g(i)$. Les droites d'ajustement auront des pentes égales à β , et une ordonnée à l'origine dont l'expression inclut α , β and c , d'où on pourra tirer α . Les valeurs de α et β sont utilisées pour transformer les gompits des cumulants standards en gompits ajustés, qui sont ensuite convertis en parités moyennes et taux de fécondité ajustés. Le niveau de la fécondité est déterminé par les points de parité les plus fiables. Ce sont généralement ceux des femmes âgées de 20-29 ou 20-34 ans, qui sont les moins sujettes à omettre des naissances et à déclarer des âges erronés comme peuvent l'être les femmes plus âgées.

Le recours au modèle relationnel de Gompertz pour calculer une distribution ajustée de la fécondité présente un certain nombre d'avantages par rapport à la méthode du quotient P/F utilisée antérieurement. Le modèle utilise une distribution de fécondité fiable pour des régimes de fécondité moyenne ou élevée (le standard de Booth). Ainsi, les taux de fécondité douteux estimés à partir des déclarations de naissances de l'année écoulée peuvent être remplacés par des valeurs-types ajustées en utilisant les points les plus fiables. La représentation graphique des deux séries de points est un guide précieux pour établir la fiabilité de chaque point ; elle peut suggérer quelles sont les données erronées et aider à identifier les tendances de la fécondité. Tous les points fiables peuvent servir à établir la distribution modèle ajustée. Le modèle fournit aussi un moyen fiable d'interpolation utile lors de l'établissement de données de parité ou de fécondité cumulée comparables

et de conversion des taux de fécondité dans des groupes d'âge non conventionnels en taux dans des groupes d'âge conventionnels.

DONNÉES NÉCESSAIRES ET HYPOTHÈSES

Tableaux des données nécessaires

- Taux de fécondité pour les 12, 24 ou 36 mois précédant l'enquête, classés par âge de la mère à l'enquête ou âge à la naissance de l'enfant, ou
 - nombre de femmes à la date du recensement ou de l'enquête par groupe quinquennal, et
 - nombre de naissances des femmes dans les 12, 24 ou 36 mois précédant l'enquête par groupe quinquennal d'âge.
- Parités moyennes des femmes, classées par groupe quinquennal d'âge de la mère, ou
 - nombre de femmes par groupe quinquennal et
 - nombre total d'enfants nés de ces femmes, par groupe quinquennal d'âge.

Hypothèses importantes

- Le schéma de fécondité standard choisi pour la procédure d'ajustement reflète de façon appropriée la forme de la distribution de la fécondité dans la population.
- Les changements de la fécondité ont été réguliers et graduels et ont affecté tous les groupes d'âge dans une mesure comparable.
- Les erreurs dans les taux de fécondité avant ajustement sont proportionnellement les mêmes parmi les femmes dans les groupes d'âge centraux (20-39 ans), de sorte que la forme de la fécondité par âge des naissances récentes déclarées est raisonnablement correcte.
- Les parités déclarées par les jeunes femmes (20-29 ou 20-34 ans) sont conformes à la réalité.

La méthode permet généralement de détecter les violations de ces hypothèses.

TRAVAIL PRÉPARATOIRE ET RECHERCHES PRÉLIMINAIRES

Avant d'entreprendre l'analyse des niveaux de fécondité en recourant à cette méthode, l'analyste doit vérifier la qualité des données au moins sur les points suivants :

- structure par âge et sexe de la population ;
- naissances déclarées au cours de l'année écoulée, et
- parités moyennes et nécessité d'une correction d'El-Badry.

AVERTISSEMENTS ET MISES EN GARDE

- Lors de l'application de cette méthode, l'analyste doit faire particulièrement attention à vérifier et à spécifier correctement la définition utilisée pour classer les mères par âge.
- Chaque fois que c'est nécessaire, les parités moyennes doivent être corrigées en utilisant la correction d'El-Badry, suite au classement erroné des femmes sans enfants dans la catégorie des parités non déclarées.
- La méthode peut être utilisée avec des données agrégées sur une période de trois ans. Une vigilance particulière est toutefois nécessaire dans l'utilisation du modèle complet (par opposition à son utilisation pour un simple lissage) avec des données largement supérieures à un an. Dans l'idéal, les nombres de personnes-années exposées au risque devraient alors être calculés. En outre, il existe une possibilité accrue de naissances multiples au cours d'une période prolongée et la formulation de la question dans le recensement ou l'enquête risque d'être mal adaptée à l'identification de tels cas.
- Si des pondérations liées au plan d'échantillonnage ont été fournies en même temps que les données, elles doivent être appliquées de manière adaptée au logiciel statistique lors de la préparation des tableaux nécessaires au modèle.
- La méthode est contre indiquée lorsque la forme de la distribution de fécondité à modéliser diffère largement de celle qui sous tend le standard de fécondité. Les paramètres α et β définissant la forme et la position de la distribution de fécondité, Zaba (1981) recommande de n'utiliser le modèle que si $-0,3 < \alpha < 0,3$ et $0,8 < \beta < 1,25$. Un autre standard devrait être choisi si α et β se situent hors de ces intervalles.
- Certaines des approximations utilisées pour établir les équations d'estimation fonctionnent moins bien dans les groupes d'âge les plus jeunes et les plus avancés que chez les femmes d'âge moyen, en particulier lorsque le schéma de fécondité déclarée diffère radicalement du standard. Les points correspondant aux déclarations des femmes de ces âges doivent donc être traités avec des précautions particulières. L'impact sur les estimations de l'indice synthétique de fécondité est toutefois limité.

APPLICATION DE LA MÉTHODE

L'application de la méthode se fait selon les étapes suivantes.

Étape 1 : Calcul des parités moyennes déclarées

Calculer les parités moyennes, ${}_5P_x$ des femmes de chaque groupe d'âge ($x, x + 5$), pour $x = 15, 20, \dots, 45$, si ça n'a pas déjà été fait lors des recherches préliminaires, ou produit comme conséquence de l'application de la correction d'El-Badry. Le calcul et la correction des parités moyennes sont décrits dans la section sur l'évaluation de la qualité des données de parité.

Étape 2 : Détermination du classement de l'âge des mères

Selon les données disponibles, les taux de fécondité peuvent être classés soit d'après l'âge des mères à la date d'enquête, soit par leur âge à la naissance de leur enfant. Le premier cas est presque systématique avec les données de recensement, où l'âge des mères est celui qu'elles avaient à la date du recensement. Le second cas est plus courant avec les données administratives tirées des systèmes d'enregistrement à l'état civil. Il est essentiel que ce classement soit connu car une définition incorrecte biaise les estimations de taux.

La feuille de calcul associée (voir le site web) au modèle peut traiter aussi bien les données sans décalage (c'est-à-dire déclarées selon l'âge des mères à la naissance) que, – dans le cas de données classées selon l'âge des mères à la date d'enquête – avec une demi-année, une année ou une année et demie de décalage (pour des périodes respectives de 12, 24 et 36 mois avant l'enquête).

Étape 3 : Calcul des taux de fécondité par âge

Les taux de fécondité par âge sont obtenus en divisant les naissances déclarées au cours de la période choisie (par exemple l'année, deux ans ou trois ans) avant la date d'enquête par le nombre de femmes dans chaque groupe d'âge.

Étape 4 : Choix du standard de fécondité à appliquer et de la variante du modèle à ajuster

Par défaut, le standard de fécondité est celui de Booth, légèrement modifié par Zaba (1981). Il est approprié dans le cas des populations à fécondité moyenne et élevée ; il est exprimé sous forme d'un schéma normalisé de fécondité cumulée (c'est-à-dire avec une fécondité totale égale à un). Les valeurs standards de $Y^s(x)$ sont déterminées en prenant les gompits de la distribution et les valeurs des parités standards, $Y^s(i)$, sont les gompits des parités associées au schéma standard de fécondité. Le choix du standard détermine les valeurs de $g()$ et $e()$ utilisées dans les procédures d'ajustement par régression qui sont tirées algébriquement de $Y^s()$.

Deux variantes du modèle relationnel de Gompertz sont présentées ici. L'option par défaut consiste à faire les mêmes hypothèses sur la nature des erreurs inhérentes aux données de fécondité que dans la méthode P/F de Brass, à savoir que les déclarations de fécondité récente souffrent d'erreurs sur la période de référence et de sous-déclaration qui sont indépendantes de l'âge, alors que les déclarations de descendance souffrent d'omissions qui s'accroissent avec l'âge. Sur la feuille de calcul, cette variante est intitulée « Forme F – Niveau P ».

Dans une seconde variante, on utilise le modèle relationnel de Gompertz pour corriger des éventuelles distorsions dans la forme de la distribution de fécondité, alors qu'on laisse le niveau inchangé. Si on suspecte des erreurs sur la période de référence ou une sous-déclaration, cette variante ne donnera pas une estimation plausible de la fécondité.

Étape 5 : Évaluer la représentation graphique des P -points et des F -points

Les représentations graphiques de $z(x) - e(x)$ en fonction de $g(x)$, et $z(i) - e(i)$ en fonction de $g(i)$ sur le même système d'axes permettent de faire apparaître des erreurs communes et des tendances dans les données (voir plus loin).

Étape 6 : Ajustement du modèle par sélection des points à utiliser

Au départ, tous les points doivent être inclus dans le modèle, sauf si la parité moyenne dans un groupe d'âge est supérieure à celle dans le groupe d'âge suivant. Dans ce cas, le gompit n'est pas défini et le modèle ne peut pas être ajusté en utilisant ce point (une telle situation ne peut pas survenir dans une cohorte réelle, mais elle peut le faire à cause d'erreurs dans les données ou dans une cohorte synthétique dans une période où la fécondité évolue rapidement).

Si les données de parité et de fécondité sont mutuellement cohérentes, les représentations graphiques des $z() - e()$ en fonction de $g()$ doivent se situer sur une droite. Les P -points et les F -points qui s'écartent d'une ligne droite doivent être exclus du modèle. Une régression par les moindres carrés ordinaires est utilisée pour ajuster des droites sur les P -points et les F -points et pour identifier, séquentiellement, les points qui ne se situent pas nettement sur une droite. Il s'agit de chercher la combinaison la plus large de P - et F -points (à peu près) alignés, et d'utiliser ceux-ci pour ajuster le modèle.

On retient ou on exclut des points en respectant les instructions suivantes :

- Une série de points contigus doit être incluse dans le modèle. Seuls des points extrêmes peuvent être exclus. (La raison est que chaque point sur le graphique est le résultat d'un calcul de ratio entre les valeurs de données adjacentes. Si l'analyse conduit à considérer une donnée comme non fiable en tant que dénominateur, il n'est pas logique de l'accepter comme numérateur du ratio suivant.)
- Il vaut mieux éliminer des *P*-points que des *F*-points, car les données de parité moyenne sont davantage sujettes à des erreurs liées à l'âge que les données de fécondité.
- Les *P*-points qui s'écartent de l'alignement construit seulement sur les autres *P*-points, ainsi que les *F*-points qui s'écartent de l'alignement construit seulement sur les autres *F*-points doivent être éliminés dès le début du processus d'ajustement.
- Il vaut mieux éliminer les *P*- et les *F*-points aux âges avancés plutôt qu'aux jeunes âges car les données *y* sont généralement moins fiables et présentent moins de cohérence entre descendance et fécondité récente. Les données relatives aux femmes de moins de 20 ans font cependant exception, car elles reposent sur des nombres faibles d'événements, ce qui rend souvent peu fiables les estimations des parités moyennes ou de la fécondité cumulée.
- Un ajustement réalisé sur un nombre élevé de points est préférable à un ajustement réalisé sur un faible nombre de points, même s'il est légèrement moins bon. La feuille de calcul produit l'écart quadratique moyen

$$RMSE = \sqrt{\frac{\sum \left((z(i) - e(i)) - \left(\alpha + (\beta - 1)^2 \frac{c}{2} + \beta g(i) \right) \right)^2}{n}}$$

à partir des points utilisés pour ajuster le modèle. Cette statistique peut aider à déterminer le nombre optimal de points sur lesquels réaliser l'ajustement, s'il y a un doute dans le choix à faire entre deux modèles concurrents. Dans ce cas, on peut choisir celui qui a le plus faible écart quadratique.

Étape 7 : Évaluer la qualité des paramètres estimés

Les valeurs de α et β , qui représentent la droite de meilleur ajustement joignant les *P*-points et les *F*-points restants, nécessitent un contrôle pour vérifier qu'elles ne s'éloignent pas de leurs valeurs centrales à un point tel que le choix du standard serait mis en cause. Un bon ajustement est obtenu si $-0,3 < \alpha < 0,3$, et si $0,8 < \beta < 1,25$.

Si les paramètres se situent en dehors de ces intervalles, au moins une des séries de données sous-jacentes pose problème ou le standard est inadéquat. Il convient alors d'essayer un autre standard (voir plus loin) ou de modifier la sélection des points avant de procéder plus avant. Si les paramètres sont toujours en dehors des intervalles ci-dessus, il faut alors considérer que la méthode n'est pas appropriée.

Étape 8 : Taux de fécondité par âge et indice synthétique de fécondité ajustés

Une fois estimés les deux paramètres du modèle, ceux-ci peuvent être appliqués au standard des parités pour fournir des valeurs ajustées $Y(i) = \alpha + \beta.Y^s(i)$. Celles-ci sont ensuite ramenées à des mesures de la proportion cumulée de fécondité réalisée par groupe d'âge i , grâce à la transformation anti-gompit. Les anti-gompits reposant sur les distributions de parité indiquent la proportion de fécondité réalisée dans le groupe d'âge. En divisant la parité observée dans chaque groupe d'âge par ces proportions, on obtient une série d'estimations de l'indice synthétique de fécondité. La moyenne de ces valeurs calculée sur le sous-ensemble des groupes d'âge qui ont été utilisés pour estimer α et β donne l'estimation ajustée de l'indice synthétique de fécondité, \hat{T} .

En appliquant les mêmes α et β aux gompits standards pour les âges qui bornent les groupes d'âge conventionnels (c'est-à-dire 20, 25 ... 50), en appliquant la transformation anti-gompit, et en multipliant par \hat{T} , on obtient une distribution ajustée de la fécondité cumulée. En faisant la différence entre les estimations successives de la fécondité cumulée et en la divisant par cinq, on obtient un schéma de fécondité ajusté pour les groupes d'âge conventionnels (15-19 ; 20-24 etc.), même si le classement des données initiales était décalé d'une demi-année.

(Si le modèle a été ajusté en utilisant seulement les *F*-points, α et β ne sont définis que par la *F*-droite. La distribution de fécondité lissée est obtenue par une série d'étapes semblable à celle décrite précédemment, sauf que les proportions ajustées sont multipliées par le niveau de fécondité estimé à partir des données récentes elles-mêmes, plutôt que par une estimation fondée sur les données de parité.)

INTERPRÉTATION ET DIAGNOSTICS

Erreurs classiques dans les données

Les points tirés des données relatives aux femmes de moins de 20 ans sont souvent peu fiables, car les événements sont

généralement peu nombreux et sujets à diverses erreurs de déclaration (par exemple, les agents recenseurs attribuent un âge plus élevé aux mères adolescentes). Il est donc courant que les droites ajustées aux P - et F - points conviennent mieux aux femmes en pleine période féconde (20-34 ans) qu'aux très jeunes et aux plus âgées. Si les P -droites et les F -droites ne convergent pas, même dans l'intervalle 20-34 ans, il est probable qu'il y a des erreurs dans au moins une des séries de données, même à ces âges, ou que des variations (importantes) ont affecté la fécondité récemment.

Une représentation graphique de l'ensemble des P -points et des F -points donne des informations sur les erreurs dans les données et sur les tendances récentes de la fécondité. En interprétant les graphiques, on se souviendra que les valeurs de $z() - e()$ (sur l'axe des ordonnées) varient avec les distributions observées de fécondité et de parité, mais pas les valeurs de $g()$ (qui ne s'appuient que sur le standard). De même, $z() - e()$ varie dans le même sens que les ratios qui le sous-tendent.

Les types de problèmes les plus courants mis en lumière par la représentation graphique sont l'omission d'enfants dans les déclarations de parité des femmes âgées, la surestimation des âges et une indication de baisse récente de la fécondité.

Zaba (1981) a utilisé des données simulées sur la base du standard de Booth pour étudier l'effet des erreurs de données et des variations de la fécondité sur les représentations graphiques. Les résultats sont présentés ci-dessous.

1. Les femmes âgées omettent des enfants en déclarant leur descendance

Si les femmes âgées omettent des enfants en déclarant leur descendance finale, les valeurs de P seront trop élevées (le dénominateur de chaque cumulant étant trop faible) par rapport à la droite attendue et les P -points auront tendance à s'incurver vers le haut aux âges élevés.

2. Sur-déclaration des naissances ou exagération des âges par les femmes âgées

Ces deux types d'erreur ont le même effet, qu'un nombre erroné de naissances soit déclaré par les femmes âgées ou que des femmes jeunes (qui ont une fécondité récente élevée) soient considérées comme plus âgées qu'elles ne le sont dans la réalité. Il en résulte une courbe F qui s'incurve vers le bas aux âges avancés.

3. Tendances de la fécondité

La divergence entre les P -points et les F -points sur le graphique peut révéler une évolution du niveau de la fécondité. Si la fécondité baisse, les cumulants de F sont plus élevés que les cumulants de P au même âge, et les F -points ont une pente plus forte que les P -points. Un diagnostic de baisse de la fécondité est possible quand les F -points s'alignent au-dessus de la droite des P -points et vice versa.

Des variations rapides de la fécondité qui affectent les jeunes âges à la maternité empêchent généralement les P -points et les F -points de s'aligner sur une même ligne, même après avoir exclu la plupart des P -points de l'ajustement. L'élimination successive des P -points qui ne s'alignent pas sur les P -points et les F -points suggère un changement rapide et récent de la fécondité aux jeunes âges.

Des graphiques diagnostiques, reposant sur le standard de Booth, sont représentés sur la figure 7.1.

Comme on le voit, si les femmes âgées omettent des naissances vivantes, les P -points à droite sur l'échelle (ceux des âges avancés) dérivent vers le haut. Quand les femmes se déclarent (ou sont déclarées) plus âgées qu'elles ne le sont, la ligne F s'incurve vers le bas aux âges avancés. Enfin, si la fécondité baisse, les F -points sont en général au dessus des P -points.

En traitant de données réelles, on est souvent confronté à un mélange d'erreurs et de tendances qui peut être beaucoup plus compliqué que les archétypes présentés ici. Des erreurs importantes peuvent masquer des tendances réelles et la méthode ne doit donc pas être appliquée aveuglément.

Quotients P/F

La méthode du quotient P/F n'est pas présentée dans ce manuel, mais les ratios qui découlent de l'application de la méthode donnent des indications utiles sur les tendances récentes de la fécondité. Ils peuvent aussi servir de guides lorsqu'il s'agit d'évaluer l'applicabilité de certaines méthodes d'estimation de la mortalité infantile.

On peut tirer facilement d'un modèle relationnel de Gompertz ajusté des pseudo-quotients pour chaque groupe d'âge

$$\frac{P}{F}(x) = \frac{{}_5P_x}{\hat{F} \cdot \exp\left(-\exp\left(\alpha_F + \beta_F Y^s(x + 2.5)\right)\right)},$$

$$x = 20, 25, \dots, 45$$

Le numérateur est la parité moyenne observée dans chaque groupe d'âge, alors que le dénominateur utilise les valeurs de α et β tirées des seuls F -points (à peu près comme dans la variante F -seulement du modèle) pour modifier le gompit standard au point médian de chaque groupe d'âge. L'anti-gompit est ensuite amené proportionnellement au niveau de l'indice synthétique de fécondité qu'impliquent les F -points retenus dans le modèle. On ne calcule pas le quotient aux plus jeunes âges, où les parités moyennes sont généralement très basses, comme l'est la fécondité cumulée à 17,5 ans, rendant ainsi le quotient instable dans ce groupe d'âge.

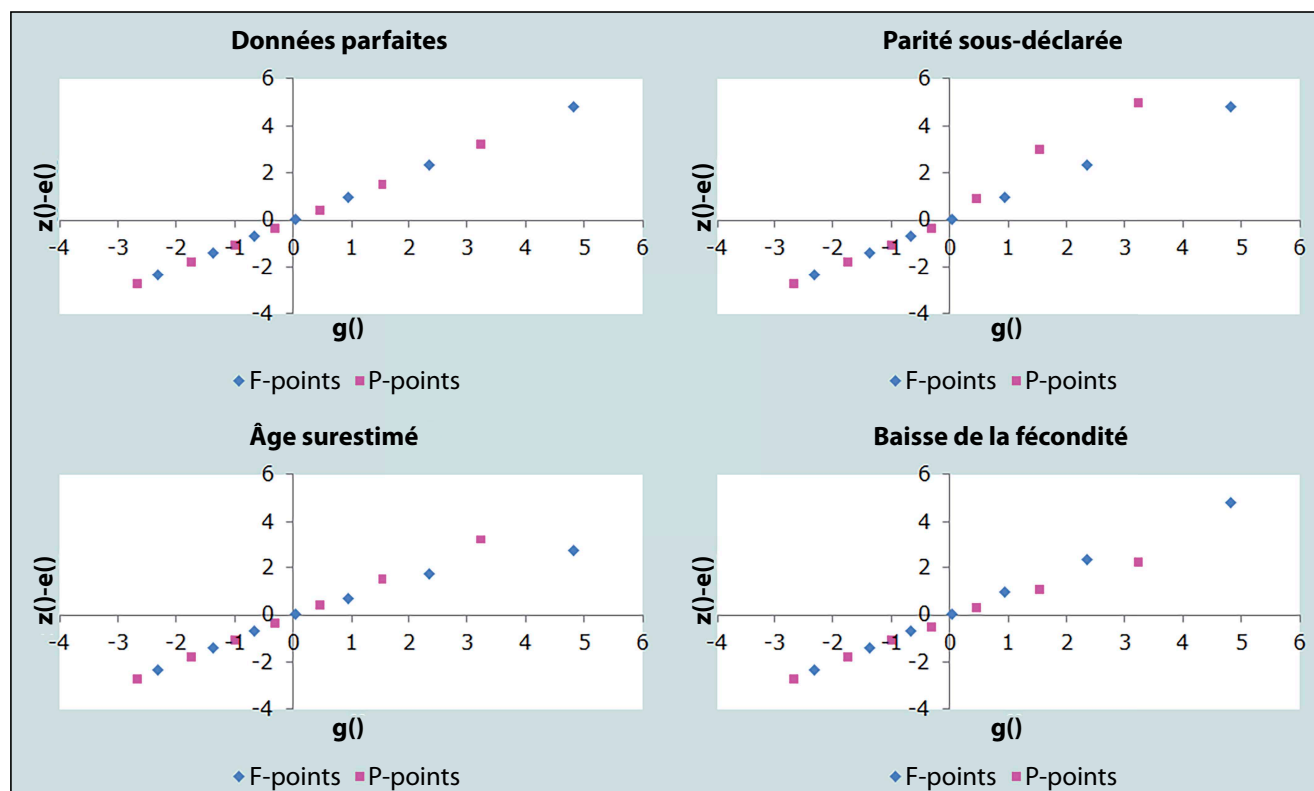
Ces quotients P/F peuvent être représentés graphiquement en ordre inversé, de sorte que le groupe d'âge le plus élevé se situe à gauche. La série des quotients P/F

peut alors être lue comme un déroulement chronologique de gauche à droite (puisque, en général, la fécondité des femmes âgées s'est produite avant celle des femmes jeunes). Des écarts importants par rapport à la tendance générale suggèrent des erreurs dans les données. La décroissance des quotients P/F représentés de cette façon révèle une divergence croissante avec l'âge entre les mesures de la fécondité par cohorte et par période ; elle est donc indicative d'une baisse de la fécondité.

EXEMPLE

Cette illustration de la méthode recourt à des données présentées dans le rapport sur la fécondité tiré du recensement de 2008 au Malawi. La méthode est mise en œuvre dans un classeur Excel disponible sur le site internet.

FIGURE 7.1 Graphiques diagnostiques reposant sur des données simulées correspondant à des erreurs courantes



Étape 1 : Calcul des parités moyennes déclarées

Les parités moyennes figurent au tableau 2.6 du rapport sur la fécondité du recensement de 2008 au Malawi. Le rapport ne précise pas si les parités ont été corrigées ou si une correction d'El-Badry a été appliquée aux données. Les données sont ici au tableau 7.1 :

TABLEAU 7.1 Mesures de la fécondité tirées du recensement de 2008 au Malawi

| Âge (à l'enquête) | Parité moyenne par femme | Taux de fécondité par période |
|-------------------|--------------------------|-------------------------------|
| 15-19 | 0,283 | 0,111 |
| 20-24 | 1,532 | 0,245 |
| 25-29 | 2,849 | 0,230 |
| 30-34 | 4,185 | 0,195 |
| 35-39 | 5,214 | 0,147 |
| 40-44 | 6,034 | 0,072 |
| 45-49 | 6,453 | 0,032 |

Étape 2 : Détermination du classement de l'âge des mères

La question sur la fécondité récente au recensement de 2008 au Malawi était « Combien de naissances vivantes au cours des 12 derniers mois ? ». Comme on ne peut pas dater la naissance de l'enfant, on peut supposer que les données sont classées par âge de la mère à la date du recensement plutôt qu'à la naissance de son enfant.

Étape 3 : Calcul des taux de fécondité par âge

Les taux de fécondité sont présentés au tableau 2.6 du rapport sur la fécondité du recensement de 2008 au Malawi. (Le calcul de ces taux, présenté au tableau 2.3 du rapport, donne à penser que ${}_5f_{20}$ était 0,250, mais les taux du tableau 2.6 sont conservés pour cet exemple de façon à permettre une meilleure comparaison des résultats obtenus).

Étape 4 : Choix du standard de fécondité à appliquer et de la variante du modèle à ajuster

En l'absence d'alternative, nous appliquons le standard de Booth, et – afin de corriger la forme et le niveau des données de fécondité, nous choisissons d'ajuster la variante Forme- F Niveau P . Les coefficients, $e()$ et $g()$, sont calculés dans les tableaux 7.2-7.4 ci-dessous.

Les valeurs du standard figurant dans la colonne [2], on calcule les gompits du standard dans la colonne [3]. Par

exemple, dans le groupe d'âge qui se termine à 19,5 ans, c'est $-\ln(-\ln(0,1140)) = -0,7753$. On notera que les valeurs cumulées s'appliquent à 14,5 ans, 19,5 ans etc., reflétant ainsi le décalage d'une demi-année dans le classement des âges des mères. Les rapports de la fécondité cumulée à deux âges successifs tirée du standard de la colonne [2] figurent à la colonne [4] et leurs gompits à la colonne [5]. Ainsi dans le groupe d'âge se terminant à 39,5 ans, on a $2,6209 = -\ln(-\ln(0,9298)) = -\ln(-\ln(0,9199/0,9893))$.

Les dérivées premières et secondes au point où $\beta = 1$, figurant aux colonnes [6] et [7], sont évaluées grâce aux formules :

$$\begin{aligned}\frac{d}{d\beta}\phi_x(1) &= \phi'_x(1) \\ &= \frac{Y^s(x+5) \cdot \exp(Y^s(x)) + Y^s(x) \cdot \exp(Y^s(x+5))}{\exp(Y^s(x)) - \exp(Y^s(x+5))}; \\ \frac{d^2}{d\beta^2}\phi_x(1) &= \phi''_x(1) \\ &= \frac{(Y^s(x) - Y^s(x+5))^2 \cdot \exp(Y^s(x) + Y^s(x+5))}{(\exp(Y^s(x)) - \exp(Y^s(x+5)))^2}\end{aligned}$$

Enfin $e(x)$ est obtenu dans la colonne [8] en retranchant la colonne [6] de la colonne [5].

Le tableau 7.3 répète les mêmes calculs, mais pour des données sans décalage ; ces valeurs sont nécessaires pour obtenir les estimations de fécondité finales, sans décalage. Le tableau 7.4 montre comment sont calculées les valeurs de $e(i)$, $g(i)$ et c à utiliser avec les données de parité, en utilisant les parités tirées du standard pour débiter dans la colonne [2].

Étape 5 : Évaluer la représentation graphique des P -points et des F -points

En respectant les instructions données plus haut, nous commençons par ajuster les modèles en utilisant tous les P -points et les F -points, respectivement.

Les résultats sont représentés sur le premier graphique à la page des **Graphiques diagnostiques** du dossier Excel associé disponibles sur le site web (Figure 7.2) :

Les deux droites ajustées aux P -points et aux F -points sont très proches l'une au dessus de l'autre, mais aucun des deux ajustements n'est particulièrement bon. Les F -points s'incurvent nettement vers le bas aux âges élevés, ce qui laisse supposer une exagération des âges déclarés, et le fait que les

TABLEAU 7.2 Calcul de $e(x)$ et $g(x)$ quand les données sont décalées d'une demi-année

| Âge x [1] | $F^s(x)/F$ [2] | $Y^s(x)$ [3] | Ratio [4] | Phi [5] | Phi' [6] | Phi'' [7] | $e(x)$ [8] | $g(x)$ [9] |
|----------------|-------------------|-----------------|---------------------|------------|-------------|--------------|---------------|---------------|
| | | =gompit[2] | = $Y^s(x)/Y^s(x+5)$ | =gompit[4] | | | =[5]–[6] | =[6] |
| 14 ½ | 0,0011 | – 1,9228 | 0,0094 | – 1,5410 | – 2,4565 | | 0,9155 | – 2,4565 |
| 19 ½ | 0,1140 | – 0,7753 | 0,3233 | – 0,1216 | – 1,4527 | 0,9563 | 1,3311 | – 1,4527 |
| 24 ½ | 0,3528 | – 0,0411 | 0,6007 | 0,6741 | – 0,7426 | 0,9632 | 1,4167 | – 0,7426 |
| 29 ½ | 0,5872 | 0,6305 | 0,7529 | 1,2592 | – 0,0364 | 0,9530 | 1,2957 | – 0,0364 |
| 34 ½ | 0,7800 | 1,3925 | 0,8479 | 1,8021 | 0,8405 | | 0,9615 | 0,8405 |
| 39 ½ | 0,9199 | 2,4830 | 0,9298 | 2,6209 | 2,1799 | | 0,4409 | 2,1799 |
| 44 ½ | 0,9893 | 4,5323 | 0,9893 | 4,5324 | 4,5315 | | 0,0010 | 4,5315 |
| Phi''–bar | | | | | | 0,9575 | | |

TABLEAU 7.3 Calcul de $e(x)$ et $g(x)$ quand les données ne sont pas sujettes à un décalage des âges

| Âge x [1] | $F^s(x)/F$ [2] | $Y^s(x)$ [3] | Ratio [4] | Phi [5] | Phi' [6] | Phi'' [7] | $e(x)$ [8] | $g(x)$ [9] |
|----------------|-------------------|-----------------|--------------|------------|-------------|--------------|---------------|---------------|
| 15 | 0,0028 | – 1,7731 | 0,0204 | – 1,3591 | – 2,3278 | | 0,9688 | – 2,3278 |
| 20 | 0,1358 | – 0,6913 | 0,3600 | – 0,0214 | – 1,3753 | 0,9582 | 1,3539 | – 1,3753 |
| 25 | 0,3773 | 0,0256 | 0,6200 | 0,7379 | – 0,6748 | 0,9629 | 1,4127 | – 0,6748 |
| 30 | 0,6086 | 0,7000 | 0,7644 | 1,3143 | 0,0393 | 0,9510 | 1,2750 | 0,0393 |
| 35 | 0,7962 | 1,4787 | 0,8559 | 1,8607 | 0,9450 | | 0,9157 | 0,9450 |
| 40 | 0,9302 | 2,6260 | 0,9378 | 2,7455 | 2,3489 | | 0,3966 | 2,3489 |
| 45 | 0,9919 | 4,8097 | 0,9919 | 4,8098 | 4,8086 | | 0,0012 | 4,8086 |
| Phi''–bar | | | | | | 0,9574 | | |

TABLEAU 7.4 Calcul des $e(i)$ et $g(i)$ à partir des données de parité

| Âge i [1] | $P^s(i)$ [2] | $Y^s(i)$ [3] | Ratio [4] | Phi [5] | Phi' [6] | Phi'' [7] | $e(i)$ [8] | $g(i)$ [9] |
|----------------|-----------------|-----------------|---------------------|------------|-------------|--------------|---------------|---------------|
| | | =gompit[2] | = $Y^s(i)/Y^s(i+1)$ | =gompit[4] | | | =[5]–[4] | =[6] |
| 0 | 0,0003 | –2,0961 | 0,0056 | –1,6449 | –2,6738 | | 1,0289 | –2,6738 |
| 1 | 0,0521 | –1,0833 | 0,2044 | –0,4622 | –1,7469 | 0,9519 | 1,2846 | –1,7469 |
| 2 | 0,2549 | –0,3124 | 0,5143 | 0,4081 | –1,0159 | 0,9638 | 1,4240 | –1,0159 |
| 3 | 0,4957 | 0,3541 | 0,7014 | 1,0367 | –0,3349 | 0,9597 | 1,3717 | –0,3349 |
| 4 | 0,7067 | 1,0579 | 0,8140 | 1,5810 | 0,4406 | | 1,1404 | 0,4406 |
| 5 | 0,8681 | 1,9561 | 0,8969 | 2,2184 | 1,5162 | | 0,7022 | 1,5162 |
| 6 | 0,9679 | 3,4225 | 0,9701 | 3,4943 | 3,2238 | | 0,2705 | 3,2238 |
| Phi''–bar | | | | | | 0,9585 | | |

P-points sont juste en dessous des *F*-points indique un léger recul de la fécondité en cours.

Étape 6 : Ajustement du modèle par sélection des points à utiliser

L'examen du graphique donne à penser que l'ajustement serait meilleur si on omettait les *P*-points et les *F*-points du dernier groupe d'âge. On le vérifie en examinant le graphique modifié (Figure 7.3) :

Les droites ne sont plus aussi proches l'une de l'autre et elles ne sont plus parallèles. L'examen visuel suggère que la suppression de l'avant-dernier *P*-point pourrait permettre à l'ensemble des points restants d'être alignés sur une seule droite (Figure 7.4).

À toutes fins utiles, on peut considérer que ces points s'alignent sur une seule droite, ce qui implique que les parités moyennes et les taux de fécondité qui sous-tendent ces points sont cohérents les uns avec les autres. L'élimination du *P*-point associé au groupe 35-39 ans permettrait peut-être d'améliorer encore l'ajustement, mais il n'est pas

nécessaire de réduire encore le nombre de points inclus dans le modèle pour une très faible amélioration de l'ajustement. L'exclusion de ce *P*-point entraîne un accroissement faible de l'écart quadratique moyen, de 0,044 à 0,045.

Nous pouvons accepter cet ajustement du modèle relationnel de Gompertz. La troisième figure sur la feuille de calcul indique que l'équation de la droite du meilleur ajustement des neuf points restants est $z() - e() = 0,9936 \cdot g() - 0,0272$.

À partir de là, la valeur de β est connue directement (0,9936), et celle de α est tirée de la formule

$$\alpha = -0.0272 - \frac{1}{2}((\beta - 1)^2 c) = -0.0272$$

où c est la moyenne des $\phi_x''(1)$ tirés du tableau 7.2 pour les données de fécondité actuelle (puisque les données sont sujettes initialement à un décalage d'une demi-année) et du tableau 7.4 pour les données de parité.

FIGURE 7.2 Graphique des $z() - e()$ en fonction de $g()$, tous les points. Recensement de 2008 au Malawi

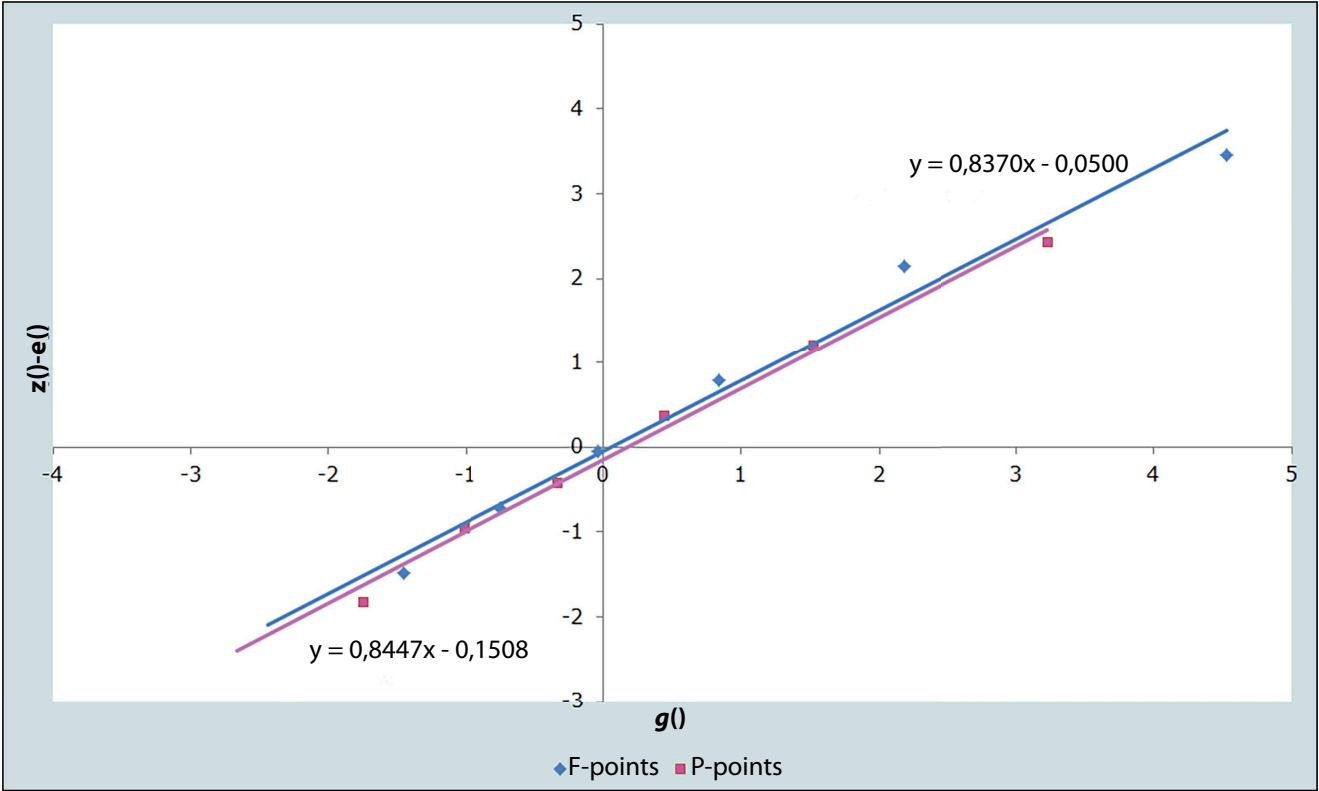


FIGURE 7.3 Graphique des $z() - e()$ en fonction de $g()$, après élimination des F- et des P-points associés au groupe 45-49 ans. Recensement de 2008 au Malawi

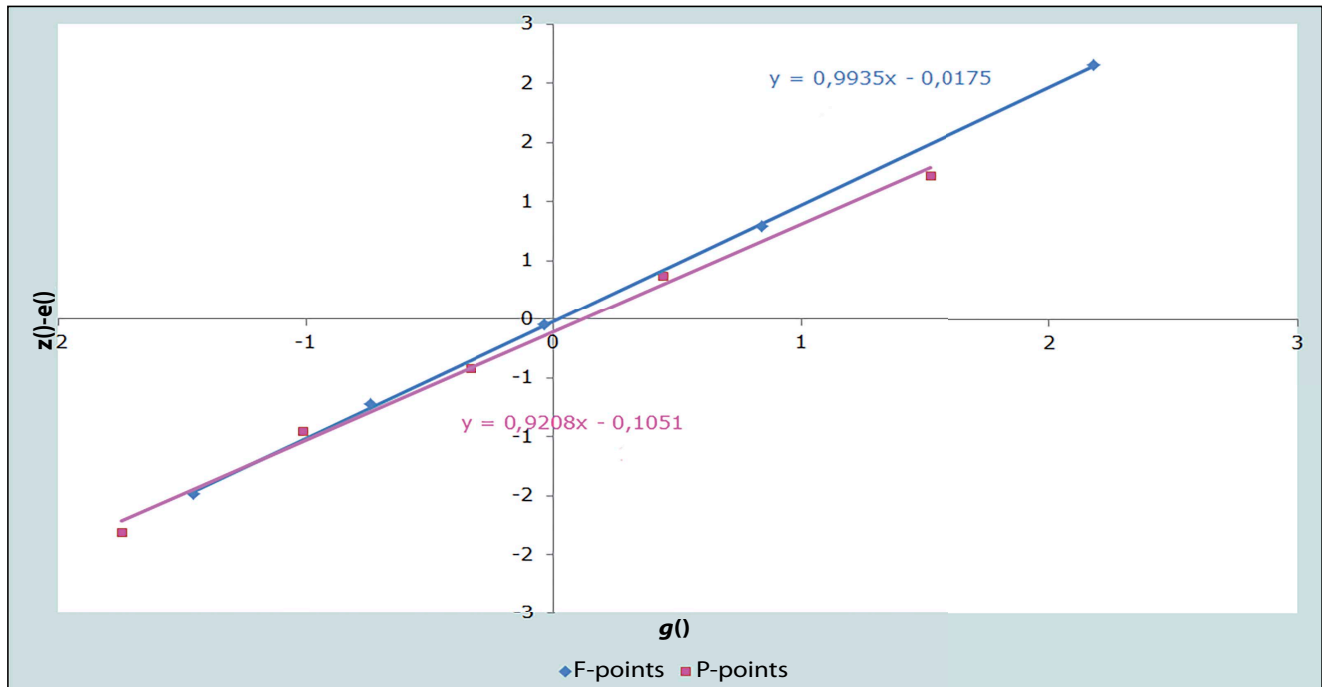
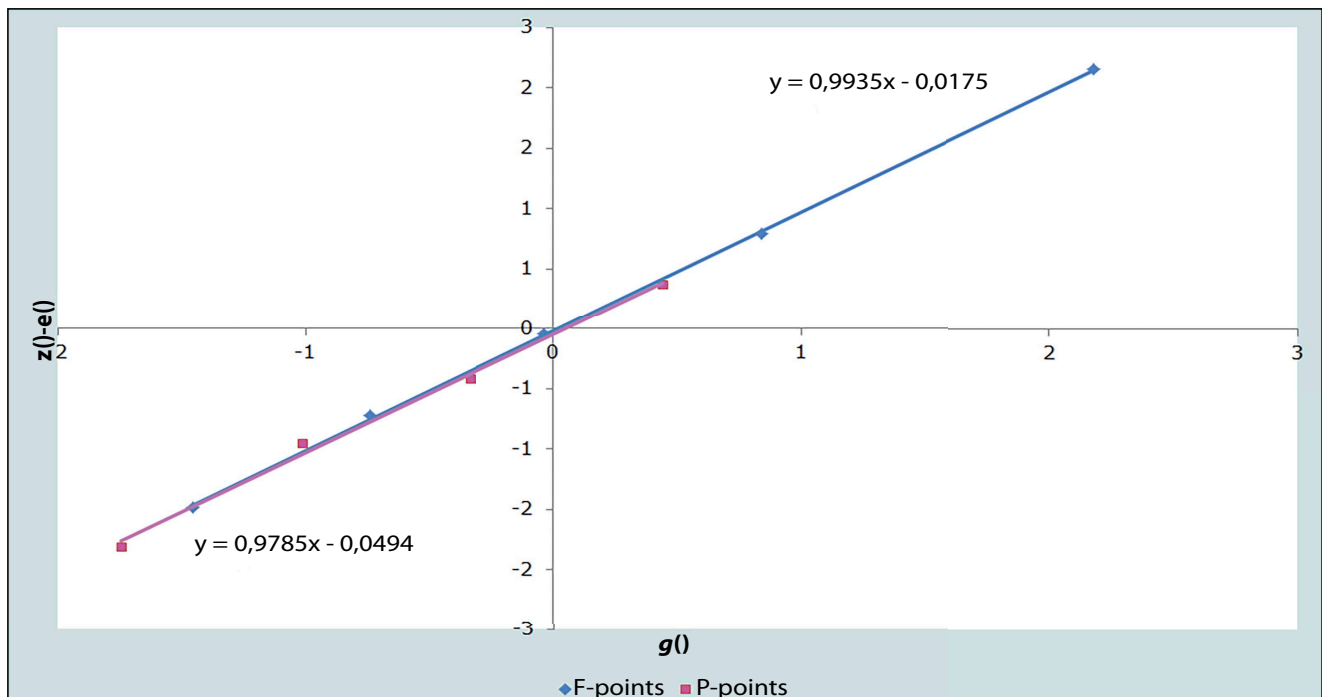


FIGURE 7.4 Graphique des $z() - e()$ en fonction de $g()$, après élimination du P-point associé au groupe 40-44 ans. Recensement de 2008 au Malawi



Étape 7 : Évaluer la qualité des paramètres estimés

Les valeurs estimées de α (– 0,0272) et β (0,9936) sont très proches des valeurs standards de 0 et 1. La valeur légèrement négative de α montre que la distribution de la fécondité observée au Malawi en 2008 est un peu plus vieille que le standard et la valeur inférieure à 1 de β suggère que la distribution est légèrement plus étalée que celle du standard.

Étape 8 : Taux de fécondité par âge et indice synthétique de fécondité ajustés

Pour déterminer le niveau global de fécondité, on applique les valeurs ajustées de α et β aux gompits des parités standards (colonne [3] du tableau 7.4) pour les groupes d'âge dont les *P*-points ont été inclus dans le modèle, et on calcule les anti-gompits (Tableau 7.5). En divisant les parités moyennes observées pour un groupe d'âge donné par l'anti-gompit ajusté, on obtient le niveau de fécondité découlant des parités moyennes (colonne [6] du tableau 5), et on déduit une estimation de l'indice synthétique de fécondité, \hat{T} , en faisant la moyenne arithmétique de ces estimations (= 5,9784).

Pour obtenir les taux de fécondité par âge associés dans des âges conventionnels, nous appliquons à nouveau α and β , mais cette fois aux gompits de fécondité actuelle, $Y^s(x)$,

TABLEAU 7.5 Calcul de l'indice synthétique de fécondité estimé, \hat{T} Recensement de 2008 au Malawi

| Groupe d'âge <i>i</i> | $Y^s(i)$ | $Y(i)$ | Anti-gompit | $P(i)$ | Niveau de fécondité déduit |
|---------------------------|----------|----------|-------------|--------|----------------------------|
| [1] | [2] | [3] | [4] | [5] | [6] |
| $= \alpha + \beta Y^s(i)$ | | | | | |
| 1 | – 1,0833 | – 1,1034 | 0,0491 | 0,283 | 5,7662 |
| 2 | – 0,3124 | – 0,3375 | 0,2462 | 1,532 | 6,2218 |
| 3 | 0,3541 | 0,3246 | 0,4854 | 2,849 | 5,8694 |
| 4 | 1,0579 | 1,0239 | 0,6982 | 4,185 | 5,9937 |
| 5 | 1,9561 | 1,9162 | 0,8631 | 5,214 | 6,0407 |
| T-chapeau | | | | | 5,9784 |

au tableau 7.3. En prenant les anti-gompits des valeurs ajustées, on obtient une distribution de la fécondité cumulée. Ces proportions sont ensuite amenées au niveau de l'indice synthétique de fécondité, \hat{T} , pour établir la distribution absolue cumulée de la fécondité. En différenciant et en divisant par 5 on obtient les taux de fécondité par âge finaux (Tableau 7.6).

Les taux de fécondité par âge figurent à la dernière colonne du tableau 7.6, pour une fécondité totale (indice synthétique, 15-49 ans) de 5,96 enfants par femme.

TABLEAU 7.6 Calcul des taux de fécondité corrigés, Malawi, Recensement de 2008

| Groupe d'âge (jusqu'à âge <i>x</i>) | $Y^s(i)$ | $Y(i)$ | Anti-gompit | Amené au niveau de T chapeau | Taux de fécondité par âge |
|--------------------------------------|----------|----------|-------------|------------------------------|-----------------------------------|
| [1] | [2] | [3] | [4] | [5] | [6] |
| $= \alpha + \beta Y^s(i)$ | | | | $F(x) = [4] * 5,9784$ | ${}_5f_{x-5} = (F(x) - F(x-5))/5$ |
| 15 | – 1,7731 | – 1,7887 | 0,0025 | 0,0151 | 0,0030 |
| 20 | – 0,6913 | – 0,7140 | 0,1298 | 0,7758 | 0,1521 |
| 25 | 0,0256 | – 0,0017 | 0,3673 | 2,1956 | 0,2840 |
| 30 | 0,7000 | 0,6683 | 0,5989 | 3,5807 | 0,2770 |
| 35 | 1,4787 | 1,4419 | 0,7894 | 4,7194 | 0,2277 |
| 40 | 2,6260 | 2,5817 | 0,9271 | 5,5428 | 0,1647 |
| 45 | 4,8097 | 4,7512 | 0,9914 | 5,9269 | 0,0768 |
| 50 | 13,8155 | 13,6984 | 1,0000 | 5,9784 | 0,0103 |

DESCRIPTION DÉTAILLÉE DE LA MÉTHODE

Introduction

Le modèle relationnel de Gompertz a été conçu à partir de la méthode du quotient P/F de Brass. Il fonctionne avec les mêmes données de base ; il utilise les données de parité des jeunes femmes pour déterminer le niveau de la fécondité, alors que la distribution de celle-ci est déterminée à partir des déclarations par les femmes de leurs naissances récentes.

Exposé mathématique

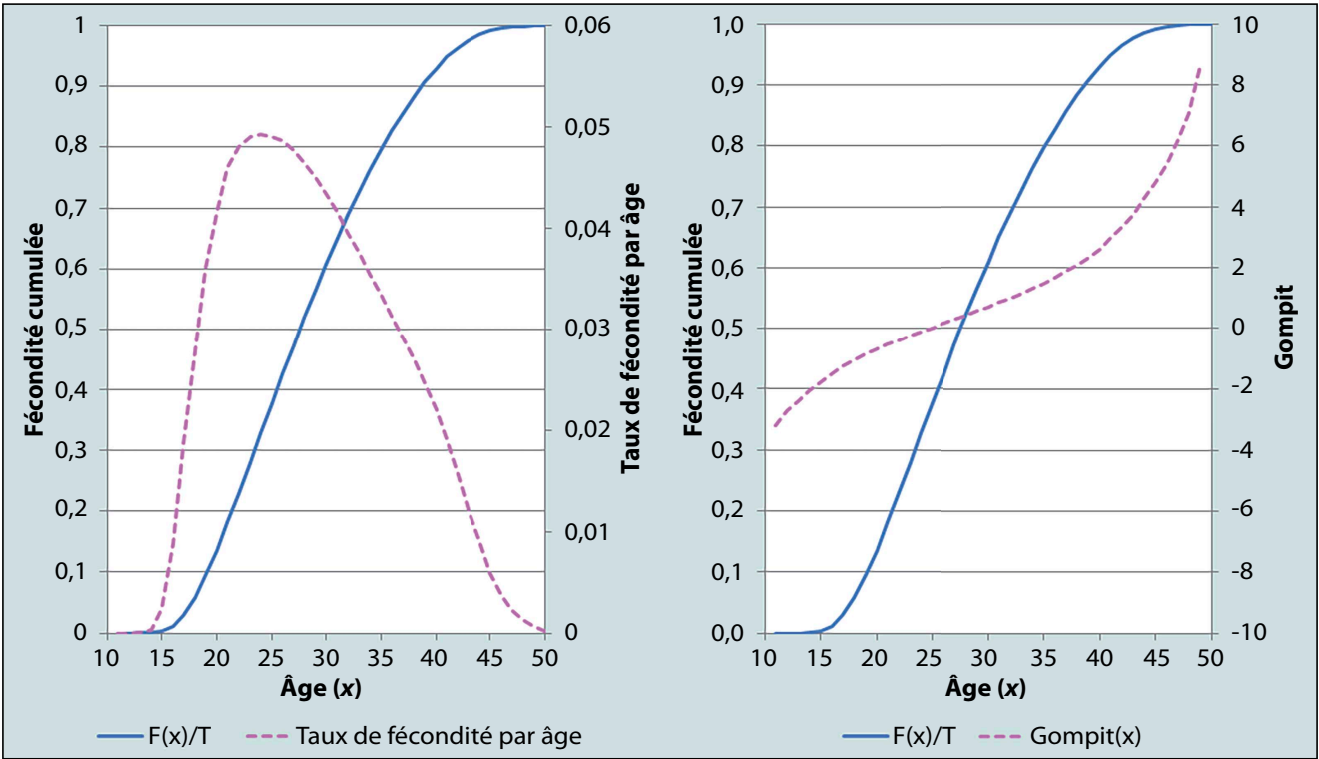
Le modèle relationnel de Gompertz pour la fécondité, développé initialement par Brass (1978), est analogue sur bien des points aux modèles logit de mortalité. Le modèle permet de décrire toute distribution de la fécondité par référence à une distribution standard et aux paramètres utilisés pour la transformer afin de produire la distribution cherchée. La transformation utilisée comme base pour établir une relation entre les deux distributions de fécondité est connue comme la transformation de Gompertz. Dans la formulation originelle du modèle, elle est réalisée sur une distribution

proportionnelle cumulée, c'est-à-dire où chaque cumulant des parités ou de la fécondité est exprimé en proportion de l'indice synthétique de fécondité (la distribution totalisée). Ce total prend la valeur de un. Les proportions transformées sont intitulées gompits et sont données par

$$Y(x) = \text{gompit}\left(\frac{F(x)}{F}\right) = -\ln\left(-\ln\left(\frac{F(x)}{F}\right)\right)$$

où $F(x)$ est la somme des taux de fécondité par âge cumulés jusqu'à l'âge x et F est l'indice synthétique de fécondité. La même relation vaut pour les parités, en remplaçant $F(x)$ par la parité moyenne par groupe d'âge, et F par la parité cumulée à 50 ans ou +. La transformation de Gompertz « étire » l'axe des âges de sorte que les gompits représentés graphiquement en fonction de l'âge s'alignent à peu près sur une droite. La transformation n'est toutefois pas parfaite, la ligne tendant à s'incurver légèrement aux deux bouts, comme on peut le voir sur la figure 7.5, où sont représentés les taux de fécondité tirés du standard de Booth (1984).

FIGURE 7.5 Effet de la transformation gompit sur une distribution de la fécondité



La transformation peut être à la base d'un modèle relationnel parce que les gompits de différentes séries de taux de fécondité en fonction de l'âge tendent à s'écarter de la linéarité de la même manière ; c'est pourquoi la relation entre deux telles séries de gompits est généralement proche de la linéarité. Le recours au modèle dans une forme relationnelle permet une estimation des paramètres du modèle par l'ajustement de droites, ce qui est une procédure simple dont les résultats sont aisés à interpréter.

La relation entre les gompits des cumulants de la fécondité de deux séries de distributions de la fécondité étant à peu près linéaire, on peut relier les gompits d'une distribution observée de la fécondité aux gompits d'une distribution standard fondée sur des données fiables, au moyen d'une relation

$$Y(x) = \alpha + \beta Y^s(x)$$

où $Y(x)$ est le gompit de la fécondité proportionnée cumulée à l'âge x , et $Y^s(x)$ est le gompit des cumulants de la fécondité standard.

Dans cette formulation, α indique dans quelle mesure les âges de fécondité dans la population diffèrent de ceux du standard, les valeurs négatives de α désignant un schéma de fécondité plus âgé. β indique dans quelle mesure l'étalement de la fécondité diffère de celui de la population standard. L'étalement de la distribution est moindre pour les valeurs supérieures à l'unité.

C'est en fait un modèle à trois paramètres. En convertissant les gompits ajustés en estimations de la fécondité cumulée, grâce à la transformation inversée, on obtient une distribution proportionnelle dont la somme est égale à un. Un troisième paramètre est nécessaire pour amener les valeurs ajustées au niveau de fécondité approprié. C'est l'indice synthétique de fécondité – celui-là même qu'on cherche à estimer – mais son estimation à partir des données observées peut ne pas être fiable à cause de déclarations erronées. C'est pourquoi Zaba (1981) a adapté la procédure originelle d'ajustement (qui n'est pas décrite ici). Nous exposons ci-dessous ses contributions et ses extensions de la méthode.

La démarche de Zaba consiste à utiliser les gompits du ratio de deux cumulants successifs de la fécondité, de façon à isoler les paramètres de forme de l'estimation du niveau de fécondité :

$$Y(x) = \text{gompit} \left(\frac{F(x)}{F(x+5)} \right) = -\ln \left(-\ln \left(\frac{F(x)}{F(x+5)} \right) \right)$$

Si le cumulant $F(x)$ suit un modèle de Gompertz de paramètres α and β , alors

$$\begin{aligned} Y(x) &= -\ln \left(\exp \left(-(\alpha + \beta Y^s(x)) \right) - \exp \left(-(\alpha + \beta Y^s(x+5)) \right) \right) \\ &= \alpha - \ln \left(\exp \left(-\beta Y^s(x) \right) - \exp \left(-\beta Y^s(x+5) \right) \right) \\ &= \alpha + \phi_x(\beta) \end{aligned}$$

où $\phi_x(\beta)$ est le second terme de l'avant-dernière ligne.

Pour les valeurs de β proches de 1, $\phi_x(\beta)$ peut être approximé par un développement en série de Taylor aux environs de $\beta = 1$:

$$\begin{aligned} \phi_x(\beta) &= \phi_x(1) + (\beta - 1) \frac{d}{d\beta} \phi_x(1) + \frac{(\beta - 1)^2}{2} \frac{d^2}{d\beta^2} \phi_x(1) + \dots \end{aligned} \quad (3)$$

À partir de la définition de $\phi_x(\beta)$, $\phi_x(1) = Y(x)$.

On peut en outre montrer que

$$\begin{aligned} \frac{d}{d\beta} \phi_x(1) &= \frac{Y^s(x+5) \cdot \exp(Y^s(x)) + Y^s(x) \cdot \exp(Y^s(x+5))}{\exp(Y^s(x)) - \exp(Y^s(x+5))}; \\ \frac{d^2}{d\beta^2} \phi_x(1) &= \frac{(Y^s(x) - Y^s(x+5))^2 \cdot \exp(Y^s(x) + Y^s(x+5))}{(\exp(Y^s(x)) - \exp(Y^s(x+5)))^2} \end{aligned} \quad (4)$$

Zaba (1981) a évalué cette dernière quantité pour différentes valeurs de x , et elle a montré qu'elle était à peu près constante dans l'intervalle $15 \leq x < 30$ ans. (On peut aussi le constater aux tableaux 7.2-7.4, où cette quantité a été dérivée). On peut donc remplacer $\phi_x'(1)$ par c , la moyenne arithmétique des quantités dans cet intervalle d'âge, et réécrire l'équation 3

$$\begin{aligned} Y_x &= \alpha + \phi_x(1) + (\beta - 1) \phi_x'(1) + (\beta - 1)^2 \frac{c}{2} \quad \text{ou} \\ Y_x + \phi_x'(1) - \phi_x(1) &= \alpha + (\beta - 1)^2 \frac{c}{2} + \beta \phi_x'(1) \end{aligned}$$

En d'autres termes, il existe une relation linéaire entre $Y_x + \phi'_x(1) - \phi_x(1)$ et $\phi'_x(1)$. Dans un travail ultérieur (Sloggett, Brass, Eldridge *et al.* 1994), Zaba a redéfini ces termes de la façon suivante :

| Terme | Terme redéfini |
|--------------------------|----------------|
| $\phi'_x(1)$ | $g(x)$ |
| $\phi_x(1) - \phi'_x(1)$ | $e(x)$ |
| $Y(x)$ | $z(x)$ |

D'où la notation révisée

$$z(x) - e(x) = \alpha + (\beta - 1)^2 \frac{c}{2} + \beta g(x)$$

qui implique une relation linéaire entre $z - e$ et g .

En appliquant le même raisonnement que ci-dessus, on peut obtenir la formulation équivalente pour $z(i) - e(i)$ en termes de α , β , c et $g(i)$.

Variantes de la procédure d'ajustement

Dans la version standard du modèle, exposée ici, les données sur la fécondité récente sont utilisées pour déterminer la forme du schéma de fécondité et le niveau de référence est fixé par les points de parité (choisis). D'autres variantes sont possibles qui privilégient une série de données par rapport à l'autre de différentes façons. Nous exposons ici celle qui ne recourt qu'aux données sur la fécondité récente.

La variante *F-seulement* privilégie les données sur la fécondité récente et les utilise à la fois pour fixer la forme et le niveau de la fécondité dans le modèle. L'analyste ne doit donc utiliser cette variante que lorsqu'il ne dispose pas de données de parité ou qu'il ne souhaite pas qu'elles jouent un rôle dans l'ajustement du modèle. Cette variante se contente de lisser les taux de fécondité observés en utilisant un modèle relationnel de Gompertz.

Il existe une autre extension du modèle relationnel de Gompertz qui ne recourt qu'aux données de parité pour estimer la fécondité à partir des accroissements de parité par cohorte [14]. Il y a aussi une version modifiée du modèle relationnel de Gompertz qui recourt aux données de deux recensements ou enquêtes [15] pour estimer la fécondité entre les enquêtes à partir de ces données.

CONSTRUCTION DES STANDARDS

Le standard de Booth

L'élaboration du standard de Booth est décrite en détail dans Booth (1984). On retiendra d'une part que le standard est destiné aux populations de fécondité moyenne ou élevée et d'autre part qu'il a été obtenu à partir d'un certain nombre de schémas produits par le modèle de fécondité de Coale-Trussell. Il est donc sujet aux contraintes imposées par ce modèle, lesquelles ne sont généralement pas matérielles.

Le standard utilisé ici n'est pas exactement celui publié par Booth. En premier lieu, le standard de Zaba (1981) diffère légèrement de celui de Booth avant 15 ans pour un meilleur ajustement des situations de fécondité très précoce. Pour en tenir compte, on peut reconstruire entièrement le tableau des coefficients présenté dans Zaba (1981) et dans Sloggett, Brass, Eldridge *et al.* (1994). Le standard utilisé ici est le même pour les coefficients sans décalage. Quand le décalage est nécessaire, de petites différences apparaissent, qui tiennent au mode d'interpolation du standard originel de Booth. Zaba (1981) a calculé les valeurs pour $F(x + 1/2)$, $F(x + 3/2)$, etc. en interpolant entre les valeurs successives de $F(x)$, $F(x + 1)$, $F(x + 2)$. Mais comme la transformation gompit rend $F(x)$ linéaire, il est plus logique d'interpoler les gompits des $F(x)$, $Y(x)$, par demi-année d'âge et d'établir ensuite les valeurs de $F(x + 1/2)$, $F(x + 3/2)$ etc. en prenant les anti-gompits appropriés.

Construction d'autres standards

Comme on l'a déjà dit, le standard de Booth était conçu pour les pays de fécondité moyenne ou forte. Pour appliquer le modèle relationnel de Gompertz à des pays de basse fécondité ou des pays dont les schémas de fécondité sont originaux, d'autres standards sont nécessaires. Nous expliquons brièvement ici comment procéder.

La démarche nécessite à la base une série de $F(x)$ qui puisse être convertie au moyen d'un gompit en une série de $Y(x)$. On en tire ensuite des valeurs de $\phi'_x(1)$, $\phi'_x(1)$ et $\phi'_x(1)$ en recourant aux relations établies par les équations 3 et 4. À partir de là, on peut calculer des tableaux de $z()$, $e()$ et $g()$. Comme signalé plus haut, les valeurs de $\phi''_x(1)$ sont à peu près constantes, pour un standard donné, entre 15 et 30 ans, aussi la moyenne des trois valeurs (15-19, 20-24, 25-29) donne-t-elle une estimation du terme constant, c .

La construction d'un nouveau standard doit s'appuyer sur une série de taux de fécondité par âge, $f^s(x)$, de bonne qualité. En se référant à l'analyse démographique traditionnelle, on peut alors définir les cumulants équivalents :

$$F^s(x) = \int_0^x f^s(a) da$$

La fonction $f(a)$ n'étant pas intégrable dans la plupart des cas, il faut il faut recourir à des techniques numériques pour approximer l'intégrale. En utilisant de façon récursive la formule du trapèze composite

$$F^s(x) \approx F^s(x-1) + \frac{1}{2} \left(f^s(x-1) + f^s(x) \right)$$

on calcule aisément le gompit $z(x)$.

En s'appuyant sur les propriétés d'un développement de Taylor décrit à l'équation 4, on peut définir les composantes de $e(x)$ et $g(x)$ et en déduire des expressions de ces quantités.

Les valeurs de $z(i)$, $g(i)$ et $e(i)$ sont définies de la même façon, en y ajoutant la nécessité d'en tirer les parités constantes associées à $F(x)$. Dans chaque groupe d'âge $[x, x+n)$, les parités sont données par

$${}_n P_x = \int_x^{x+n} F(a) da$$

RÉFÉRENCES

- Booth H. 1980. "The estimation of fertility from incomplete cohort data by means of the transformed Gompertz model." Unpublished PhD thesis, London: University of London.
- Booth H. 1984. "Transforming Gompertz' function for fertility analysis: The development of a standard for the relational Gompertz function", *Population Studies* **38**(3):495-506. doi: <http://dx.doi.org/10.2307/2174137> [14]
- Brass W. 1978. *The relational Gompertz model of fertility by age of woman*. London: Centre for Population Studies, London School of Hygiene and Tropical Medicine.
- Sloggett A, W Brass, SM Eldridge, IM Timæus, P Ward and B Zaba (eds). 1994. *Estimation of Demographic Parameters from Census Data*. Tokyo: Statistical Institute for Asia and the Pacific.
- US Census Bureau. 1997. *Population Analysis Spreadsheets for Excel*. Washington, D.C: US Bureau of the Census. <http://www.census.gov/population/international/software/pas> [15] Accessed: 1 January 2013.
- Zaba B. 1981. *Use of the Relational Gompertz Model in Analysing Fertility Data Collected in Retrospective Surveys*. Centre for Population Studies Research Paper 81-2. London: Centre for Population Studies, London School of Hygiene & Tropical Medicine.

qui peut aussi être évalué en recourant à la formule du trapèze composite.

AUTRES LECTURES ET RÉFÉRENCES

En plus des textes déjà cités, les travaux sur le modèle relationnel de Gompertz sont peu nombreux. C'est sans doute dû en partie au fait qu'il a été décrit par Booth (1984) peu après la publication du *Manuel X*, un exposé cohérent du mode d'application du modèle n'apparaissant ensuite que dans le manuel de l'ISAP (Sloggett, Brass, Eldridge *et al.* 1994). La méthode a été appliquée dans de nombreux cas à travers le monde, mais pas dans la forme décrite ici.

Par exemple, la suite de feuilles de calcul (PASEX) préparée par le US Census Bureau (1997) offre une version simplifiée du modèle, l'utilisateur étant forcé d'ajuster P et F par des droites en utilisant seulement 2 P -points et 2 F -points, ou 3 de chaque, sans beaucoup s'attacher à la cohérence interne des points choisis. C'est la voie adoptée par l'Office Statistique du Malawi dans son analyse des données de fécondité du recensement de 2008. Étant donné la forte cohérence de ces données pour l'ensemble des femmes âgées de moins de 40 ans, les résultats présentés dans ce rapport (indice synthétique de fécondité = 6,0) ne diffèrent guère de ceux présentés plus haut à titre d'exemple. Avec des données de moindre qualité, une telle congruence des résultats de diverses applications ne serait pas assurée.

Chapitre 8 PROBABILITÉS D'AGRANDISSEMENT DES FAMILLES

Tom A Moultrie and Basi Zaba

INTRODUCTION

Quand les femmes ou les parents pensent à « avoir des enfants », il s'agit généralement de commencer à former une famille ou d'avoir un autre enfant. La présente section traite des aspects séquentiels de la constitution de la descendance, en s'attachant aux proportions de mères ayant atteint une parité ou un rang de naissance donné et à la proportion de celles qui accroissent ensuite leur parité d'au moins un enfant supplémentaire.

L'information sur le rang de naissance s'obtient aisément à partir de deux questions simples du recensement : L'une sur la descendance atteinte par les femmes et l'autre sur les naissances au cours de l'année écoulée. Grâce à la première question, on peut décomposer les naissances de l'année passée par rang de naissance. On obtient de cette manière des informations supplémentaires sur les tendances de la fécondité et on peut évaluer les changements de la distribution de la descendance par parité en calculant les probabilités d'agrandissement des familles.

Une probabilité d'agrandissement des familles désigne la proportion de femmes qui passent d'une parité à la suivante. Les probabilités peuvent être calculées pour des cohortes de femmes définies par l'âge ou le mariage. On considère généralement des cohortes d'âge, les probabilités d'agrandissement étant dans ce cas calculées à partir de la distribution par parité des femmes d'un groupe d'âge donné. Pour les cohortes ayant achevé leur vie féconde, ces mesures restent fixes, sous condition que les femmes âgées ne soient pas soumises à une mortalité différentielle par parité. Mais pour les cohortes encore dans leur période féconde, les mesures changent du fait que des femmes passent à une parité plus élevée. Les mesures relatives aux jeunes femmes souffrent donc à la fois d'effets de censure et de sélection, les femmes ayant eu des enfants rapidement étant surreprésentées dans les combinaisons âge-parité avant la fin de la période

féconde. Les comparaisons de probabilités d'agrandissement entre femmes jeunes et âgées sont donc trompeuses.

Brass (1985) décrit une technique pour projeter les probabilités d'agrandissement jusqu'à la fin de la période reproductive, afin que les données de parité des jeunes femmes puissent être utilisées. L'intérêt de décomposer la fécondité par parité est que différentes probabilités d'agrandissement peuvent répondre à différents facteurs, de sorte que l'évolution de certaines probabilités d'agrandissement peut nous renseigner sur les processus en œuvre dans les variations de la fécondité d'une population. Par exemple, les probabilités d'agrandissement mesurent aisément les restrictions de fécondité (ou leur absence) liées au nombre d'enfants déjà nés. Elles peuvent aussi être moins sujettes que d'autres indicateurs plus courants à certains types d'erreur sur les données (par exemple les erreurs sur la période de référence). Le calcul de taux de fécondité de premier rang, réalisé lors de la construction de probabilités d'agrandissement projetées, peut à la fois donner des indications sur la qualité des données et sur l'évolution de la fécondité.

Le calcul des probabilités d'agrandissement est simple pour les cohortes qui ont terminé leur vie féconde. La comparaison de cohortes successives peut donner des informations sur les tendances de la fécondité, mais des conclusions plus fiables peuvent être tirées s'il est possible de comparer les probabilités d'agrandissement des mêmes cohortes à travers plusieurs recensements. Pour les jeunes femmes, qui fournissent des informations sur des tendances plus récentes de la fécondité, une procédure plus élaborée est nécessaire. Il s'agit de prévoir la distribution par parité que les jeunes femmes auraient si elles conservaient les taux actuels de fécondité par âge et rang jusqu'à la fin de leur période féconde. Cette distribution par parité permet de calculer les probabilités d'agrandissement projetées, qui expriment les variations de fécondité implicites dans la combinaison des histoires géné-

siques jusqu'à ce jour et des taux de fécondité actuels par âge et rang. Un des avantages des probabilités d'agrandissement projetées est qu'il s'agit de ratios, les erreurs dans les taux de fécondité par âge de chaque rang dues à des sous-déclarations tendant à se compenser, quand ces taux sont utilisés pour projeter les probabilités d'agrandissement, à l'exception du passage de la parité nulle à la première naissance.

Nous introduisons deux nouveaux concepts. Le premier, $M(i)$, proportion de femmes ayant atteint la parité i , est la proportion de celles ayant au moins i enfants. Le second est la probabilité d'agrandissement $a(i)$, qui mesure la proportion de femmes qui passent d'une parité i à la suivante, $i + 1$. Les proportions projetées de femmes atteignant la parité i , $M^*(i)$, et les probabilités d'agrandissement $a^*(i)$ désignent respectivement les proportions de femmes dont on prévoit qu'elles atteindront au moins la parité i , et qu'elles passeront de la parité i à la parité $i + 1$ d'ici à la fin de leur période féconde.

ESTIMATIONS OBTENUES

Les mesures suivantes peuvent être calculées pour tout groupe d'âge de femmes :

- La proportion observée de femmes ayant atteint la parité i , ${}_5M_x(i)$, est la proportion de femmes âgées de x à $x + 5$ qui ont eu i naissances ou plus. La proportion projetée correspondante,

$${}_5M_x^*(i)$$

est la proportion de femmes âgées de x à $x + 5$ dont on prévoit qu'elles auront eu i naissances ou plus à la fin de leur vie reproductive. Après la fin de la période féconde, pour $x \geq 50$, les proportions observées et attendues sont égales.

- Les probabilités d'agrandissement observées de rang i pour les femmes âgées de x à $x + 5$ sont notées

$${}_5a_x(i)$$

et représentent la proportion de femmes ayant au moins i enfants qui en ont eu au moins un de plus. Les probabilités d'agrandissement projetées sont notées

$${}_5a_x^*(i)$$

Une fois encore, pour $x \geq 50$, les probabilités observées et attendues sont égales.

Les identités suivantes rattachent ces mesures les unes aux autres et les relient à la distribution des femmes par parité.

Désignons par $N(i)$ le nombre de femmes dans la population de parité exacte i et par N le nombre total de femmes dans la population.

Désignons par $W(i)$ le nombre de femmes dans la population ayant atteint la parité i ou plus et par π la parité la plus élevée atteinte dans la population.

Le nombre de femmes qui ont atteint la parité i ou plus est obtenu en sommant les $N(i)$ à partir de i :

$$W(i) = \sum_{j=i}^{\pi} N(j) = N(i) + N(i + 1) + \dots + N(\pi)$$

de sorte que la proportion correspondante $M(i)$ est égale à

$$M(i) = \frac{1}{N} \cdot \sum_{j=i}^{\pi} N(j)$$

Le nombre de femmes de parité 0 ou plus est

$$W(0) = \sum_{j=0}^{\pi} N(j) = N$$

De sorte que la proportion à la parité zéro ou plus, $M(0) = N/N = 1$.

La parité moyenne dans l'ensemble de la population est

$$P = \frac{1}{N} \cdot \sum_{j=1}^{\pi} j \cdot N(j) = \frac{N(1)}{N} + \frac{2 \cdot N(2)}{N} + \frac{3 \cdot N(3)}{N} + \dots + \frac{\pi \cdot N(\pi)}{N} = \frac{1}{N} \cdot \sum_{j=1}^{\pi} W(j) = \sum_{j=1}^{\pi} M(j)$$

Pour les femmes qui ont achevé leur vie féconde, la parité moyenne est égale à la descendance finale (*completed family size, CFS*), l'équivalent de l'indice synthétique de fécondité dans une cohorte réelle :

$$CFS(x \text{ to } x + 5) = \sum_{j=1}^{\pi} {}_5M_x(j)$$

Pour $x \geq 50$.

Les probabilités d'agrandissement peuvent être écrites en termes de nombres ou de proportions :

$$a(i) = \frac{W(i + 1)}{W(i)} = \frac{W(i + 1) / N}{W(i) / N} = \frac{M(i + 1)}{M(i)}$$

Inversement, la proportion de femmes atteignant la parité $i + 1$ ou plus peut être exprimée en termes de probabilités d'agrandissement $a(i)$ pour les transitions aux rangs de naissance inférieurs :

$$M(i+1) = a(i) \cdot M(i) = a(i) \cdot a(i-1) \cdot M(i-2) \\ = a(i) \cdot a(i-1) \cdot a(i-2) \cdot \dots \cdot a(0) = \prod_{j=1}^i a(j)$$

où le symbole \prod désigne un produit.

DONNÉES NÉCESSAIRES ET HYPOTHÈSES

Tableau des données nécessaires

- Probabilités d'agrandissement observées
 - Parité par groupe d'âge des femmes âgées de 45-49 ans ou plus
- Probabilités d'agrandissement projetées
 - Parité par groupe d'âge des femmes âgées de 45-49 ans ou moins, et
 - Nombre d'enfants nés au cours de l'année précédant le recensement, classé par âge des mères (en groupes quinquennaux) et par nombre d'enfants déjà nés. Étant donné les hypothèses ci-dessous, une répartition exacte des naissances de la dernière année n'est pas requise. Il est recommandé d'utiliser les naissances de chacun des 12 mois révolus précédant la date du recensement.

Hypothèses importantes

- Les femmes ont eu une naissance au plus au cours de l'année écoulée.
- Les taux de fécondité par âge et rang vont continuer à s'appliquer à l'avenir. Cette hypothèse peut être irréaliste. Elle indique que les probabilités d'agrandissement projetées mesurent les conséquences de la combinaison entre la descendance atteinte par les femmes et les taux de fécondité actuels par âge et rang. Les probabilités d'agrandissement ne doivent pas être traitées comme des prévisions lorsque les cohortes sont encore susceptibles d'avoir encore un nombre important d'enfants additionnels du rang considéré ou lorsque la fécondité connaît une évolution rapide.

TRAVAIL PRÉPARATOIRE ET RECHERCHES PRÉLIMINAIRES

Avant d'entreprendre l'analyse des niveaux de fécondité par la méthode des probabilités d'agrandissement, on doit s'interroger sur la qualité des données dans les domaines suivants :

- structure par âge et sexe de la population ;
- naissances déclarées au cours de l'année écoulée ; et
- parités moyennes et nécessité d'une correction d'El-Badry.

PRÉCAUTIONS ET MISES EN GARDE

Si nécessaire, les parités doivent être celles estimées après application de la correction d'El-Badry quand des femmes sans enfants ont été considérées à tort comme n'ayant pas déclaré leur parité.

Si des pondérations liées au plan d'échantillonnage ont été fournies en même temps que les données, elles doivent être appliquées de manière adaptée au logiciel statistique utilisé pour préparer les tableaux nécessaires au modèle.

Les probabilités d'agrandissement projetées ne doivent être calculées qu'à partir des recensements ou d'autres bases de données incluant un grand nombre de femmes (au moins 10 000 dans chaque groupe d'âge), car le calcul de taux de fécondité par âge et rang fiables (nécessaires à l'établissement des probabilités d'agrandissement projetées) requiert des estimations robustes des taux pour chaque combinaison d'âge et de parité.

APPLICATION DE LA MÉTHODE

Le calcul des probabilités d'agrandissement passe par un certain nombre d'étapes, mais le principe qui guide le calcul n'est pas difficile à saisir. Les taux de fécondité par âge et rang qui sont calculés à l'aide des naissances de l'année écoulée servent à projeter la distribution de parité qu'on observerait dans chaque cohorte, si ces taux continuaient à s'appliquer pendant le reste de la période féconde de la cohorte. Ceci est fait en calculant par rang de naissance l'équivalent de l'indice synthétique de fécondité, c'est-à-dire le cumul sur l'ensemble des groupes d'âge des taux de fécondité pour les naissances de rang i . Ces taux servent à calculer les naissances supplémentaires de rang i qui viendraient s'ajouter dans chaque génération de femmes dans l'intervalle entre leur âge actuel et la fin de leur période reproductive. C'est une mesure synthétique du moment car elle cumule les taux d'une période récente. La proportion supplémentaire de femmes qui atteindraient la parité n ou plus est ajoutée à la proportion actuelle de femmes de rang i ou plus ; il en résulte la proportion finale de femmes de parité i ou plus. Ces distributions de parité projetées cumulées servent à calculer les probabilités d'agrandissement projetées, qui permettent l'analyse des tendances de la fécondité. En désignant par ${}_5P_x$ la parité moyenne des femmes âgées entre x et $x+5$ ans, la descendance finale projetée des femmes âgées de x à $x+5$ est donc

$${}_5PTFR_x = {}_5P_x + (TFR - \sum_{i=15,5}^x {}_5ASFR_i)$$

Rappel :

- PTFR = indice synthétique projeté
- P = parité moyenne (longitudinal)
- TFR = indice synthétique de fécondité (transversal)
- ASFR = taux de fécondité par âge

Étape 1 : Élaborer un tableau des enfants déjà nés par groupe d'âge de la mère pour chaque groupe d'âge où des données existent

Classer le nombre de femmes dans chaque groupe d'âge $[x, x + 5)$, ${}_5N_x$, par nombre d'enfants déjà nés, i , pour élaborer une série ${}_5N_x(i)$. Si nécessaire, les données doivent avoir été corrigées pour éliminer les parités peu vraisemblables. De même, une correction d'El-Badry doit avoir été appliquée.

Étape 2 : Élaborer un tableau du nombre de naissances au cours de l'année écoulée, par âge de la mère et parité

$${}_5B_x(i)$$

L'hypothèse à ce stade est que les femmes ont eu au maximum un enfant au cours des 12 derniers mois, et qu'elles n'ont eu ni grossesses ni accouchements multiples au cours de la période. Il s'ensuit l'hypothèse implicite qu'une naissance au cours de l'année passée issue d'une femme aujourd'hui de parité i est une naissance de rang i .

Étape 3 : Calculer les proportions de femmes ayant déjà atteint la parité i et les probabilités d'agrandissement pour chaque groupe d'âge

Le nombre de femmes âgées entre x et $x + 5$ qui ont donné naissance à i enfants ou plus est égal à

$${}_5W_x(i) = \sum_{j=i}^{\pi} {}_5N_x(j)$$

Les proportions de l'ensemble des femmes âgées de $[x, x + 5)$ qui ont eu i naissances ou plus sont calculées ainsi :

$${}_5M_x(i) = \frac{1}{{}_5N_x} \cdot \sum_{j=i}^{\pi} {}_5N_x(j)$$

Les proportions de femmes avec i naissances ou plus doivent être calculées pour toutes les cohortes, qu'elles aient ou non terminé leur vie féconde. Les femmes de parité zéro doivent être incluses dans le dénominateur, mais pas celles de parité inconnue (après application de la correction d'El-Badry, si nécessaire).

La probabilité d'agrandissement entre les parités i et $i + 1$ est alors définie comme

$${}_5a_x(i) = \frac{{}_5M_x(i+1)}{{}_5M_x(i)}$$

Le calcul des probabilités d'agrandissement pour les femmes qui ont terminé leur vie reproductive s'arrête ici. Les étapes suivantes décrivent comment calculer les probabilités d'agrandissement projetées pour les cohortes de femmes plus jeunes.

Étape 4 : Calculer les taux de fécondité par âge et rang

À partir des tableaux de naissances des 12 derniers mois par groupe d'âge et parité de la mère (étape 2), on calcule les taux de fécondité par âge et rang (en anglais, *Age-Order Specific Fertility Rates, AOSFR*) des femmes des groupes d'âge 15-19, ... 45-49,

$${}_5AOSFR_x(i) = \frac{{}_5B_x(i)}{{}_5N_x}$$

Cette fois encore, les femmes de parité non déclarée doivent être exclues du dénominateur. (Le dénominateur étant le nombre total de femmes dans le groupe d'âge (${}_5N_x$), et non une exposition au risque de connaître les événements du numérateur, les taux ne sont pas de véritables taux d'incidence, ou taux de 1^{re} catégorie).

Étape 5 : Calculer les taux de fécondité cumulés par rang

Les taux de fécondité par âge et rang cumulés (en anglais, *Total Order Fertility Rates, TOFR*) jusqu'à $x + 5$ ans pour le rang i sont calculés ainsi : ${}_5TOFR_x(i) = 5 \cdot \sum_{j=15,55}^x AOSFR_j(i)$.

Il en résulte que ${}_5TOFR_{45}(i) - {}_5TOFR_x(i)$ est une mesure de la proportion additionnelle de femmes qui atteindront la parité i entre $x + 5$ ans et la fin de la période reproductive, sous l'hypothèse que la fécondité future restera la même que la fécondité actuelle.

Étape 6 : Calculer la distribution par âge des taux de fécondité par rang et l'interpoler en âges conventionnels

Les taux cumulés calculés à l'étape précédente s'appliquent à 19,5 ans, 24,5 ans, etc. (le décalage d'une demi-année tenant compte du fait que le classement par âge de la mère a été établi à la date du recensement et non à la date de naissance de l'enfant). Par ailleurs, les proportions de femmes ayant atteint

chaque parité et les probabilités d'agrandissement s'appliquent (à peu près) au point médian de chaque groupe d'âge 17,5 ans, 22,5 ans, etc. C'est pourquoi les taux par âge et rang doivent être interpolés de sorte qu'ils s'appliquent à l'âge central de chaque groupe d'âge. L'interpolation est réalisée en exprimant les taux de fécondité cumulés par âge et rang comme des proportions des indices synthétiques par rang et en pratiquant une interpolation linéaire sur les gompits des cumulants.

Pour chaque rang de naissance, il faut calculer la proportion de l'indice synthétique par rang réalisée avant d'atteindre la limite supérieure du groupe d'âge. Puisque nous avons utilisé des groupes d'âge décalés dans notre exemple, les limites supérieures sont abaissées d'une demi-année. Les proportions sont calculées en faisant la somme des taux jusqu'à la limite supérieure du groupe d'âge, et en exprimant cette somme en proportion de l'indice synthétique, c'est-à-dire

$${}_5\theta_x(i) = \frac{{}_5TOFR_x(i)}{{}_5TOFR_{45}(i)}$$

Comme on l'a déjà dit, ces proportions s'appliquent à 19,5 ans, 24,5 ans, etc. À cause de la courbure du schéma de fécondité, une interpolation linéaire directe pour obtenir les proportions à 22,5 ans, 27,5 ans, etc. n'est pas appropriée. Mais une transformation gompit (double log négatif) fait d'une fonction sigmoïde une droite. Les gompits des proportions sont donc calculés, puis interpolés, avant d'être ramenés à leur échelle originelle par recours aux anti-gompits. Ainsi, en utilisant l'astérisque pour désigner le passage aux points médians conventionnels

$${}_5\theta_x^*(i) = \exp \left[-\exp \left(\frac{-\left(0.4 \cdot \left(-\ln(-\ln({}_5\theta_{x-5}(i))) \right) \right)}{+0.6 \cdot \left(-\ln(-\ln({}_5\theta_x(i))) \right)} \right) \right] \quad (1)$$

Ceci vaut pour $x = 20, 25, \dots, 40$.

Pour éviter d'introduire une composante projetée trop importante dans les probabilités d'agrandissement projetées, nous ne tenons pas compte dans nos calculs des estimations de ${}_5\theta_x^*(i)$ qui indiquent des proportions de fécondité par rang atteintes au point médian d'un groupe d'âge inférieures à 0,3. Nous traitons comme un peu hasardeux les calculs reposant sur des valeurs de ${}_5\theta_x^*(i) < 0,5$.

Étape 7 : Calculer l'accroissement de parité attendu

La proportion additionnelle des femmes atteignant chaque parité dans l'intervalle entre le point médian du groupe d'âge et la fin de la période féconde est égale à

$${}_5TOFR_{45}(i) - {}_5TOFR_x(i) = {}_5TOFR_{45}(i) \cdot (1 - {}_5\theta_x^*(i))$$

Étape 8 : Calculer les probabilités d'agrandissement cumulées projetées et les probabilités d'agrandissement projetées

Les probabilités d'agrandissement cumulées projetées sont calculées en ajoutant l'accroissement de parité futur obtenu à la précédente étape aux probabilités d'agrandissement cumulées actuelles obtenues à l'étape 1. La proportion de femmes âgées de x à $x + 5$ ans dont on prévoit qu'elles atteindront au moins la parité i est égale à

$${}_5M_x^*(i) = {}_5M_x(i) + {}_5TOFR_{45}(i) \cdot (1 - {}_5\theta_x^*(i))$$

Et la probabilité d'agrandissement projetée entre les parités i et $i + 1$ est égale à

$${}_5a_x^*(i) = \frac{{}_5M_x^*(i + 1)}{{}_5M_x^*(i)}$$

INTERPRÉTATION DES RÉSULTATS

L'interprétation des variations des probabilités d'agrandissement peut être compliquée par des erreurs dans les données et doit tenir compte des hypothèses qui sous-tendent les calculs. S'il y a des erreurs dans les données, il est prudent de ne pas sur-interpréter les résultats. Les probabilités d'agrandissement qui incluent une proportion substantielle de composante projetée doivent être interprétées avec une prudence redoublée, car elles risquent de refléter des effets de période qui cesseront à l'avenir. Si par exemple les taux de fécondité par âge et rang reflètent un retard des naissances dans la période où les données sont recueillies, et que la fécondité s'accroît ensuite par compensation, de sorte que les femmes retrouvent les parités atteintes par les cohortes antérieures, les probabilités d'agrandissement projetées seront trompeuses. De même, un déclin futur de la fécondité réduira les parités finales des femmes même si les taux actuels prédisent une stabilité. Les probabilités d'agrandissement projetées pour les femmes âgées sont plus fiables car elles incluent une part plus faible de fécondité projetée.

La probabilité projetée de premier rang, ${}_5M_x^*(1)$, est souvent biaisée du fait d'erreurs sur la période de référence qui ne sont pas compensées comme elles le sont pour les probabilités des autres rangs. Les proportions projetées de femmes ayant au moins un enfant sont affectées par les modifications de l'âge à la première naissance. La mesure étant

prise dans une cohorte fictive à partir des taux du moment, des variations dans le calendrier des premières naissances affectent aussi le taux cumulé, qui peut prendre une valeur extérieure à l'intervalle des valeurs attendues pour la plupart des cohortes réelles dans les pays en développement. Du fait de fluctuations de court terme et d'effets de calendrier, la probabilité de premier rang (du moment) peut éventuellement dépasser l'unité. Ceci ne doit pas être confondu avec, ou interprété comme, une probabilité d'agrandissement supérieure à 1, ce qui est par définition impossible.

La probabilité de premier rang projetée peut aussi donner une indication sur la qualité des données. Les taux indiquent la proportion de femmes dont on prévoit qu'elles deviendront mères au cours de leur vie reproductive, dans une cohorte fictive soumise à ces taux. Des erreurs sur la période de référence peuvent se traduire par une proportion invraisemblable de femmes dont on prévoit qu'elles deviendront mères. Dans des cas extrêmes d'erreurs, la proportion peut dépasser éventuellement l'unité. Des taux cumulés inférieurs à 0,9 laissent présumer des erreurs dans les données ou des variations importantes dans le calendrier des premières naissances, mais on a pu observer de telles valeurs dans des populations affectées par des niveaux élevés de stérilité primaire et dans des populations actuelles où l'infécondité volontaire est fréquente. On a souvent observé une hausse de la proportion (projetée) de femmes dont on s'attend à ce qu'elles deviennent mères, ${}_5PPR_x^{(*)}(0,1)$, dans les populations qui ont connu un recul de la stérilité primaire.

Les probabilités d'agrandissement de rang plus élevé peuvent être sujettes à de fortes fluctuations dues à des erreurs d'échantillonnage dans les taux de fécondité par âge et rang, ceux-ci étant calculés sur des nombres de naissances relativement faibles. Quand d'importantes erreurs d'échantillonnage sont dues aux faibles nombres de femmes, on peut envisager de faire la moyenne de probabilités d'agrandissement projetées contiguës – deux ou plus – afin de percevoir les tendances plus clairement.

L'examen de l'évolution des probabilités d'agrandissement au fil des cohortes (projetées pour les cohortes de femmes âgées de 40-44 ans et moins et effectives pour les cohortes de 45-49 ans et plus) peut éclairer les tendances de la fécondité. Le recul d'une probabilité d'agrandissement (projetée) indique un recul de la proportion de femmes ayant une naissance supplémentaire et, toutes choses égales par ailleurs, une baisse de la fécondité. Des variations des probabilités d'agrandissement à différentes parités peuvent

tenir à divers facteurs, certaines probabilités étant affectées davantage que d'autres. Identifier celles qui varient peut être utile pour comprendre les modifications de la fécondité dans la population. Par exemple, la diffusion de la limitation des naissances par une contraception liée à la parité se traduira par un recul des probabilités d'agrandissement à des parités moyennes (disons, 3 à 6). Une variation des probabilités d'agrandissement à des rangs plus élevés reflète souvent des variations de fécondité indépendantes de la parité, par exemple des modifications de la durée d'allaitement.

EXEMPLE

Nous analysons les données du recensement de 1998 au Cambodge réalisé le 3 mars 1998. La méthode est présentée dans le fichier Excel associé disponibles à <http://demographi-cestimation.iussp.org/fr>.

Étape 1 : Élaborer un tableau des enfants déjà nés par groupe d'âge de la mère pour chaque groupe d'âge où des données existent

Les données du recensement, non corrigées, figurent au tableau 8.1.

Les données suivent un schéma classique d'augmentation des parités moyennes avec l'âge jusqu'à 45-49 ans. Par la suite, les parités moyennes n'augmentent que légèrement. Les données indiquent donc que les femmes de plus de 50 ans ont omis de déclarer certains de leurs enfants déjà nés ou que la fécondité a été à peu près constante parmi les femmes ayant terminé leur vie féconde dans les 10-15 dernières années.

Étape 2 : Élaborer un tableau du nombre de naissances au cours de l'année écoulée, par âge de la mère et parité

$${}_5B_x(i)$$

Les données du recensement de 1998 au Cambodge sur le nombre d'enfants nés au cours de l'année précédant le recensement, par groupe d'âge de la mère et parité, figurent au tableau 8.2.

D'après les données, l'indice synthétique de fécondité est de l'ordre de 3,25 enfants par femmes, soit nettement moins que la descendance finale (5,566) de la génération des femmes âgées de 45-49 ans. Ceci suggère soit une baisse rapide de la fécondité encours dans le pays, soit une omission systématique des naissances récentes dans le recensement.

TABLEAU 8.1 Nombre de femmes par groupe d'âge et parité, Cambodge, recensement de 1998

| Parité (i) | Groupe d'âge | | | | | | | | | |
|----------------|--------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| | 15-19 | 20-24 | 25-29 | 30-34 | 35-39 | 40-44 | 45-49 | 50-54 | 55-59 | 60-64 |
| 0 | 635 940 | 200 620 | 105 350 | 55 300 | 33 900 | 22 970 | 16 430 | 9 790 | 7 640 | 6 880 |
| 1 | 30 340 | 94 810 | 79 570 | 38 780 | 21 150 | 16 570 | 13 460 | 9 600 | 8 020 | 7 450 |
| 2 | 5 980 | 62 580 | 113 640 | 66 750 | 33 320 | 21 920 | 17 220 | 12 840 | 9 690 | 8 020 |
| 3 | 1 650 | 21 680 | 88 750 | 75 660 | 45 680 | 27 750 | 20 250 | 15 170 | 11 180 | 8 120 |
| 4 | 1 200 | 6 890 | 44 480 | 72 370 | 56 100 | 35 450 | 24 530 | 18 090 | 13 350 | 10 160 |
| 5 | 630 | 2 060 | 16 580 | 51 600 | 57 050 | 39 100 | 26 660 | 19 660 | 16 100 | 12 040 |
| 6 | 0 | 1 120 | 5 630 | 30 250 | 50 960 | 39 180 | 27 540 | 19 860 | 16 180 | 12 250 |
| 7 | 0 | 510 | 2 110 | 13 390 | 35 780 | 33 190 | 25 850 | 18 880 | 16 170 | 13 690 |
| 8 | 0 | 480 | 1 070 | 5 970 | 20 580 | 24 920 | 23 180 | 16 940 | 14 830 | 11 970 |
| 9 | 0 | 0 | 440 | 1 820 | 9 590 | 15 860 | 16 130 | 12 680 | 11 360 | 9 200 |
| 10 | 0 | 0 | 240 | 880 | 4 840 | 10 420 | 12 790 | 10 720 | 9 800 | 7 750 |
| 11 | 0 | 0 | 120 | 320 | 1 760 | 4 430 | 6 670 | 5 370 | 4 850 | 3 970 |
| 12 | 0 | 0 | 80 | 150 | 760 | 3 220 | 5 350 | 4 810 | 4 780 | 4 230 |
| 13 | 0 | 0 | 40 | 60 | 250 | 990 | 1 960 | 1 790 | 1 790 | 1 440 |
| 14 | 0 | 0 | 0 | 40 | 170 | 620 | 1 020 | 970 | 760 | 540 |
| 15 | 0 | 0 | 0 | 10 | 10 | 250 | 510 | 460 | 440 | 490 |
| 16 | 0 | 0 | 0 | 10 | 30 | 130 | 230 | 160 | 140 | 210 |
| 17 | 0 | 0 | 0 | 0 | 10 | 40 | 80 | 50 | 70 | 70 |
| 18 | 0 | 0 | 0 | 10 | 10 | 30 | 30 | 30 | 40 | 40 |
| TOTAL | 675 740 | 390 750 | 458 100 | 413 370 | 371 950 | 297 040 | 239 890 | 177 870 | 147 190 | 118 520 |
| Parité moyenne | 0,082 | 0,862 | 1,965 | 3,150 | 4,369 | 5,086 | 5,566 | 5,758 | 5,974 | 5,969 |

TABLEAU 8.2 Naissances déclarées au cours de l'année précédant le recensement par groupe d'âge de la mère et parité, Cambodge, recensement de 1998

| Parité (i) | Groupe d'âge | | | | | | |
|---------------------------|--------------|--------|--------|--------|-------|-------|-------|
| | 15-19 | 20-24 | 25-29 | 30-34 | 35-39 | 40-44 | 45-49 |
| 0 | | | | | | | |
| 1 | 13 100 | 23 490 | 11 500 | 3 170 | 1 040 | 320 | 190 |
| 2 | 2 340 | 19 180 | 21 320 | 7 830 | 2 010 | 730 | 150 |
| 3 | 410 | 8 370 | 22 640 | 11 470 | 3 790 | 850 | 230 |
| 4 | 210 | 2 430 | 14 330 | 14 110 | 5 200 | 1 450 | 300 |
| 5 | 100 | 600 | 6 220 | 12 440 | 6 740 | 1 830 | 480 |
| 6 | 0 | 280 | 2 020 | 8 360 | 8 120 | 2 870 | 560 |
| 7 | 0 | 90 | 750 | 4 430 | 7 220 | 3 240 | 840 |
| 8 | 0 | 90 | 360 | 2 100 | 4 930 | 3 020 | 750 |
| 9 | 0 | 0 | 90 | 640 | 2 800 | 2 310 | 690 |
| 10 | 0 | 0 | 40 | 300 | 1 390 | 1 570 | 640 |
| 11 | 0 | 0 | 30 | 160 | 630 | 810 | 490 |
| 12 | 0 | 0 | 10 | 40 | 200 | 480 | 320 |
| 13 | 0 | 0 | 10 | 20 | 80 | 250 | 140 |
| 14 | | 0 | 0 | 10 | 60 | 100 | 70 |
| 15 | | | 0 | 0 | 0 | 50 | 30 |
| 16 | | | | 0 | 10 | 10 | 30 |
| 17 | | | | | 0 | 0 | 10 |
| 18 | | | | 0 | 0 | 0 | 10 |
| Taux de fécondité par âge | 0,024 | 0,140 | 0,173 | 0,157 | 0,119 | 0,067 | 0,025 |

Étape 3 : Calculer les proportions de femmes ayant déjà atteint la parité i et les probabilités d'agrandissement pour chaque groupe d'âge

Les proportions de femmes atteignant la parité i ou plus sont calculées à partir des données du tableau 8.1 et figurent au tableau 8.3. Les parités supérieures ou égales à 12 sont regroupées.

La proportion des femmes ayant eu i naissances ou plus figure au tableau 8.3. Ainsi, alors que 80,4 % des femmes âgées de 45-49 ans ont eu au moins trois naissances, elles ne sont que 71,9 % à en avoir eu quatre ou plus. Les probabilités d'agrandissement qui en découlent, ${}_5a_x(i)$, figurent au tableau 8.4.

Environ 93,2 % des femmes âgées de 45-49 ans au Cambodge en 1998 ont eu au moins un enfant et 89,5 % des mères de trois enfants dans le même groupe d'âge en ont ensuite eu un quatrième.

Étape 4 : Calculer les taux de fécondité par âge et rang

Les naissances déclarées au cours de l'année passée (par âge et parité) du tableau 8.2 sont divisées par le nombre total de femmes dans chaque groupe d'âge (les totaux par groupe d'âge du tableau 8.1). Les taux de fécondité par âge et rang qui en résultent figurent au tableau 8.5. Par exemple ${}_5AOSFR_{25}(1) = 11\,500/458\,100 = 0,0251$.

Étape 5 : Calculer les taux de fécondité cumulés par rang

À partir des données présentées au tableau 8.5, on obtient les taux de fécondité par âge et rang cumulés en sommant sur chaque ligne jusqu'au groupe d'âge désiré. Par exemple, le taux par âge et rang cumulé jusqu'à la fin du groupe des 25-29 ans pour la parité 3 est égale à $5 \times (0,0006 + 0,0214 + 0,0494) = 0,3572$.

Étape 6 : Calculer la distribution par âge des taux de fécondité par rang et l'interpoler en âges conventionnels

À partir des résultats de l'étape 5, la proportion de fécondité par rang atteinte à l'issue de chaque groupe d'âge est obtenue en divisant les taux de fécondité par rang cumulés par l'indice synthétique de fécondité par rang qui figure dans la dernière colonne du tableau 8.5. Les résultats sont au tableau 8.6.

La proportion de la fécondité atteinte à 49,5 ans est supposée égale à un. Avec le même exemple qu'à l'étape précédente, la proportion de fécondité de troisième rang atteinte à 29,5 ans est $0,3572/0,5660=0,6311$.

Les âges des mères étant ceux au moment du recensement, et les parités moyennes calculées à l'étape 1 s'appliquant à 17,5 ans, 22,5 ans etc., nous devons rendre les deux séries de données comparables en interpolant entre deux valeurs pour avoir des proportions à 22,5 ans, 27, 5 ans

TABLEAU 8.3 Proportion de femmes ayant atteint chaque parité par groupe d'âge, Cambodge, recensement de 1998

| Parité (i) | Groupe d'âge | | | | | | | | | |
|----------------|--------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| | 15-19 | 20-24 | 25-29 | 30-34 | 35-39 | 40-44 | 45-49 | 50-54 | 55-59 | 60-64 |
| 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 1 | 0,0589 | 0,4866 | 0,7700 | 0,8662 | 0,9089 | 0,9227 | 0,9315 | 0,9450 | 0,9481 | 0,9420 |
| 2 | 0,0140 | 0,2439 | 0,5963 | 0,7724 | 0,8520 | 0,8669 | 0,8754 | 0,8910 | 0,8936 | 0,8791 |
| 3 | 0,0051 | 0,0838 | 0,3483 | 0,6109 | 0,7624 | 0,7931 | 0,8036 | 0,8188 | 0,8278 | 0,8114 |
| 4 | 0,0027 | 0,0283 | 0,1545 | 0,4279 | 0,6396 | 0,6997 | 0,7192 | 0,7335 | 0,7518 | 0,7429 |
| 5 | 0,0009 | 0,0107 | 0,0574 | 0,2528 | 0,4888 | 0,5803 | 0,6169 | 0,6318 | 0,6611 | 0,6572 |
| 6 | 0,0000 | 0,0054 | 0,0212 | 0,1280 | 0,3354 | 0,4487 | 0,5058 | 0,5213 | 0,5517 | 0,5556 |
| 7 | 0,0000 | 0,0025 | 0,0090 | 0,0548 | 0,1984 | 0,3168 | 0,3910 | 0,4096 | 0,4418 | 0,4522 |
| 8 | 0,0000 | 0,0012 | 0,0043 | 0,0224 | 0,1022 | 0,2051 | 0,2833 | 0,3035 | 0,3320 | 0,3367 |
| 9 | 0,0000 | 0,0000 | 0,0020 | 0,0080 | 0,0469 | 0,1212 | 0,1866 | 0,2082 | 0,2312 | 0,2357 |
| 10 | 0,0000 | 0,0000 | 0,0010 | 0,0036 | 0,0211 | 0,0678 | 0,1194 | 0,1370 | 0,1540 | 0,1581 |
| 11 | 0,0000 | 0,0000 | 0,0005 | 0,0015 | 0,0081 | 0,0327 | 0,0661 | 0,0767 | 0,0874 | 0,0927 |
| 12+ | 0,0000 | 0,0000 | 0,0003 | 0,0007 | 0,0033 | 0,0178 | 0,0383 | 0,0465 | 0,0545 | 0,0592 |

etc. la formule est donnée par l'équation 1 et les résultats figurent au tableau 8.7.

Étant donné l'incertitude entourant la projection des probabilités d'agrandissement futures, les cellules du tableau 8.7 dont les valeurs sont inférieures à 0,3 (c'est-à-dire où moins de 30 % de la fécondité par rang a été atteinte

à l'âge indiqué) sont indiquées en gris et seront exclues des calculs ultérieurs. Les valeurs supérieures à 0,3 mais inférieures à 0,5 sont indiquées en italique et doivent être traitées avec circonspection dans les calculs ultérieurs. Les calculs reposant sur ces valeurs seront présentés entre parenthèses dans les tableaux suivants.

TABLEAU 8.4 Probabilités d'agrandissement observées par groupe d'âge de la mère et parité, Cambodge, recensement de 1998

| Parité (i) | Groupe d'âge | | | | | | | | | |
|------------|--------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| | 15-19 | 20-24 | 25-29 | 30-34 | 35-39 | 40-44 | 45-49 | 50-54 | 55-59 | 60-64 |
| 0 | 0,0589 | 0,4866 | 0,7700 | 0,8662 | 0,9089 | 0,9227 | 0,9315 | 0,9450 | 0,9481 | 0,9420 |
| 1 | 0,2377 | 0,5013 | 0,7744 | 0,8917 | 0,9374 | 0,9395 | 0,9398 | 0,9429 | 0,9425 | 0,9333 |
| 2 | 0,3679 | 0,3435 | 0,5840 | 0,7909 | 0,8949 | 0,9149 | 0,9180 | 0,9190 | 0,9263 | 0,9230 |
| 3 | 0,5259 | 0,3378 | 0,4437 | 0,7004 | 0,8389 | 0,8822 | 0,8950 | 0,8958 | 0,9082 | 0,9156 |
| 4 | 0,3443 | 0,3770 | 0,3717 | 0,5909 | 0,7642 | 0,8294 | 0,8578 | 0,8613 | 0,8794 | 0,8846 |
| 5 | 0,0000 | 0,5060 | 0,3698 | 0,5063 | 0,6862 | 0,7732 | 0,8199 | 0,8251 | 0,8345 | 0,8454 |
| 6 | | 0,4692 | 0,4214 | 0,4283 | 0,5915 | 0,7060 | 0,7730 | 0,7858 | 0,8008 | 0,8140 |
| 7 | | 0,4848 | 0,4854 | 0,4091 | 0,5151 | 0,6473 | 0,7244 | 0,7409 | 0,7513 | 0,7446 |
| 8 | | 0,0000 | 0,4623 | 0,3560 | 0,4586 | 0,5909 | 0,6589 | 0,6862 | 0,6965 | 0,7001 |
| 9 | | | 0,5217 | 0,4485 | 0,4498 | 0,5593 | 0,6397 | 0,6577 | 0,6662 | 0,6707 |
| 10 | | | 0,5000 | 0,4054 | 0,3827 | 0,4824 | 0,5534 | 0,5599 | 0,5677 | 0,5864 |
| 11 | | | 0,5000 | 0,4667 | 0,4133 | 0,5438 | 0,5792 | 0,6063 | 0,6232 | 0,6388 |

TABLEAU 8.5 Taux de fécondité par âge et rang et indice synthétique de fécondité par rang (tofr), cambodge, recensement de 1998

| Groupe d'âge | | | | | | | | |
|------------------------------|--------|--------|--------|--------|--------|--------|--------|------------------|
| Parité (<i>i</i>) | 15-19 | 20-24 | 25-29 | 30-34 | 35-39 | 40-44 | 45-49 | TOFR(<i>i</i>) |
| 0 | | | | | | | | |
| 1 | 0,0194 | 0,0601 | 0,0251 | 0,0077 | 0,0028 | 0,0011 | 0,0008 | 0,5847 |
| 2 | 0,0035 | 0,0491 | 0,0465 | 0,0189 | 0,0054 | 0,0025 | 0,0006 | 0,6326 |
| 3 | 0,0006 | 0,0214 | 0,0494 | 0,0277 | 0,0102 | 0,0029 | 0,0010 | 0,5660 |
| 4 | 0,0003 | 0,0062 | 0,0313 | 0,0341 | 0,0140 | 0,0049 | 0,0013 | 0,4603 |
| 5 | 0,0001 | 0,0015 | 0,0136 | 0,0301 | 0,0181 | 0,0062 | 0,0020 | 0,3582 |
| 6 | 0,0000 | 0,0007 | 0,0044 | 0,0202 | 0,0218 | 0,0097 | 0,0023 | 0,2959 |
| 7 | 0,0000 | 0,0002 | 0,0016 | 0,0107 | 0,0194 | 0,0109 | 0,0035 | 0,2320 |
| 8 | 0,0000 | 0,0002 | 0,0008 | 0,0051 | 0,0133 | 0,0102 | 0,0031 | 0,1632 |
| 9 | 0,0000 | 0,0000 | 0,0002 | 0,0015 | 0,0075 | 0,0078 | 0,0029 | 0,0996 |
| 10 | 0,0000 | 0,0000 | 0,0001 | 0,0007 | 0,0037 | 0,0053 | 0,0027 | 0,0625 |
| 11 | 0,0000 | 0,0000 | 0,0001 | 0,0004 | 0,0017 | 0,0027 | 0,0020 | 0,0346 |
| 12+ | 0,0000 | 0,0000 | 0,0000 | 0,0002 | 0,0009 | 0,0030 | 0,0025 | 0,0335 |
| Taux de fécondité par âge | 0,0239 | 0,1396 | 0,1731 | 0,1574 | 0,1189 | 0,0670 | 0,0247 | 3,5231 |

TABLEAU 8.6 Proportion de la fécondité par rang atteinte à l'issue de chaque groupe d'âge, Cambodge, recensement de 1998

| Parité (i) | Groupe d'âge | | | | | | |
|------------|--------------|--------|--------|--------|--------|--------|--------|
| | 19,5 | 24,5 | 29,5 | 34,5 | 39,5 | 44,5 | 49,5 |
| 0 | | | | | | | |
| 1 | 0,1658 | 0,6799 | 0,8945 | 0,9601 | 0,9840 | 0,9932 | 1,0000 |
| 2 | 0,0274 | 0,4153 | 0,7832 | 0,9329 | 0,9756 | 0,9951 | 1,0000 |
| 3 | 0,0054 | 0,1946 | 0,6311 | 0,8762 | 0,9663 | 0,9915 | 1,0000 |
| 4 | 0,0034 | 0,0709 | 0,4107 | 0,7815 | 0,9334 | 0,9864 | 1,0000 |
| 5 | 0,0021 | 0,0235 | 0,2130 | 0,6331 | 0,8861 | 0,9721 | 1,0000 |
| 6 | 0,0000 | 0,0121 | 0,0866 | 0,4284 | 0,7973 | 0,9606 | 1,0000 |
| 7 | 0,0000 | 0,0050 | 0,0402 | 0,2712 | 0,6895 | 0,9245 | 1,0000 |
| 8 | 0,0000 | 0,0071 | 0,0311 | 0,1868 | 0,5928 | 0,9042 | 1,0000 |
| 9 | 0,0000 | 0,0000 | 0,0099 | 0,0876 | 0,4654 | 0,8556 | 1,0000 |
| 10 | 0,0000 | 0,0000 | 0,0070 | 0,0650 | 0,3639 | 0,7866 | 1,0000 |
| 11 | 0,0000 | 0,0000 | 0,0095 | 0,0654 | 0,3103 | 0,7046 | 1,0000 |
| 12+ | 0,0000 | 0,0000 | 0,0065 | 0,0318 | 0,1724 | 0,6201 | 1,0000 |

TABLEAU 8.7 Proportion de fécondité par rang atteinte au point médian de chaque intervalle d'âge, Cambodge, recensement de 1998

| Parité (i) | Groupe d'âge | | | | | |
|------------|--------------|--------|--------|--------|--------|--------|
| | 22,5 | 27,5 | 32,5 | 37,5 | 42,5 | 47,5 |
| 0 | | | | | | |
| 1 | 0,4897 | 0,8326 | 0,9409 | 0,9769 | 0,9904 | 0,9999 |
| 2 | 0,2135 | 0,6652 | 0,8915 | 0,9634 | 0,9906 | 0,9999 |
| 3 | 0,0739 | 0,4655 | 0,8044 | 0,9428 | 0,9852 | 0,9999 |
| 4 | 0,0275 | 0,2526 | 0,6624 | 0,8916 | 0,9742 | 0,9998 |
| 5 | 0,0102 | 0,1104 | 0,4751 | 0,8139 | 0,9506 | 0,9998 |
| 6 | | 0,0452 | 0,2738 | 0,6811 | 0,9228 | 0,9997 |
| 7 | | 0,0197 | 0,1539 | 0,5410 | 0,8640 | 0,9996 |
| 8 | | 0,0183 | 0,1061 | 0,4345 | 0,8232 | 0,9996 |
| 9 | | | 0,0430 | 0,2965 | 0,7449 | 0,9995 |
| 10 | | | 0,0311 | 0,2221 | 0,6527 | 0,9994 |
| 11 | | | 0,0341 | 0,1937 | 0,5671 | 0,9993 |
| 12+ | | | 0,0181 | 0,1001 | 0,4473 | 0,9993 |

Étape 7 : Calculer l'accroissement de parité attendu

La proportion additionnelle des femmes atteignant chaque parité dans l'intervalle entre le point médian du groupe d'âge et la fin de la période féconde est donnée par

$${}_5TOFR_{45}(i) - {}_5TOFR_x(i) = {}_5TOFR_{45}(i) \cdot (1 - \theta_x^*(i))$$

comme le montre le tableau 8.8.

Ainsi, l'accroissement futur de fécondité de rang 3 prévu après 30 ans est égal à $0,5660 \times (1 - 0,4655) = 0,3025$. Les valeurs entre parenthèses sont moins fiables, étant fondées sur des proportions de fécondité par âge et rang atteintes au point médian de ce groupe d'âge inférieures à 0,50.

Étape 8 : Calculer les probabilités d'agrandissement cumulées projetées et les probabilités d'agrandissement projetées

Les probabilités d'agrandissement cumulées projetées sont obtenues en additionnant les accroissements futurs de fécondité par rang présentés au tableau 8.8 à la proportion de femmes ayant atteint chacune des parités telle que calculée à l'étape 3 et présentée au tableau 8.3.

Les probabilités d'agrandissement projetées pour les cohortes encore en âge d'avoir des enfants peuvent maintenant être comparées aux probabilités d'agrandissement finales des femmes âgées de 50 ans ou plus. Les résultats figurent au tableau 8.10.

Dans le groupe d'âge 25-29 ans, la proportion projetée avec une parité 3 et plus est de 0,3483 (cf. tableau 8.3) + 0,3025 (cf. tableau 8.8) = 0,6508. Les valeurs entre parenthèses sont moins fiables, étant fondées sur des proportions de fécondité par âge et rang atteintes au point médian de ce groupe d'âge inférieures à 0,50.

TABLEAU 8.8 Proportion additionnelle de femmes atteignant la parité *i* d'ici la fin de la période féconde par groupe d'âge de la mère, Cambodge, recensement de 1998

| Parité (<i>i</i>) | Groupe d'âge | | | | |
|---------------------|--------------|----------|----------|----------|----------|
| | 22,5 | 27,5 | 32,5 | 37,5 | 42,5 |
| 0 | | | | | |
| 1 | (0,2984) | 0,0979 | 0,0346 | 0,0135 | 0,0056 |
| 2 | | 0,2118 | 0,0686 | 0,0232 | 0,0059 |
| 3 | | (0,3025) | 0,1107 | 0,0324 | 0,0084 |
| 4 | | | 0,1554 | 0,0499 | 0,0119 |
| 5 | | | (0,1880) | 0,0666 | 0,0177 |
| 6 | | | | 0,0944 | 0,0228 |
| 7 | | | | 0,1065 | 0,0316 |
| 8 | | | | (0,0923) | 0,0289 |
| 9 | | | | | 0,0254 |
| 10 | | | | | 0,0217 |
| 11 | | | | | 0,0150 |
| 12+ | | | | | (0,0185) |

Les probabilités d'agrandissement projetées sont calculées en divisant les proportions devant atteindre chaque parité successive dans une cohorte d'âge (Tableau 8.10).

PROBABILITÉS D'AGRANDISSEMENT PROJETÉES

La probabilité d'agrandissement de la parité 0 à la parité 1 indique la proportion de femmes ayant eu au moins un enfant. La probabilité d'agrandissement projetée souffre d'un fort biais, comme nous le verrons plus loin. La proportion de femmes sans enfants est restée à peu près constante au fil des cohortes. Mais on voit à la figure 8.1 que les probabilités d'agrandissement aux parités 2 et suivantes reculent sensiblement dans les cohortes successives.

Les proportions projetées de femmes qui devraient atteindre chaque parité d'après le recensement de 1998 au Cambodge sont représentées à la figure 8.2.

Les conséquences de probabilités d'agrandissement réduites dans les cohortes jeunes apparaissent sur les proportions de femmes qui devraient atteindre chaque parité. Alors que 55 % des femmes âgées de 60-64 ans ont eu 6 enfants ou plus, cette proportion dépasse à peine la moitié (50,6 %) dans la cohorte dont la période féconde vient de se terminer (femmes âgées de 45-49 ans).

On voit ici les prémices d'une baisse de la fécondité qui a débuté lorsque la cohorte aujourd'hui âgée de 50-54 ans a eu ses enfants, soit environ 25-30 ans avant le recensement.

TABLEAU 8.9 Proportions projetées devant atteindre chaque parité, Cambodge, recensement de 1998

| Parité (<i>i</i>) | Groupe d'âge | | | | | | | | |
|---------------------|--------------|----------|----------|----------|----------|--------|--------|--------|--------|
| | 20-24 | 25-29 | 30-34 | 35-39 | 40-44 | 45-49 | 50-54 | 55-59 | 60-64 |
| 0 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 | 1,0000 |
| 1 | (0,7850) | 0,8679 | 0,9008 | 0,9224 | 0,9283 | 0,9315 | 0,9450 | 0,9481 | 0,9420 |
| 2 | | 0,8081 | 0,8410 | 0,8752 | 0,8728 | 0,8754 | 0,8910 | 0,8936 | 0,8791 |
| 3 | | (0,6508) | 0,7216 | 0,7948 | 0,8014 | 0,8036 | 0,8188 | 0,8278 | 0,8114 |
| 4 | | | 0,5833 | 0,6895 | 0,7115 | 0,7192 | 0,7335 | 0,7518 | 0,7429 |
| 5 | | | (0,4408) | 0,5554 | 0,5980 | 0,6169 | 0,6318 | 0,6611 | 0,6572 |
| 6 | | | | 0,4298 | 0,4715 | 0,5058 | 0,5213 | 0,5517 | 0,5556 |
| 7 | | | | 0,3049 | 0,3483 | 0,3910 | 0,4096 | 0,4418 | 0,4522 |
| 8 | | | | (0,1945) | 0,2339 | 0,2833 | 0,3035 | 0,3320 | 0,3367 |
| 9 | | | | | 0,1466 | 0,1866 | 0,2082 | 0,2312 | 0,2357 |
| 10 | | | | | 0,0895 | 0,1194 | 0,1370 | 0,1540 | 0,1581 |
| 11 | | | | | 0,0477 | 0,0661 | 0,0767 | 0,0874 | 0,0927 |
| 12+ | | | | | (0,0363) | 0,0383 | 0,0465 | 0,0545 | 0,0592 |

Première maternité

D’après les données cambodgiennes, la proportion des femmes de 45-49 ans ayant eu au moins un enfant était de 0,9315. Les proportions projetées, classées selon l’âge actuel de la mère, figurent à la première ligne du tableau 8.10.

Il y a certainement eu d’importants changements dans le calendrier de la première maternité dans les cohortes jeunes, mais les résultats pour les femmes entre 20 et 30 ans ainsi que, très probablement, pour celles âgées de 30-34 ans, sont invraisemblables pour des raisons déjà évoquées dans la présentation de leur mode de calcul. Les probabilités d’agrandissement projetées pour les femmes entre 20 et 30 ans sont fortement affectées à la fois par des modifications dans le calendrier de la fécondité et par l’importance de la composante projetée. Ces résultats ne doivent pas être pris en compte ; une mesure de la première maternité par la méthode des tables de survie doit leur être préférée. Mais ces méthodes et leurs applications débordent des objectifs de ce manuel.

À l’exception de cet indice, le calcul des probabilités d’agrandissement projetées doit éliminer l’effet des erreurs sur la période de référence dans la mesure où celles-ci sont indépendantes de la parité. On le voit dans les estimations des probabilités d’agrandissement projetées pour les autres parités.

Les données pour les cohortes âgées de 20-24 à 40-44 ans incluent une composante projetée d’autant plus importante que la cohorte est jeune. Les données des cohortes plus âgées reflètent une fécondité achevée. Les probabilités d’agrandissement projetées pour les femmes de 30-44 ans font apparaître un recul systématique de la proportion de femmes qui devraient atteindre la parité suivante d’ici à la fin de leur période féconde. C’est le signe d’un processus en cours de baisse de la fécondité.

TABLEAU 8.10 Probabilités d’agrandissement projetées, Cambodge, recensements de 1998

| Parité (i) | Groupe d’âge | | | | | | | | |
|------------|--------------|----------|----------|----------|----------|--------|--------|--------|--------|
| | 20-24 | 25-29 | 30-34 | 35-39 | 40-44 | 45-49 | 50-54 | 55-59 | 60-64 |
| 0 | (0,7850) | 0,8679 | 0,9008 | 0,9224 | 0,9283 | 0,9315 | 0,9450 | 0,9481 | 0,9420 |
| 1 | | 0,9312 | 0,9337 | 0,9488 | 0,9403 | 0,9398 | 0,9429 | 0,9425 | 0,9333 |
| 2 | | (0,8053) | 0,8580 | 0,9081 | 0,9182 | 0,9180 | 0,9190 | 0,9263 | 0,9230 |
| 3 | | | 0,8083 | 0,8676 | 0,8878 | 0,8950 | 0,8958 | 0,9082 | 0,9156 |
| 4 | | | (0,7558) | 0,8055 | 0,8404 | 0,8578 | 0,8613 | 0,8794 | 0,8846 |
| 5 | | | | 0,7738 | 0,7885 | 0,8199 | 0,8251 | 0,8345 | 0,8454 |
| 6 | | | | 0,7094 | 0,7388 | 0,7730 | 0,7858 | 0,8008 | 0,8140 |
| 7 | | | | (0,6379) | 0,6715 | 0,7244 | 0,7409 | 0,7513 | 0,7446 |
| 8 | | | | | 0,6266 | 0,6589 | 0,6862 | 0,6965 | 0,7001 |
| 9 | | | | | 0,6104 | 0,6397 | 0,6577 | 0,6662 | 0,6707 |
| 10 | | | | | 0,5326 | 0,5534 | 0,5599 | 0,5677 | 0,5864 |
| 11 | | | | | (0,7611) | 0,5792 | 0,6063 | 0,6232 | 0,6388 |

FIGURE 8.1 Probabilités d'agrandissement projetées par parité et groupe d'âge, Cambodge, recensement de 1998

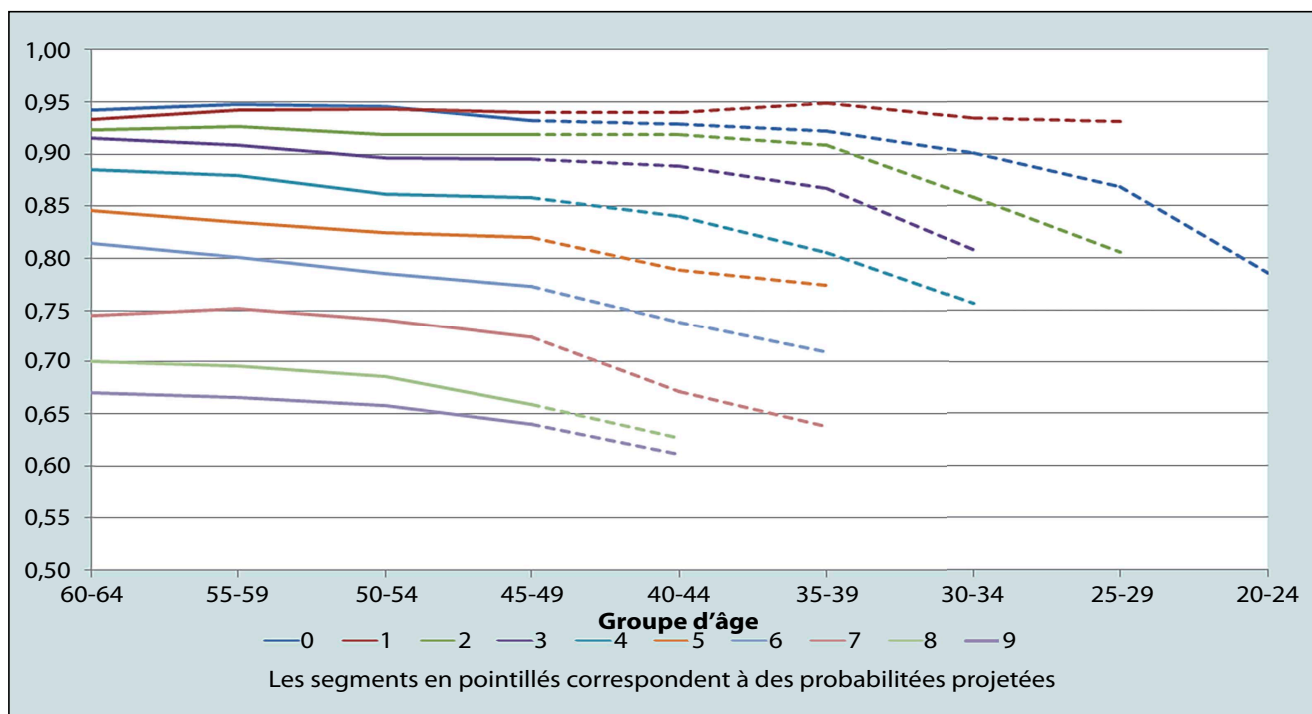
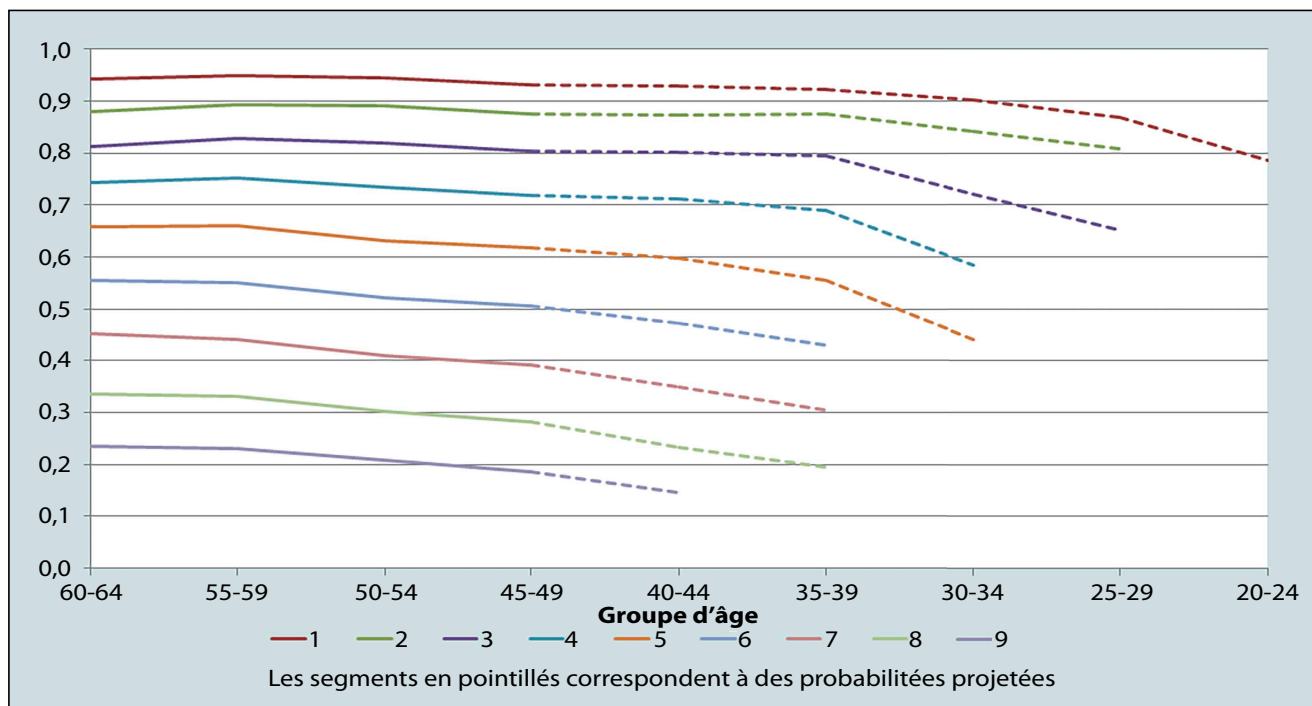


FIGURE 8.2 Proportion projetée de femmes atteignant chaque parité par groupe d'âge, Cambodge, recensement de 1998



AUTRES LECTURES ET RÉFÉRENCES

On trouvera dans Brass and Juárez (1983) l'exposé de méthodes semblables à celles présentées ci-dessus, mais utilisant les données détaillées d'histoires génésiques et accordant davantage d'attention aux effets de censure et de sélection.

De nombreux travaux portent sur l'utilisation et les problèmes associés aux probabilités d'agrandissement du moment (par opposition aux mesures prises dans les cohortes). Ceci déborde des objectifs du présent manuel. Le lecteur intéressé peut consulter Whelpton (1946), Feeney and Yu (1987), Ní Bhrolcháin (1987), Hinde (1998), et Bongaarts and Feeney (1998).

Bongaarts J and G Feeney. 1998. "On the quantum and tempo of fertility", *Population and Development Review* **24**(2):271-291. doi: <http://dx.doi.org/10.2307/2807974>

Brass W. 1985. *Advances in Methods for Estimating Fertility and Mortality from Limited and Defective Data*. London: Centre for Population Studies, London School of Hygiene and Tropical Medicine.

Brass W and F Juárez. 1983. "Censored cohort parity progression ratios from birth histories", *Asian and Pacific Census Forum* **10**(1):5-12

Feeney G and J Yu. 1987. "Period parity progression measures of fertility in China", *Population Studies* **41**(1):77-102. doi: <http://dx.doi.org/10.1080/0032472031000142546>

Hinde A. 1998. *Demographic Methods*. London: Arnold

Ní Bhrolcháin M. 1987. "Period parity progression ratios and birth intervals in England and Wales, 1941-1971: A synthetic life table analysis", *Population Studies* **41**(1):103-125. doi: <http://dx.doi.org/10.1080/0032472031000142556>

Preston SH, P Heuveline and M Guillot. 2001. *Demography: Measuring and Modelling Population Processes*. Oxford: Blackwell.

Whelpton PK. 1946. "Reproduction rates adjusted for age, parity, fecundity, and marriage", *Journal of the American Statistical Association* **41**(236):501-516. doi: <http://dx.doi.org/10.1080/01621459.1946.10501893>

Chapitre 9 ESTIMATION DE LA FÉCONDITÉ PAR LA SURVIE INVERSE

Ian M Timæus and Tom A Moultrie

DESCRIPTION DE LA MÉTHODE

La survie inverse est une méthode d'estimation de la fécondité à partir des données recueillies dans un recensement ou une enquête en une seule vague, qui peut être utilisée même si aucune question n'a été posée directement concernant la fécondité. Dans une population fermée à la migration, la population à un âge x est composée des survivants des naissances survenues dans cette population il y a x années révolues. Il en résulte que le nombre de naissances survenues il y a x années peut être calculé si on peut estimer les probabilités de survie de la table de mortalité de la naissance à l'âge x (c'est-à-dire $\frac{L_x}{l_0}$). On parle aussi de projection rétrospective ou rétro-projection.

En ramenant la population à ce qu'était son effectif l'année de sa naissance et en divisant le nombre de naissances par une estimation de la population totale cette année-là, on obtient le taux brut de natalité, et en divisant par une estimation du nombre des femmes d'âge fécond, on obtient le taux global de fécondité. En combinant les estimations rétrospectives des naissances et des femmes selon l'âge avec des estimations, ou des hypothèses raisonnables, sur la distribution de la fécondité par âge, on peut aussi estimer l'indice synthétique de fécondité.

Les questions sur les naissances de l'année écoulée ou sur la date de naissance du dernier né des femmes ne permettent qu'une estimation de la fécondité actuelle. Les méthodes de survie inverse peuvent fournir des estimations de fécondité pour les 15 dernières années. En outre, à la différence des estimations de fécondités tirées d'histoires génésiques, qui ne sont généralement recueillies qu'auprès des femmes âgées de 15 à 49 ans, les estimations obtenues par la survie inverse ne sont pas tronquées de plus en plus sévèrement aux âges élevés à mesure qu'elles sont calculées pour des périodes plus anciennes. Une distribution des enfants par année

d'âge permet de produire une série annuelle d'estimations de la fécondité. Toutefois, en pratique, les données sur l'âge recueillies dans les pays en développement sont rarement suffisamment précises pour qu'une série chronologique puisse être établie de façon fiable.

PRÉCAUTIONS ET MISES EN GARDE

Les méthodes décrites ici dépendent étroitement de la précision de la distribution par âge de la population déclarée. Des erreurs dans la distribution par âge, telles que l'omission de nourrissons et de jeunes enfants et l'attraction exercée par la déclaration de certains âges, peuvent avoir un impact important sur les résultats obtenus. Comme le note le *Manuel X*, « dans la mesure où les ensembles de données disponibles comportent bien trop fréquemment des erreurs ou des omissions de ce genre, il arrive souvent que les méthodes de la projection rétrospective ne permettent pas d'obtenir des estimations fiables de la fécondité. Leur utilité provient alors essentiellement du fait qu'elles fournissent des estimations originales de la fécondité qui permettent de mesurer de degré de vraisemblance des estimations qui ont été calculées par d'autres moyens » (UN Population Division 1983 : 178) (Division de la Population des NU 1984 : 186).

Le sous enregistrement des nourrissons et des jeunes enfants est courant dans les recensements et les enquêtes des pays en développement. Ce sous dénombrement des enfants, en particulier ceux de deux ans ou moins, se traduit par des estimations trop basses de la fécondité dans la même proportion. Les méthodes de projection rétrospective tendent ainsi à sous-estimer la fécondité actuelle et donnent une impression exagérée de baisse de la fécondité.

En outre, une mauvaise déclaration des âges, en particulier l'attraction par certains âges, peut se traduire par des séries chronologiques erratiques des estimations de fécondité.

Des recherches sur la qualité des données par âge et sexe sont indispensables avant de procéder à des estimations de la fécondité par l'application d'une des méthodes de survie inverse. Celles-ci doivent être utilisées avec prudence s'il existe des doutes sur la qualité des données.

Bien que l'opération soit possible, il n'est généralement pas recommandé de procéder à des estimations de fécondité par projection rétrospective remontant à plus de 15 ans avant le recueil des données. La première raison est que la migration et le sous-enregistrement différentiel risquent de biaiser les estimations fondées sur la distribution par âge des adultes âgés de 15 ans ou plus. En outre, des estimations de fécondité remontant plus loin en arrière dépendent de plus en plus de la précision des évaluations de la mortalité des femmes âgées.

Comme les résultats dépendent des hypothèses faites sur la survie, ils sont peu sensibles aux hypothèses sur la mortalité des enfants ou des femmes âgées de moins de 65 ans. Par exemple, si la probabilité de décéder d'un enfant né il y a une douzaine d'années était de 20 % et était sous-estimée d'un quart (soit une probabilité de 16 %), il en résulte une surestimation du nombre de naissances de $(100 - 16)/(100 - 20)$, donc seulement de 5 %. De même, la probabilité de décéder de la mère a peu de chances d'être largement supérieure à 20 %, même si celle-ci était âgée à la naissance de son enfant. Si cette probabilité était sous-estimée d'un quart, la fécondité ne serait surestimée que de 4,8 %. En outre, si les biais dans les estimations de mortalité vont dans la même direction pour les enfants et les mères, les erreurs qui en résultent dans les estimations de fécondité se compensent et s'annulent partiellement.

MÉTHODE 1 : ESTIMATION DU TAUX BRUT DE NATALITÉ PAR PÉRIODE QUINQUENNALE

Données nécessaires et hypothèses

Tableaux de données nécessaires

Pour calculer les taux moyens de natalité pour les périodes quinquennales précédant le recensement ou l'enquête, les données suivantes sont nécessaires :

- la population recensée âgée de moins de 15 ans, classée par groupes d'âge quinquennaux ;
- la population totale recensée à deux dates, ou la population totale recensée à une seule date et une estimation du taux de croissance ; et
- des estimations de la survie des enfants tirées de tables de mortalité, ${}_5L_0$, ${}_5L_5$ et ${}_5L_{10}$.

Hypothèses importantes

La population est supposée avoir été fermée à la migration pendant toute la période couverte par les estimations par rétroprojection. Toutefois, comme les enfants migrent généralement avec leur mère, les erreurs au numérateur et au dénominateur des taux estimés se compensent largement. Le biais n'est important que si les flux migratoires sont abondants et que les migrants ont une fécondité différente du reste de la population.

Application de la méthode

Étape 1 : Calculer les survivants de la table de mortalité des enfants

Pour projeter rétrospectivement la population des trois groupes quinquennaux 0-4, 5-9 et 10-14 ans, on n'a besoin que des valeurs de ${}_5L_0$, ${}_5L_5$ et ${}_5L_{10}$, les personnes-années vécues par la population stationnaire entre la naissance et 5 ans, et entre 5 et 10 ans, et 10 et 15 ans, respectivement. La source de ces mesures peut être une table de mortalité effectivement construite pour la population étudiée ou une table-type jugée adéquate (par exemple une des tables-types régionales de Princeton). Étant donné le haut degré d'agrégation entre les âges, on peut supposer que la mortalité est constante au long de chaque période quinquennale précédant le recensement ou l'enquête. Si la mortalité est surestimée, la fécondité sera surestimée (puisque la population dénombrée sera constituée des survivants de cohortes de naissances trop importantes) et vice versa.

Comme la mortalité évolue rapidement au cours de la première année de vie, ${}_5L_0$ doit être calculé comme ${}_5L_0 = {}_1L_0 + {}_4L_1$ si on utilise une table de mortalité effective du pays considéré. Si cette table fait la distinction entre filles et garçons, une table doit être calculée pour les deux sexes réunis moyennant une hypothèse appropriée sur le rapport de masculinité à la naissance. Soit la formule

$$l_x = \frac{\frac{S_0}{100} \cdot l_x^m + l_x^f}{\frac{100 + S_0}{100}}$$

où l_x^m et l_x^f sont les valeurs de l_x pour les garçons et les filles respectivement, et S_0 est le rapport de masculinité à la naissance (nombre de naissances de garçons pour 100 naissances de filles).

Si une table-type de mortalité est utilisée pour représenter la mortalité dans l'enfance, ${}_5L_0$, ${}_5L_5$, and ${}_5L_{10}$ sont obtenus par interpolation entre les valeurs appropriées dans la série des tables types.

Étape 2 : Estimation des populations au milieu de la période

Pour estimer des taux annuels de natalité pour les trois périodes quinquennales précédant le recensement, il faut disposer d’une estimation de la population totale au milieu de chacune des trois périodes. En l’absence d’une information détaillée directement utilisable (par exemple les résultats d’un recensement réalisé 10 ans plus tôt), et étant donné l’inévitable perte de précision due à l’agrégation des effectifs de population et de la mortalité en groupes quinquennaux, une estimation précise n’est pas nécessaire. Le mode le plus facile d’estimation de la population totale d’années auparavant, $N(t - d)$, où $d = 2,5$ ans, $7,5$ et $12,5$, est de supposer un taux de croissance constant, r , et de l’appliquer à la population dénombrée au temps t , $N(t)$, soit

$$N(t - d) = N(t) \cdot \exp(-d \cdot r)$$

Le taux de croissance, r , peut être estimé à partir d’informations sur l’effectif de la population totale, P , à deux points du temps, t_0 et t_1 , de sorte que

$$r = \frac{\ln\left(\frac{P_1}{P_0}\right)}{(t_1 - t_0)}$$

Ainsi, par exemple, la population 2,5 ans avant le recensement le plus récent est égale à

$$N(t - 2,5) = N(t) \cdot \exp(-2,5r)$$

ce qu’on peut étendre aux estimations de la population à d’autres points du temps.

Étape 3 : Estimer les naissances de chaque période quinquennale antérieure au recensement

Soit $B(t - 5, t)$ le nombre de naissances survenues chaque année au cours de la période $t - 5$ à t années avant le recensement, où $t = 0, 5$ ou 10 . Soit aussi ${}_5N_x(t)$ la population dénombrée âgée entre x et $x + 5$ ans à la date de recensement. Il s’ensuit que le nombre moyen de naissances dans chacune des périodes quinquennales de $t - 5$ à t est égal à

$$B(t - 5, t) = {}_5N_0(t) \cdot l_0 / {}_5L_0$$

$$B(t - 10, t - 5) = {}_5N_5(t) \cdot l_0 / {}_5L_5$$

$$\text{et} \qquad B(t - 15, t - 10) = {}_5N_{10}(t) \cdot l_0 / {}_5L_{10}$$

Le taux brut de natalité (en anglais, *Crude Birth Rate, CBR*) de chaque période est alors calculé en divisant le nombre de naissances de chaque période par la population correspondante en milieu de période, telle que calculée à l’étape 2. Soit

$$CBR(t - 5, t) = \frac{B(t - 5, t)}{N(t - 2.5)}$$

et de même pour les périodes antérieures.

EXEMPLE

Les données du recensement de 2008 au Cambodge nous permettent d’illustrer les principes de l’estimation de la fécondité par la projection rétrospective.

Étape 1 : Calculer les survivants de la table de mortalité des enfants

La série des survivants est tirée de la table de mortalité de l’OMS pour le Cambodge en 2000. Les valeurs de ${}_nL_x$ (sexes réunis) dans cette table de mortalité sont présentées au tableau 9.1 pour une racine de 100 000.

TABLEAU 9.1 Valeurs tirées de la table de mortalité de l’OMS pour le Cambodge, 2000

| Âge | <i>n</i> | <i>x</i> | <i>l_x</i> | <i>d_x</i> | <i>nL_x</i> |
|-----|----------|----------|----------------------|----------------------|-----------------------|
| < 1 | 1 | 0 | 100 000 | 7 956 | 94 431 |
| 1-4 | 4 | 1 | 92 044 | 2 684 | 361 734 |
| 5-9 | 5 | 5 | 89 360 | 887 | 444 581 |

À partir de ces données, on peut calculer les grandeurs suivantes

$$\frac{l_0}{{}_5L_0} = \frac{100\,000}{94\,431 + 361\,734} = 0,21922$$

$$\frac{l_0}{{}_5L_5} = \frac{100\,000}{444\,581} = 0,22493$$

Étape 2 : Estimation des populations au milieu de la période

D’après les échantillons des recensements cambodgiens diffusés par IPUMS International, le Cambodge comptait 11 412 540 habitants le 3 mars 1998. Le 3 mars 2008, c’est

13 401 210. La population de moins de 5 ans en 2008 est de 1 736 790 et celle de 5-9 ans 1 472 470.

Comme les recensements ont été réalisés précisément à 10 ans d'écart, le taux de croissance intercensitaire est égal à

$$r = \frac{\ln\left(\frac{13401210}{11412540}\right)}{10} = 1,60632$$

pour 100 par an.

La population 2,5 ans et 7,5 ans avant le second recensement est estimée à :

$$N(t - 2,5) = N(t) \cdot \exp(-2,5r) = 12\,873\,706$$

$$N(t - 7,5) = N(t) \cdot \exp(-7,5r) = 11\,880\,172$$

Étape 3 : Estimer les naissances de chaque période quinquennale antérieure au recensement

Le nombre estimé de naissances pour chacune des périodes des cinq années précédant le recensement est égal à

$$B(t - 5, t) = 1\,736\,790 \times 0,21922 = 380\,739.$$

Pour les cinq années précédentes, c'est

$$B(t - 10, t - 5) = 1\,472\,471 \times 0,22493 = 331\,202.$$

En divisant les naissances par la population au milieu de la période, on obtient des taux bruts de natalité de 29,6 naissances pour 1 000 pour le Cambodge en 2003-2008 et 27,9 naissances pour 1 000 en 1998-2003.

L'ordre de grandeur de ces taux est comparable à ceux de l'US Census Bureau (28 en 2000 et 27 en 2005), et de la Division de la Population des NU (25,1 naissances pour 1 000 en 2000-2005). Les estimations par la projection rétrospective suggèrent cependant une légère hausse de la fécondité pendant les années 1990. Il se pourrait que ce résultat soit faux et qu'il résulte d'erreurs dans les estimations initiales de la population ou au choix d'une table de mortalité inappropriée.

MÉTHODE 2 : ESTIMATION DU TAUX GLOBAL DE FÉCONDITÉ ET DE L'INDICE SYNTHÉTIQUE DE FÉCONDITÉ

La seconde application du concept de survie inverse décrite ici permet d'obtenir des estimations du taux global de fécondité et de l'indice synthétique de fécondité jusqu'à

15 ans avant le recensement ou l'enquête. Pour calculer le taux global de fécondité, il suffit d'estimer rétrospectivement la taille de la population adulte féminine à partir du nombre de femmes dénombrées, en tenant compte de la mortalité adulte. Mais pour calculer l'indice synthétique de fécondité, il ne suffit pas de connaître le nombre de naissances chaque année, mais aussi l'âge des mères de ces enfants. Une façon relativement simple de procéder, si cette information n'est pas disponible, consiste à répartir les naissances par groupe d'âge des mères d'après des estimations indépendantes de la distribution par âge de la fécondité.

Données nécessaires et hypothèses

Tableaux de données nécessaires

Pour le calcul du taux global de fécondité par année, les données suivantes sont nécessaires :

- Des tableaux de la population (sexes réunis) de 0 à 14 ans par année d'âge.
- Des tableaux de la population féminine de 15 à 64 ans par groupes d'âges quinquennaux.
- Les survivants des tables de mortalité des cohortes, L_x , pour les enfants de 0 à 14 ans (sexes réunis)
- Les probabilités de survie, ${}_5L_x/{}_5L_x$ des femmes adultes pour chacune des trois périodes quinquennales précédant le recensement ou l'enquête.

La mise en œuvre de la méthode dans le dossier Excel associé disponibles sur <http://demographicestimation.iussp.org/fr> est possible avec des estimations de mortalité obtenues soit par référence aux paramètres de période α et β des tables-types relationnelles appropriées, soit à partir des valeurs calculées ${}_5q_0$ (pour les enfants) et ${}_{45}q_{15}$ (pour les femmes adultes) pour chacune des trois périodes quinquennales précédant le recensement ou l'enquête.

Pour estimer l'indice synthétique de fécondité, il faut aussi disposer

- soit d'une seule distribution de la fécondité par âge supposée applicable à l'ensemble de la période couverte par nos estimations,
- soit de deux distributions de la fécondité par âge, l'une applicable à une date raisonnablement proche du recensement ou de l'enquête et l'autre à une date antérieure d'environ 15 ans.

Les calculs peuvent être menés à partir d'une série de taux de fécondité ou à partir des paramètres d'un modèle relationnel de Gompertz ajusté à un schéma type de fécondité.

Seule la forme estimée de la distribution de fécondité est fondée sur ces schémas de fécondité. C'est le nombre estimé de naissances rapporté à la population des femmes d'âge fécond qui détermine presque entièrement les estimations du taux global et de l'indice synthétique de fécondité.

Hypothèses importantes

La population est supposée avoir été fermée à la migration pendant toute la période couverte par les estimations par rétroprojection. Toutefois, comme les enfants migrent généralement avec leur mère, les erreurs au numérateur et au dénominateur des taux estimés se compensent largement. Le biais n'est important que si les flux migratoires sont abondants et que les migrants ont une fécondité différente du reste de la population.

Application de la méthode

Étape 1 : Estimer le nombre de naissances chaque année précédant le recensement

La population recensée à l'âge x est composée des survivants des naissances qui sont survenues au cours de la période de 12 mois centrée sur la date $x + 0,5$ ans avant le recensement. Algébriquement,

$$B_{x+0,5} = \frac{N_x}{{}^cL_x}, 0 \leq x \leq 14$$

Dans ce calcul, la mesure des survivants, cL_x , est prise dans une *cohorte*. Elle dépend de la mortalité aux âges successifs dans les années successives conduisant au recensement. Il se peut que les estimations par cohorte de la mortalité soient tirées du recensement ou de l'enquête utilisé pour estimer la fécondité. Ce peuvent être des estimations indirectes à partir des données sur les enfants déjà nés et les enfants survivants, dans le cas d'un recensement, ou des estimations directes tirées d'une analyse par cohorte des histoires génésiques, dans le cas d'une enquête sur la fécondité. Brass (1979) décrit une procédure simple pour estimer directement L_x à partir des proportions d'enfants survivants selon le groupe d'âge de leur mère.

Si on ne dispose pas de taux de survie par cohorte, ceux-ci peuvent être tirés d'estimations de la mortalité du moment par année d'âge pour chacune de périodes quinquennales précédant le recensement ou l'enquête.

Si on travaille avec un système logit relationnel de tables-types de mortalité, on doit définir α_T et β_T comme para-

mètres générant une table de mortalité pour la période T , où $T = 0$ désigne la période 0-4 ans avant le recensement, $T = 5$ la période 5-9 ans avant et $T = 10$ la période 10-14 ans avant. Pour un standard donné, désigné par l'exposant s ,

$$Y_{x,T} = \alpha_T + \beta_T \cdot Y_x^s$$

où Y est la fonction logit :

$$Y_x = \frac{1}{2} \ln \left(\frac{1 - l_x}{l_x} \right) \quad (1)$$

où l_x ou ${}_x p_0$ désigne la proportion de survivants de la naissance à l'anniversaire x dans une table de mortalité de racine 1. L'équation 1 implique que

$$l_x = (1 + \exp(2Y_x))^{-1}$$

et que

$$l_{x,T} = \left(1 + \exp \left(2 \left(\alpha_T + \beta_T \cdot Y_x^s \right) \right) \right)^{-1} \quad (2)$$

Au-delà de la première année de vie, on peut approximer $L_{x,T}$, les personnes années vécues entre x et $x + 1$ dans la période T en supposant que la survie recule linéairement sur une échelle logistique, et donc que le logit de $L_{x,T}$ est la moyenne de $Y_{x,T}$ et $Y_{x+1,T}$. De l'équation 2

$$L_{x,T} \approx l_{x+0,5,T} = \left(1 + \exp \left(2 \left(\frac{Y_{x,T} + Y_{x+1,T}}{2} \right) \right) \right)^{-1} \\ = \left(1 + \exp(Y_{x,T} + Y_{x+1,T}) \right)^{-1}, 0 < x \leq 14 \quad (3)$$

Les valeurs de l_x dans les réseaux de tables-types de mortalité sont souvent données par année d'âge jusqu'à 5 ans, puis de cinq en cinq ans. Si c'est le cas pour les enfants de 5 ans ou plus, on peut supposer que le logit de la fonction de survie recule linéairement sur l'ensemble de l'intervalle d'âge x à $x + 5$. Ainsi par exemple, si une table de mortalité donne des valeurs pour $x = 5$ et $x = 10$, la valeur estimée de $L_{9,T}$ est donnée par

$$L_{9,T} \approx l_{9,5,T} = \left(1 + \exp \left(2 \left(\frac{1}{10} Y_{5,T} + \frac{9}{10} Y_{10,T} \right) \right) \right)^{-1} \\ = \left(1 + \exp(0,2Y_{5,T} + 1,8Y_{10,T}) \right)^{-1}$$

Pour la première année de vie, on doit tenir compte de la concentration des décès dans les premiers jours et les premières semaines. Dans les populations où la mortalité est moyenne ou élevée, on peut approximer les personnes-années vécues dans la première année par

$$L_{0,T} = 0,3 + 0,7 \left(\exp(2Y_{1,T}) \right)^{-1} \quad (4)$$

Les probabilités de survie d'un âge au suivant, $P_{x,T}$, au cours de la période T , sont tirés du rapport entre les valeurs successives de $L_{x,T}$:

$$P_{x,T} = \frac{L_{x,T}}{L_{x-1,T}}, 0 < x \leq 14 \quad (5)$$

$$P_{0,T} = L_{0,T}$$

Une fois obtenue la série des enfants survivants par année d'âge et périodes quinquennales, que ce soit par la procédure qui vient d'être indiquée ou par une autre, une estimation des survivants par année d'âge par cohorte annuelle peut être calculée comme suit. On rappelle que $P_{a,T}$ est la probabilité de survie entre les âges a et $a+1$ au cours de la période T (où $T=0, 5$ ou 10 , selon qu'il s'agit de la période 0-4, 5-9 ou 10-14 ans avant le recensement). On définit $S_{a,t}$ la probabilité de survie entre les âges a et $a+1$ au cours de la période t à $t+1$ années avant le recensement, $0 \leq t \leq 14$. En utilisant l'interpolation linéaire pour estimer la survie pour les années intermédiaires

$$S_{a,t} = P_{a,0}, 0 \leq t \leq 2$$

$$S_{a,t} = P_{a,T}, \left(1 - \frac{t-T-2}{5} \right) + P_{a,T+5} \left(\frac{t-T-2}{5} \right) 2 < t-T < 8$$

$$S_{a,t} = P_{a,10}, 12 < t < 15$$

La proportion de naissances survenues x à $x+1$ années plus tôt qui survivent au moment du recensement, cL_x , peut alors être calculée comme

$${}^cL_x = S_{0,x} \cdot S_{1,x-1} \cdots S_{x-1,1} \cdot S_{x,0} \quad (6)$$

Le nombre de naissances de chaque année avant le recensement, centré sur le point médian de cette année-là (c'est-à-dire 6 mois avant la date du recensement), est donc

$$B_{x+0,5} = \frac{N_x}{{}^cL_x}, 0 \leq x \leq 14$$

où N_x désigne le nombre d'enfants âgés de x déclarés au recensement.

Étape 2 : Estimer les populations en milieu d'année des femmes par groupe quinquennal d'âge

Le calcul de la survie des femmes âgées de 15 à 64 ans au moment du recensement est simple à réaliser car la mortalité est généralement faible aux âges adultes. Des estimations même approximatives de la mortalité suffisent donc pour une estimation satisfaisante de la population passée à partir de la population recensée.

La variation absolue de la mortalité au fil des âges dans un groupe quinquennal d'âge est faible. On peut donc approximer ${}_5L_x$ par une interpolation linéaire entre Y_x et Y_{x+5} . Ceci signifie qu'on peut estimer la survie entre les groupes d'âge quinquennaux au temps T (avec $T=5, 10$ et 15) comme

$${}_5P_{x,T} = \frac{{}_5L_{x+5}}{{}_5L_x} = \frac{1 + \exp(2\alpha_T + \beta_T(Y_x^s + Y_{x+5}^s))}{1 + \exp(2\alpha_T + \beta_T(Y_{x+5}^s + Y_{x+10}^s))} \quad (7)$$

En débutant avec la population recensée en $T=0$, on peut calculer le nombre de femmes dans chaque groupe quinquennal d'âge $T+5$ ans avant le recensement à partir du nombre en T :

$${}_5N_{x,T+5}^f = \frac{{}_5N_{x+5,T}^f}{{}_5P_{x,T}}, 10 \leq x < 60, T = 0, 5, 10 \quad (8)$$

En outre, comme la structure par âge d'une population ne change que lentement, on peut calculer les populations des femmes âgées de 10-14, 15-19, ..., 60-64 ans pour chacune des années avant la date du recensement par interpolation linéaire entre les estimations de populations 0, 5, 10 et 15 ans avant le recensement produites grâce à l'équation 8. Par exemple, pour estimer le nombre de femmes âgées de 20-24 ans 8,5 ans avant le recensement, la formule serait

$${}_5N_{20,8,5} = 0,3({}_5N_{20,5}) + 0,7({}_5N_{20,10})$$

Étape 3 : Calculer les taux globaux de fécondité

Le taux global de fécondité pour l'année centrée sur $x-0,5$ ans avant le recensement est

$$GFR_{x+0,5} = \frac{B_{x+0,5}}{\sum_{a=15,5}^{45} {}_5N_{a,x+0,5}}, 0 \leq x \leq 14 \quad (9)$$

où le dénominateur est le nombre total de femmes âgées de 15 à 49 ans au point médian de l'année au cours de laquelle les naissances sont survenues.

Étape 4 : Estimer les taux par âge et l'indice synthétique de fécondité

Au-delà du calcul des taux globaux de fécondité, une extension naturelle consiste à utiliser un schéma de distribution de la fécondité par âge dans la population étudiée pour estimer l'indice synthétique de fécondité par une procédure analogue à celle de la standardisation indirecte.

Ce schéma de fécondité peut provenir de données sur les naissances récentes recueillies dans le même recensement que celui analysé par les méthodes de projection rétrospective. Comme on n'en tirera que des informations sur la distribution par âge de la fécondité, il n'est pas nécessaire d'ajuster les données pour tenir compte des erreurs sur la période de référence. Si une autre distribution est connue à partir d'un recensement ou d'une enquête antérieure (si possible d'environ 15 ans) dans la même population, on peut interpoler entre les deux distributions, ou extrapoler si nécessaire, afin d'estimer la forme de la distribution de la fécondité pour chacune des années dont on souhaite estimer l'indice synthétique de fécondité.

Mais même si la fécondité a changé, il n'est pas essentiel d'avoir deux distributions de fécondité, car la distribution par âge de la fécondité évolue graduellement et les estimations de l'indice synthétique sont peu sensibles à la précision des hypothèses sur la forme de la courbe de fécondité. Aussi une seule distribution (de préférence vers le milieu de la période d'estimation) suffit-elle.

Si, pour estimer la fécondité, on a utilisé un modèle relationnel de Gompertz, en combinaison avec le standard choisi, les paramètres α et β calculés lors de l'estimation de la fécondité définissent la forme de la courbe de fécondité. En conséquence, les valeurs de α et β peuvent être utilisées pour déterminer la forme de la (des) courbe(s) de fécondité utilisées dans l'estimation de l'indice synthétique.

Après avoir estimé la proportion de la fécondité totale qui survient dans chaque groupe d'âge pour chaque année antérieure au recensement, on peut appliquer ces

proportions à la population des femmes de chaque groupe d'âge chaque année pour estimer le nombre de naissances qu'auraient eu les femmes de ce groupe d'âge si l'indice synthétique avait été égal à un enfant par femme.

Donc, une fois qu'on a choisi une courbe de fécondité (ramenée à un indice synthétique égal à 1), ${}_5f_{a,x+0,5}^*$, pour chaque groupe d'âge ($a = 15, 20, \dots, 45$) pour chacune des 15 années (x) antérieures au recensement, le nombre de naissances attendu des femmes de chaque groupe d'âge pour chaque année est donné par

$${}_5B_{a,x+0,5}^* = {}_5N_{a,x+0,5} \cdot {}_5f_{a,x+0,5}^*, 0 \leq x \leq 14$$

Il s'ensuit que le nombre total de naissances qui serait survenu l'année x , si l'indice synthétique avait été égal à 1, est de

$$B_{x+0,5}^* = \sum_{a=15,5}^{45} {}_5N_{a,x+0,5} \cdot {}_5f_{a,x+0,5}^*, 0 \leq x \leq 14$$

Mais l'étape 2 a permis d'estimer le nombre effectif de naissances chaque année, $B_{x+0,5}$. L'indice synthétique de fécondité estimé pour chaque année est donc égal au rapport entre B et B^*

$$TF_{x+0,5} = \frac{B_{x+0,5}}{B_{x+0,5}^*}, 0 \leq x \leq 14$$

On peut obtenir des estimations des taux de fécondité par âge pour l'année en multipliant les taux proportionnels, ${}_5f_{a,x}^*$, pour chaque groupe d'âge (a) et chaque année (x) par l'estimation de l'indice synthétique pour cette année. On remarquera cependant que ces résultats se contentent de reproduire la distribution de fécondité par âge introduite dans les calculs : la méthode ne donne pas de nouvelle information sur la distribution par âge de la fécondité.

EXEMPLE

L'exemple, mis en œuvre dans le dossier Excel associé disponibles sur <http://demographicestimation.iussp.org/fr>, consiste à estimer l'évolution de la fécondité au Cambodge à partir du recensement de 2008. L'échantillon de micro-données est celui mis à disposition par IPUMS International et la date de référence du recensement est le 3 mars 2008. Dans une notation décimale, cette date correspond à 2008,17.

Les nombres d'enfants âgés de moins de 15 ans par année d'âge, et les nombres de femmes âgées de 10 à 64 ans par groupe quinquennal tirés des données figurent au tableau 9.2.

TABLERAU 9.2 Nombres d'enfants par année d'âge, et de femmes par groupe quinquennal, Cambodge, recensement de 2008

| Âge | Enfants | Groupe d'âge | Femmes |
|-----|---------|--------------|---------|
| 0 | 281 260 | 10-14 | 815 930 |
| 1 | 261 320 | 15-19 | 780 320 |
| 2 | 268 410 | 20-24 | 697 160 |
| 3 | 286 810 | 25-29 | 626 430 |
| 4 | 278 990 | 30-34 | 361 650 |
| 5 | 293 760 | 35-39 | 435 880 |
| 6 | 293 490 | 40-44 | 393 760 |
| 7 | 302 060 | 45-49 | 352 520 |
| 8 | 315 970 | 50-54 | 294 280 |
| 9 | 267 190 | 55-59 | 230 200 |
| 10 | 326 980 | 60-64 | 160 590 |
| 11 | 280 260 | | |
| 12 | 354 120 | | |
| 13 | 356 920 | | |
| 14 | 354 830 | | |

La mortalité pour la période de cinq ans précédant le recensement est tirée des tables de mortalité 2009 OMS pour le Cambodge. Les survivants aux anniversaires x , ainsi que les logits correspondants, pour les enfants de moins de 15 ans (sexes réunis) et pour les femmes de 10 à 65 ans figurent au tableau 9.3.

Cette table de mortalité a été utilisée comme standard sur lequel ajuster des tables types de mortalité relationnelles pour les enfants et les adultes séparément pour les périodes antérieures de 5-9 et 10-14 ans au recensement aux tables de mortalité de l'OMS pour le Cambodge en 2000 et 1990. Les valeurs ajustées de α et β pour les enfants et les adultes figurent au tableau 9.4.

Les estimations des taux de fécondité par âge pour mi-1997 et mi-2007 sont empruntées aux *Perspectives de Population Mondiale de la Division de la Population des NU* (révision de 2010). Elles figurent au tableau 9.5.

Étape 1 : Estimer le nombre de naissances chaque année précédant le recensement

En appliquant les équations 3 et 4 aux valeurs logit du standard présenté au tableau 9.3, en utilisant les valeurs

TABLERAU 9.3 Table de mortalité OMS pour le Cambodge en 2009 utilisée comme standard

| Enfants | $l(x)$ | Logit, $Y(x)$ |
|---------|--------|---------------|
| 0 | 1,0000 | |
| 1 | 0,9320 | - 1,3085 |
| 5 | 0,9125 | - 1,1723 |
| 10 | 0,9048 | - 1,1256 |
| 15 | 0,8997 | - 1,0968 |
| Femmes | | |
| 10 | 0,9138 | - 1,1805 |
| 15 | 0,9099 | - 1,1561 |
| 20 | 0,9049 | - 1,1266 |
| 25 | 0,8980 | - 1,0875 |
| 30 | 0,8889 | - 1,0399 |
| 35 | 0,8760 | - 0,9774 |
| 40 | 0,8603 | - 0,9089 |
| 45 | 0,8419 | - 0,8361 |
| 50 | 0,8179 | - 0,7511 |
| 55 | 0,7847 | - 0,6467 |
| 60 | 0,7374 | - 0,5161 |
| 65 | 0,6732 | - 0,3613 |

TABLERAU 9.4 Paramètres des tables types de mortalité relationnelles pour les périodes quinquennales avant le recensement de 2008, la table de mortalité OMS 2009 pour le Cambodge étant le standard

| | | Période avant le recensement (années) | | |
|---------|----------|---------------------------------------|------|-------|
| | | 0-4 | 5-9 | 10-14 |
| Enfants | α | 0,00 | 0,22 | 0,35 |
| | β | 1,00 | 1,10 | 1,17 |
| Adultes | α | 0,00 | 0,08 | 0,00 |
| | β | 1,00 | 0,97 | 0,87 |

TABLERAU 9.5 Taux de fécondité pour le Cambodge utilisés dans la projection rétrospective

| Groupe d'âge | Années | |
|--------------|--------|--------|
| | 1997,5 | 2007,5 |
| 15-19 | 0,053 | 0,042 |
| 20-24 | 0,197 | 0,154 |
| 25-29 | 0,214 | 0,148 |
| 30-34 | 0,184 | 0,112 |
| 35-39 | 0,133 | 0,071 |
| 40-44 | 0,068 | 0,030 |
| 45-49 | 0,013 | 0,003 |

TABLEAU 9.6 Calcul des survivants et des probabilités de survie des cohortes par âge, Cambodge 1983-2008

| x | $L_{x,0}$ | $L_{x,5}$ | $L_{x,10}$ | $P_{x,0}$ | $P_{x,5}$ | $P_{x,10}$ | cL_x |
|-----|-----------|-----------|------------|-----------|-----------|------------|-----------|
| 0 | 0,9524 | 0,9439 | 0,9397 | 0,9524 | 0,9439 | 0,9397 | 0,9524 |
| 1 | 0,9298 | 0,9170 | 0,9106 | 0,9763 | 0,9716 | 0,9691 | 0,9298 |
| 2 | 0,9252 | 0,9112 | 0,9039 | 0,9951 | 0,9936 | 0,9926 | 0,9252 |
| 3 | 0,9197 | 0,9041 | 0,8958 | 0,9941 | 0,9923 | 0,9910 | 0,9181 |
| 4 | 0,9145 | 0,8974 | 0,8881 | 0,9944 | 0,9926 | 0,9914 | 0,9104 |
| 5 | 0,9118 | 0,8939 | 0,8840 | 0,9970 | 0,9960 | 0,9954 | 0,9049 |
| 6 | 0,9102 | 0,8919 | 0,8818 | 0,9983 | 0,9978 | 0,9974 | 0,9003 |
| 7 | 0,9087 | 0,8899 | 0,8795 | 0,9983 | 0,9978 | 0,9974 | 0,8954 |
| 8 | 0,9071 | 0,8879 | 0,8771 | 0,9983 | 0,9977 | 0,9973 | 0,8911 |
| 9 | 0,9056 | 0,8858 | 0,8747 | 0,9983 | 0,9977 | 0,9973 | 0,8871 |
| 10 | 0,9043 | 0,8842 | 0,8728 | 0,9986 | 0,9981 | 0,9978 | 0,8834 |
| 11 | 0,9033 | 0,8828 | 0,8713 | 0,9989 | 0,9985 | 0,9983 | 0,8801 |
| 12 | 0,9022 | 0,8815 | 0,8698 | 0,9989 | 0,9985 | 0,9983 | 0,8767 |
| 13 | 0,9012 | 0,8802 | 0,8682 | 0,9989 | 0,9985 | 0,9982 | 0,8741 |
| 14 | 0,9002 | 0,8789 | 0,8667 | 0,9989 | 0,9985 | 0,9982 | 0,8719 |

TABLEAU 9.7 Nombre estimé de naissances pour chacune des 15 années précédant le recensement de 2008 au Cambodge

| Milieu d'année précédant la date du recensement | Naissances |
|---|------------|
| 2007,67 | 295 328 |
| 2006,67 | 281 062 |
| 2005,67 | 290 118 |
| 2004,67 | 312 409 |
| 2003,67 | 306 453 |
| 2002,67 | 324 644 |
| 2001,67 | 325 997 |
| 2000,67 | 337 355 |
| 1999,67 | 354 583 |
| 1998,67 | 301 184 |
| 1997,67 | 370 118 |
| 1996,67 | 318 451 |
| 1995,67 | 403 924 |
| 1994,67 | 408 337 |
| 1993,67 | 406 976 |

de α et β du tableau 9.4, on obtient les estimations de $L_{x,T}$ et $P_{x,T}$ qui figurent au tableau 9.6.

La survie de cohorte jusqu'à la date du recensement pour les enfants âgés de 0 à 14 ans est ensuite estimée à partir des équations 5 et 6 (dernière colonne du tableau 9.6).

Enfin, le nombre de naissances pour chaque année antérieure au recensement est obtenu en divisant le nombre d'enfants recensés en 2008 à chaque âge par la valeur appropriée de cL_x (Tableau 9.7).

Étape 2 : Estimer les populations en milieu d'année des femmes par groupe quinquennal d'âge

Les probabilités de survie des femmes par groupe quinquennal d'âge (tableau 9.8) sont obtenues en appliquant les valeurs de α et β pour chaque période de cinq ans (tableau 9.4) aux logits du standard (tableau 9.3), au moyen de l'équation 7.

Les estimations de population des femmes par groupe quinquennal d'âge 5, 10 et 15 ans avant le recensement (tableau 9.9) sont obtenues par application des probabilités de survie (tableau 9.8) à la population dénombrée des femmes âgées de 10 à 64 ans, au moyen de l'équation 8.

TABLEAU 9.8 Probabilités de survie, ${}_5P_{x,T}$, par groupe quinquennal d'âge et période de cinq ans avant le recensement, Cambodge 1993-1998

| Groupe d'âge | $T=0$ | $T=5$ | $T=10$ |
|--------------|--------|--------|--------|
| 10-14 | 0,9951 | 0,9942 | 0,9945 |
| 15-19 | 0,9935 | 0,9922 | 0,9927 |
| 20-24 | 0,9912 | 0,9895 | 0,9902 |
| 25-29 | 0,9878 | 0,9855 | 0,9866 |
| 30-34 | 0,9838 | 0,9810 | 0,9826 |
| 35-39 | 0,9804 | 0,9772 | 0,9793 |
| 40-44 | 0,9752 | 0,9714 | 0,9744 |
| 45-49 | 0,9658 | 0,9609 | 0,9654 |
| 50-54 | 0,9501 | 0,9438 | |
| 55-59 | 0,9271 | | |

L'interpolation linéaire entre les quatre colonnes de droite du tableau 9 permet d'estimer aisément la population par groupe d'âge quinquennal pour les 15 années centrées sur les dates 6, 18, 30 ... 174 mois avant le recensement, c'est-à-dire en 2007,67 ; 2006,67 ; ... ; 1993,67. Un extrait des résultats est présenté au tableau 10.

TABLEAU 9.9 Nombre de femmes par groupe quinquennal d'âge 0, 5, 10 et 15 ans avant le recensement de 2008 au Cambodge

| Groupe d'âge | 2008,17 | 2003,17 | 1998,17 | 1993,17 |
|--------------|---------|---------|---------|---------|
| 10-14 | 815 930 | 784 134 | 705 852 | 640 512 |
| 15-19 | 780 320 | 701 746 | 636 977 | 372 754 |
| 20-24 | 697 160 | 632 014 | 370 014 | 454 008 |
| 25-29 | 626 430 | 366 133 | 449 548 | 414 964 |
| 30-34 | 361 650 | 443 047 | 409 395 | 376 470 |
| 35-39 | 435 880 | 401 624 | 369 909 | 320 300 |
| 40-44 | 393 760 | 361 474 | 313 672 | 258 772 |
| 45-49 | 352 520 | 304 701 | 252 137 | 190 119 |
| 50-54 | 294 280 | 242 285 | 183 546 | |
| 55-59 | 230 200 | 173 224 | | |
| 60-64 | 160 590 | | | |

Étape 3 : Calculer les taux globaux de fécondité

Les taux globaux de fécondité sont calculés grâce à l'équation 9, les naissances du tableau 9.7 et la population des femmes de 15-49 ans (tableau 9.10).

Étape 4 : Estimer les taux par âge et l'indice synthétique de fécondité

Par interpolation linéaire entre les deux séries de taux de fécondité du tableau 9.5, on peut calculer les distributions de fécondité – pour un indice synthétique d'un enfant par femme – pour chacune des années précédant le recensement. Le tableau 9.12 donne un extrait de la série, illustrant une partie des calculs.

TABLEAU 9.11 Estimations du taux global de fécondité (GFR), Cambodge 1993-2008

| Milieu d'année | GFR (15-49) (pour mille) |
|----------------|--------------------------|
| 2007,67 | 82 |
| 2006,67 | 80 |
| 2005,67 | 85 |
| 2004,67 | 93 |
| 2003,67 | 94 |
| 2002,67 | 102 |
| 2001,67 | 106 |
| 2000,67 | 112 |
| 1999,67 | 121 |
| 1998,67 | 106 |
| 1997,67 | 134 |
| 1996,67 | 119 |
| 1995,67 | 156 |
| 1994,67 | 163 |
| 1993,67 | 168 |

Pour chaque année, la multiplication du nombre de femmes dans chaque groupe d'âge par la distribution de fécondité correspondante pour ce groupe d'âge indique ce que serait le nombre d'enfants dans ce groupe d'âge si l'indice synthétique de fécondité était égal à un enfant par femme.

Mais le nombre de naissances annuelles a déjà été estimé (tableau 9.7). Le rapport entre ces estimations et celles faites dans l'hypothèse d'un indice synthétique égal à 1 représente l'estimation de l'indice synthétique de fécondité de chaque année.

TABLEAU 9.10 Nombre de femmes par groupe quinquennal d'âge par année

| Groupe d'âge | 2007,67 | 2006,67 | 2005,67 | 2004,67 | 2003,67 | 2002,67 |
|--------------|---------|---------|---------|---------|---------|---------|
| 10-14 | 812 750 | 806 391 | 800 032 | 793 672 | 787 313 | 776 305 |
| 15-19 | 772 463 | 756 748 | 741 033 | 725 318 | 709 604 | 695 269 |
| 20-24 | 690 645 | 677 616 | 664 587 | 651 558 | 638 528 | 605 814 |
| 25-29 | 600 400 | 548 341 | 496 281 | 444 222 | 392 162 | 374 474 |
| 30-34 | 369 790 | 386 069 | 402 349 | 418 628 | 434 907 | 439 682 |
| 35-39 | 432 454 | 425 603 | 418 752 | 411 901 | 405 050 | 398 453 |
| 40-44 | 390 531 | 384 074 | 377 617 | 371 160 | 364 703 | 356 694 |
| 45-49 | 347 738 | 338 174 | 328 610 | 319 046 | 309 483 | 299 444 |

TABLEAU 9.12 Distributions de fécondité interpolées par groupe quinquennal d'âge et par année

| Groupe d'âge | 2007,67 | 2006,67 | 2005,67 | 2004,67 | 2003,67 | 2002,67 |
|--------------|---------|---------|---------|---------|---------|---------|
| 10-14 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0000 |
| 15-19 | 0,0123 | 0,0126 | 0,0128 | 0,0131 | 0,0133 | 0,0136 |
| 20-24 | 0,0456 | 0,0465 | 0,0474 | 0,0483 | 0,0492 | 0,0502 |
| 25-29 | 0,0496 | 0,0500 | 0,0503 | 0,0506 | 0,0509 | 0,0513 |
| 30-34 | 0,0426 | 0,0423 | 0,0421 | 0,0418 | 0,0416 | 0,0413 |
| 35-39 | 0,0310 | 0,0304 | 0,0299 | 0,0293 | 0,0287 | 0,0282 |
| 40-44 | 0,0158 | 0,0153 | 0,0148 | 0,0143 | 0,0138 | 0,0133 |
| 45-49 | 0,0031 | 0,0029 | 0,0028 | 0,0026 | 0,0024 | 0,0022 |

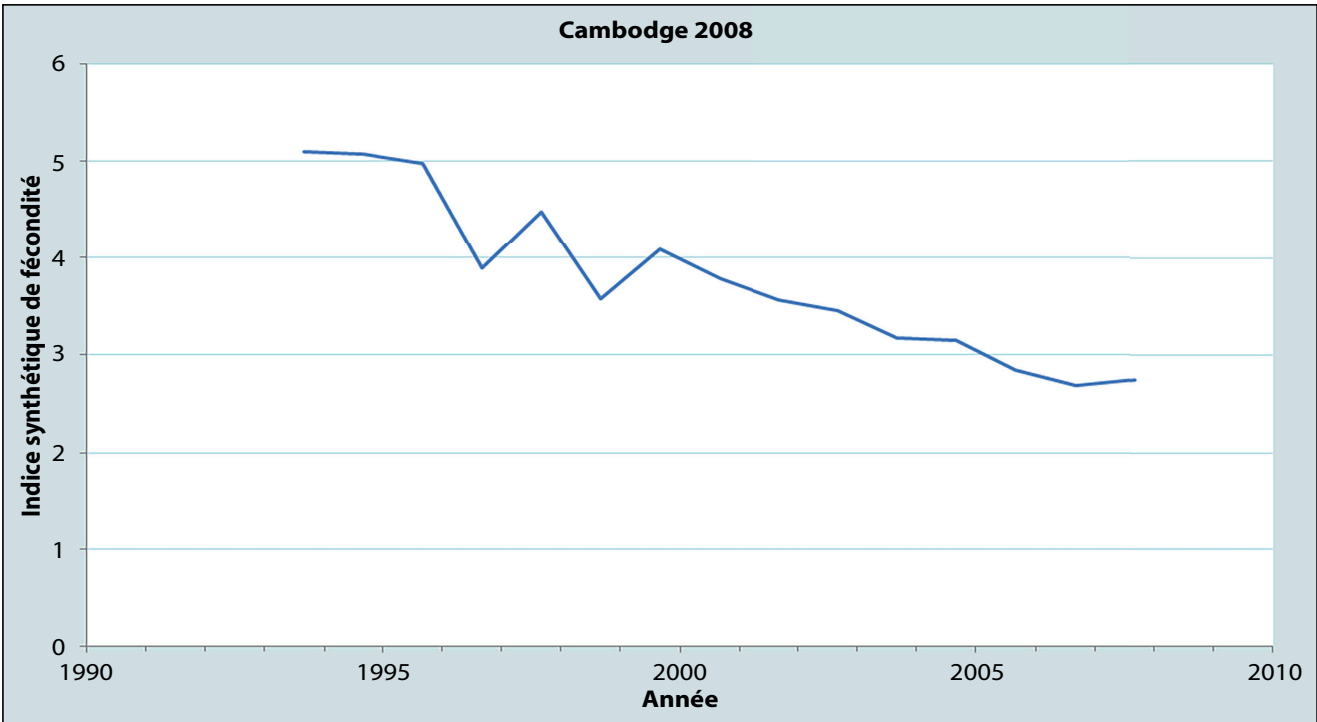
Interprétation

Les résultats de l'application des méthodes de survie inverse aux données du recensement de 2008 au Cambodge sont représentés Figure 9.1. Le graphique fait apparaître clairement le recul de la fécondité au Cambodge au cours des 15 années antérieures à 2008. Alors que le taux de natalité estimé par la méthode 1 augmente légèrement entre deux périodes quinquennales avant le recensement, l'indice synthé-

TABLEAU 9.13 Estimations de l'indice synthétique de fécondité, Cambodge 1993-2007

| Milieu d'année | Indice synthétique |
|----------------|--------------------|
| 2007,67 | 2,76 |
| 2006,67 | 2,69 |
| 2005,67 | 2,85 |
| 2004,67 | 3,15 |
| 2003,67 | 3,18 |
| 2002,67 | 3,46 |
| 2001,67 | 3,56 |
| 2000,67 | 3,78 |
| 1999,67 | 4,09 |
| 1998,67 | 3,58 |
| 1997,67 | 4,48 |
| 1996,67 | 3,89 |
| 1995,67 | 4,98 |
| 1994,67 | 5,07 |
| 1993,67 | 5,10 |

FIGURE 9.1 Évolution de l'indice synthétique de fécondité estimé par projection rétrospective à partir du recensement du Cambodge de 2008



tique de fécondité recule continuellement. On voit aussi sur la figure 9.1 l'impact sur la structure par âge, et par conséquent sur les estimations de fécondité, d'une préférence pour la déclaration de certains âges : 8, 10 et 12, plutôt que 9 ou 11 ans.

En général, les principales difficultés susceptibles de se produire lors de l'estimation de la fécondité par les méthodes de survie inverse ne proviennent ni de la procédure de projection rétrospective elle-même ni des hypothèses sur la distribution par âge de la fécondité, mais d'erreurs dans la déclaration des âges ou de l'omission par les répondants de certains jeunes enfants membres du ménage. Ces omissions et l'exagération des âges des jeunes enfants produisent souvent une fausse impression d'accélération de la baisse de la fécondité dans les quelques années précédant le recensement. Rien ne permet de penser que de telles erreurs de déclarations se sont produites au Cambodge.

Étant donné que dans la plupart des populations contemporaines les décès dans les quinze premières années de vie sont relativement rares, les erreurs dans les hypothèses sur la mortalité des enfants ont un impact modeste sur les estimations de fécondité. Une surestimation de la mortalité des enfants entraîne une surestimation de la fécondité passée et réciproquement. Comme la proportion cumulée d'enfants décédés augmente au fil des âges, les erreurs sur la mortalité affectent davantage les estimations les plus anciennes et créent un léger biais dans les estimations de l'évolution de la fécondité. Il est souvent plus facile de déterminer s'il existe de tels biais quand on peut représenter plusieurs séries d'estimations de la fécondité sur un même graphique, plutôt qu'une seule comme on l'a fait ici.

EXTENSIONS DE LA MÉTHODE

La méthode mère-progéniture est une extension bien établie des méthodes de projection rétrospective pour estimer la fécondité. Une présentation complète de la méthode a été donnée par Cho and Retherford (1986). Elle conduit aussi à des estimations de l'indice synthétique de fécondité, mais au lieu de s'appuyer sur des estimations indépendantes de la distribution par âge de la fécondité pour répartir les naissances entre les mères elle cherche à relier chaque enfant recensé à sa mère. Ceci permet de classer les enfants à la fois selon leur âge et celui de leur mère. Ces effectifs peuvent ensuite être rétro-projetés pour déterminer le nombre d'enfants nés par année selon l'âge de leur mère à leur naissance.

Dans les enquêtes où des informations détaillées ont été recueillies sur les relations familiales au sein du ménage, et en particulier si un identifiant des mères qui vivent avec leurs enfants a été porté sur l'enregistrement de chaque enfant, un exercice d'appariement est possible, même s'il passe par un algorithme complexe. Des hypothèses sont nécessaires concernant l'âge des mères d'enfants adoptifs et des orphelins de mère et, s'ils sont nombreux, il peut être problématique de supposer qu'ils peuvent être répartis par groupe d'âge des mères au prorata.

Si l'information sur les relations au sein du ménage est sommaire, l'appariement des enfants et de leur mère peut devenir problématique, surtout dans les populations où les ménages regroupent souvent des familles étendues. En particulier, si la mère ne réside pas dans le ménage, les enfants risquent d'être rattachés à tort à une tante, une grand-mère ou une autre femme plutôt que de rester sans lien maternel. Ces erreurs ont généralement pour conséquence d'exagérer la fécondité estimée des femmes âgées. Les divorces, les remariages et les mariages polygames accroissent aussi le risque d'identifier à tort une femme comme mère d'un enfant.

Nous ne pouvons pas donner de conseil ici sur les algorithmes détaillés qu'il faudrait utiliser pour rattacher les enfants à leur mère quand on applique la méthode des enfants du ménage pour estimer la fécondité, car la procédure adoptée doit être exactement adaptée aux questions qui ont été posées et aux procédures de travail sur le terrain. Par exemple, l'appariement est plus aisé si on a demandé aux enquêteurs de lister ensemble les membres de chaque unité familiale restreinte, plutôt que de regrouper les membres du ménage par âge ou par genre.

Comme un nombre important de jeunes gens quittent le foyer à la fin de leur adolescence (y compris, dans la plupart des cultures, les jeunes femmes qui se marient), la recommandation de n'appliquer les méthodes de projection rétrospective qu'à l'estimation de la fécondité des 15 années précédant le recensement vaut à plus forte raison encore pour les estimations obtenues par la méthode des enfants du ménage.

Malgré ces éventuelles restrictions, l'expérience montre que la méthode mère-progéniture peut donner des estimations de fécondité de même qualité que celles calculées à partir de données d'histoire génésique. En fait, comme elles évitent le problème de troncature des taux de fécondité des femmes âgées à mesure que les estimations remontent dans le temps et comme les données d'histoire génésique doivent être recueillies auprès des femmes en personne et que celles

qui ne peuvent pas être interrogées ont en moyenne une fécondité moindre que les autres, certains analystes pensent que les estimations produites par la méthode des enfants du ménage sont meilleures que les estimations directes tirées d'histoires génésiques dans la plupart des cas (Avery, St. Clair, Levin *et al.* 2013).

Par ailleurs, les méthodes de projection rétrospective décrites ici en détail fonctionnent généralement bien au niveau national. Elles ne nécessitent pas un accès à des données individuelles et elles peuvent aboutir à des estimations plus précises que la méthode mère-progéniture dans les cas où les appariements erronés des enfants à leur mère sont fréquents.

La méthode mère-progéniture est cependant particulièrement appropriée pour les études de fécondité différentielle selon les caractéristiques de la mère comme son instruction, sa résidence ou son état matrimonial. Comme elle peut être appliquée à des données de recensement, la méthode peut permettre d'estimer la fécondité de sous-groupes trop peu nombreux pour être étudiés par des données d'enquête de fécondité. En outre, comme les informations sur les caractéristiques des femmes sont généralement beaucoup plus nombreuses dans les recensements que sur les bulletins de naissance, la méthode reste valable même dans les pays où l'enregistrement des naissances est complet.

RÉFÉRENCES

- Avery C, T St. Clair, M Levin and K Hill. 2013. "The 'Own Children' fertility estimation procedure: A reappraisal", *Population Studies* **67**(2): 171-183. <http://dx.doi.org/10.1080/00324728.2013.769616>
- Brass W. 1979. "Evaluation of birth and death registration using age distributions and child survivorship data", *Asian and Pacific Census Forum* **5**(3):9-11,20. <http://hdl.handle.net/10125/3563>
- Cho LJ, RD Retherford and MK Cho. 1986. *The Own-Child Method of Fertility Estimation*. Honolulu: University of Hawaii Press.
- UN Population Division. 1983. *Manual X: Indirect Techniques for Demographic Estimation*. New York: United Nations, Department of Economic and Social Affairs, ST/ESA/SER.A/81. <http://www.un.org/esa/population/techcoop/DemEst/manual10/manual10.html>
- Division de la Population des Nations Unies. 1984. *Manuel X. Techniques indirectes d'estimation démographique*. New York : Nations Unies, Département des affaires économiques et sociales internationales, ST/ESA/SER.A/81. http://unstats.un.org/unsd/demographic/standmeth/handbooks/Manuel_X-fr.pdf

Méthodes fondées sur plusieurs recensements

CHAPITRE 10 MODÈLES RELATIONNELS

SYNTHÉTIQUES DE GOMPERTZ

Tom A Moultrie

DESCRIPTION DE LA MÉTHODE

Le modèle relationnel synthétique de Gompertz est une extension de la méthode relationnelle de Gompertz pour l'estimation des taux par âge et de l'indice synthétique de fécondité. Il s'appuie sur deux séries des données de parité, recueillies en différents points du temps, et sur des estimations de la fécondité du moment pour la période intermédiaire faites à partir des déclarations sur les naissances récentes classées par âge de la mère.

La méthode tient compte explicitement des variations de la fécondité ; elle est conçue pour des recensements ou des enquêtes réalisés à 5 ou 10 ans d'écart. Dans ces cas, les survivantes d'une cohorte de femmes observées à la première date peuvent être identifiées à la seconde, et la variation de la parité moyenne de la cohorte peut être calculée. Il en résulte, pour différentes cohortes, entre les deux dates, une série d'accroissements de parité qui peuvent être cumulés, déterminant ainsi les parités moyennes d'une cohorte fictive qui connaîtrait la fécondité représentée par les accroissements de parité observés.

Quand de telles données ne sont pas aisément disponibles, ou qu'elles ne sont pas fiables, une procédure plus simple, et généralement acceptable, consiste à calculer des taux de fécondité par âge pour la première et la dernière année de la période et à estimer les taux de la période entière comme moyenne arithmétique de ces deux séries. Si on ne dispose pas d'enregistrement des naissances, mais que les deux enquêtes ou les deux recensements ont collecté des données sur les naissances de l'année écoulée, les taux de fécondité par âge peuvent être approximés de la même façon, en faisant la moyenne des taux observés au début et à la fin de la période. Si les naissances au cours des 12 mois précédant chaque enquête sont clas-

sées par âge de la mère au moment de l'enquête, les taux de fécondité observés correspondent à des groupes d'âge décalés de six mois. L'analyse devra en tenir compte.

Une fois les parités et les taux de fécondité correspondants calculés pour la période entre les deux enquêtes, le cumul et l'interpolation des taux et leur comparaison avec les parités moyennes sont réalisés exactement comme il a été décrit dans la présentation du modèle relationnel de Gompertz conventionnel.

DONNÉE NÉCESSAIRES

Les données nécessaires sont :

- Le nombre d'enfants déjà nés classés par groupe quinquennal d'âge de la mère, tiré des deux enquêtes ou des deux recensements réalisés à cinq ou dix ans d'écart.
- SOIT le nombre de naissances au cours de l'année précédant chaque recensement classées par groupe quinquennal d'âge de la mère, SOIT les naissances enregistrées par groupe d'âge de la mère pour chaque année de la période intercensitaire. Si on ne dispose pas des données sur les naissances classées par âge de la mère aux points terminaux de la période intercensitaire, on peut utiliser un schéma approprié de fécondité par âge se référant à peu près au milieu de la période.
- Le nombre de femmes dans chaque groupe quinquennal d'âge aux deux enquêtes ou aux deux recensements.
- La population totale à chaque enquête ou recensement, si on doit calculer le taux brut de natalité, ou si on doit évaluer la complétude relative des données tirées du système d'enregistrement à l'état civil.

HYPOTHÈSES

La plupart des hypothèses sont celles sur lesquelles repose le modèle relationnel de Gompertz :

- Le schéma de fécondité standard choisi pour la procédure d'ajustement reflète de façon appropriée la forme de la distribution de la fécondité dans la population.
- Les changements intercensitaires de la fécondité ont été réguliers et graduels et ont affecté tous les groupes d'âge de manière comparable.
- Les erreurs dans les taux de fécondité avant ajustement sont proportionnellement les mêmes parmi les femmes dans les groupes d'âge centraux (20-39 ans), de sorte que la forme de la fécondité par âge des naissances récentes déclarées est raisonnablement correcte.
- Les parités déclarées par les jeunes femmes (20-29 ans) sont conformes à la réalité.

Le calcul des parités moyennes dans la cohorte synthétique suppose que la mortalité et la migration sont sans effet sur les distributions de parité effectives. En d'autres termes, on suppose que la parité moyenne des femmes qui décèdent ou qui migrent entre les deux enquêtes ne diffère pas significativement de la parité moyenne aux mêmes âges des femmes encore survivantes et présentes en fin de période.

TRAVAIL PRÉPARATOIRE ET RECHERCHES PRÉLIMINAIRES

Avant d'entreprendre l'analyse des niveaux de fécondité en recourant à cette méthode, l'analyste doit vérifier la qualité des données au moins sur les points suivants :

- structure par âge et sexe de la population ;
- naissances déclarées au cours de l'année écoulée, et
- parités moyennes et nécessité d'une correction d'El-Badry.

PRÉCAUTIONS ET MISES EN GARDE

Il est essentiel que les séries de taux de fécondité par âge dont on doit prendre la moyenne soient exprimées dans une même classification des âges. Si ce n'est pas le cas initialement, parce qu'une série se réfère à des groupes d'âge décalés de six mois et pas l'autre, la première doit être préalablement ajustée (par exemple, en appliquant la variante F-seul du modèle relationnel de Gompertz). En règle générale, on doit éviter de combiner des estimations des taux de fécondité par âge tirées de différentes sources (par exemple état civil et recensement), car les schémas de fécondité risquent d'être déformés de façon différente.

Si on ne dispose pas de taux de fécondité par âge pour le début et la fin de la période, on peut utiliser une série de taux se référant à peu près au milieu de celle-ci. On doit se souvenir que, dans l'application de la méthode relationnelle de Gompertz, seule importe la distribution des taux par âge

de la période intercensitaire, de sorte que si cette distribution est plus ou moins constante dans cet intervalle, la date de référence exacte des taux utilisés importe peu.

Si on utilise des données de naissances enregistrées, des variations dans la complétude des données par groupe d'âge au fil du temps peuvent déformer la distribution de la fécondité. Si c'est le cas, la méthode doit être appliquée avec circonspection.

APPLICATION DE LA MÉTHODE

L'application de la méthode se fait selon les étapes suivantes.

Étape 1 : Calcul des parités moyennes déclarées

Calculer les parités moyennes, ${}_5P_x(t_1)$ et ${}_5P_x(t_2)$, des femmes de chaque groupe d'âge $[x, x+5)$ aux dates des deux recensements (t_1 et t_2), pour $x = 15, 20 \dots 45$. Pour faciliter l'exposé, nous désignons la parité moyenne dans chaque groupe d'âge i au temps t par $P(i, t) = {}_5P_x(t)$ où $i = (x/5 - 2)$. Les parités moyennes tirées du premier recensement ou enquête sont donc notées $P(i, 1)$ et celles tirées du second par $P(i, 2)$.

Étape 2 : Calcul des parités moyennes pour une cohorte fictive

Le mode de calcul des parités dépend de la longueur de l'intervalle intercensitaire.

a) Intervalle de cinq ans

Si l'intervalle entre les deux séries de données est de cinq ans, les survivantes du groupe d'âge i au premier recensement appartiennent au groupe d'âge $i+1$ au deuxième, et l'accroissement de parité entre les deux recensements est égal à $P(i+1, 2) - P(i, 1)$. Ces accroissements peuvent être calculés pour chaque groupe d'âge, et les parités de la cohorte fictive sont alors obtenues par cumul des accroissements successifs. Ainsi, si on désigne par $\Delta P(i+1)$ l'accroissement de parité pour la cohorte du groupe d'âge i aux premier recensement, et par $P(i, s)$ la parité du groupe d'âge i dans la cohorte fictive (où s renvoie à « synthétique »), on a $\Delta P(i+1) = P(i+1, 2) - P(i, 1)$ pour $i = 1, \dots, 6$, et donc

$$P(i, s) = \sum_{j=1}^i \Delta P(j)$$

On considère que l'accroissement de parité $\Delta P(i+1)$ pour le groupe d'âge le plus jeunes ($i = 0$) est égal à $P(1, 2)$,

en supposant que $P(0,1)$, parité moyenne des femmes de 10-14 ans au premier recensement, est nulle. Si la fécondité évolue rapidement, cette valeur de $\Delta P(1)$ reflète donc des taux du moment plus proches de l'enquête que du milieu de l'intervalle, donnant un poids un peu exagéré à l'évolution de la fécondité.

b) Intervalle de dix ans

Si la période entre les recensements ou les enquêtes est de 10 ans, les survivantes de la cohorte appartenant au groupe d'âge i lors de la première enquête seront dans le groupe d'âge $(i + 2)$ lors de la seconde. Les parités de la cohorte fictive sont alors obtenues en cumulant deux séries parallèles d'accroissements de parité. Une fois encore, pour les groupes d'âge les plus jeunes, on considère que $\Delta P(1)$ est égal à $P(1,2)$ et que $\Delta P(2)$ est égal à $P(2,2)$. Les autres accroissements de parité sont égaux à $\Delta P(i + 2) = P(i + 2, 2) - P(i, 1)$ pour $i = 1, \dots, 5$.

Les parités de la cohorte fictive pour les groupes d'âge ayant un numéro pair sont obtenues en cumulant les accroissements de parité pour les groupes d'âge ayant un numéro pair, alors que ceux pour les groupes d'âge ayant un numéro impair sont obtenus en cumulant les accroissements de parité pour les groupes d'âge ayant un numéro impair. Donc

$$\begin{aligned} P(1,s) &= \Delta P(1) = P(1,2) \\ P(2,s) &= \Delta P(2) = P(2,2) \\ P(3,s) &= \Delta P(1) + \Delta P(3) \\ P(4,s) &= \Delta P(2) + \Delta P(4) \\ P(5,s) &= \Delta P(1) + \Delta P(3) + \Delta P(5) \\ P(6,s) &= \Delta P(2) + \Delta P(4) + \Delta P(6) \\ P(7,s) &= \Delta P(1) + \Delta P(3) + \Delta P(5) + \Delta P(7) \end{aligned}$$

Étape 3 : Calcul des taux de fécondité actuels

La méthode de calcul des taux, notés $f(i)$, où i désigne les groupes d'âge comme précédemment, dépend des données disponibles.

a) Données tirées d'un système d'enregistrement à l'état civil

Une procédure possible consiste à calculer les taux de fécondité par âge se rapportant à peu près à la première et à la dernière année de la période entre les deux recensements en utilisant les données sur le nombre de naissances déclarées l'année précédant chaque recensement. Dans ce cas, on

divise le nombre de naissances déclarées pour chaque groupe d'âge de la mère par le nombre de femmes recensées au même âge. On obtient ensuite les taux de fécondité par âge pour la période intercensitaire en faisant la moyenne arithmétique des taux par âge des deux années.

Une autre méthode consiste à calculer un taux moyen de fécondité par âge pour l'ensemble de la période à partir des taux de chaque année si ceux-ci sont disponibles à partir de l'état civil. Il s'agit alors de faire la somme des naissances déclarées pour chaque groupe d'âge de la mère et de la diviser par le nombre de personnes-années vécues (en faisant la moyenne de l'effectif des groupes d'âge au début et à la fin de l'intervalle, et en la multipliant par le nombre d'années au cours de la période intercensitaire).

Les taux de fécondité par âge obtenus à partir de l'état civil sont, par définition, classés par âge de la mère au moment de son accouchement.

b) Données tirées des recensements (enquêtes) ayant servi au calcul des parités moyennes à l'étape 2

Si les données sont tirées des déclarations que les femmes ont faites des naissances qu'elles ont eues au cours de l'année précédant le recensement qui a servi à calculer les parités moyennes, la moyenne arithmétique des deux séries de taux est à nouveau utilisée comme estimation de la fécondité de la période intercensitaire. Toutefois, la distribution des taux de fécondité ainsi calculés résulte d'une classification des mères selon leur âge au recensement et non selon leur âge à la naissance de l'enfant, en moyenne six mois auparavant. On doit tenir compte de ce décalage dans les âges lors de l'application du modèle relationnel de Gompertz.

Le processus d'ajustement du modèle relationnel de Gompertz aux données a déjà été décrit dans la section consacrée au modèle. Les seuls points de différence sont les suivants :

- Les estimations s'appliquent au point médian de la période, c'est-à-dire 2,5 ou 5 ans avant le second recensement.
- La feuille de calcul porte seulement sur l'application conventionnelle du modèle relationnel de Gompertz, les parités servant à fixer le niveau et la distribution de la fécondité pour la période servant à déterminer la forme de la courbe de fécondité.
- Si les données sont classées par âge de la mère à la date du recensement (c'est-à-dire que les données sur la fécondité récente ont été obtenues à partir du recensement ou de l'enquête qui a aussi fourni les parités moyennes, et non

à partir de l'état civil), le dossier Excel associé disponibles sur <http://demographicestimation.iussp.org/fr> vaut seulement pour le cas où les données sur la fécondité récente couvrent les 12 mois précédant le recensement ou l'enquête.

Les étapes pertinentes sont rappelées ci-dessous.

Étape 4 : Choix du standard de fécondité à appliquer dans le modèle

Par défaut, le standard de fécondité est celui de Booth, légèrement modifié par Zaba (1981). Il est approprié dans le cas des populations à fécondité moyenne et élevée ; il est exprimé sous forme d'un schéma normalisé de fécondité cumulée (c'est-à-dire avec une fécondité totale égale à un). Les valeurs standards de $Y^s(x)$ sont déterminées en prenant les gompits de la distribution. Les valeurs des parités standards, $Y^s(i)$, sont les gompits des parités associées au schéma standard de fécondité. Le choix du standard détermine les valeurs de $g()$ et $e()$ utilisées dans les procédures d'ajustement par régression.

Étape 5 : Évaluer la représentation graphique des P -points et des F -points

Les représentations graphiques de $z(x) - e(x)$ en fonction de $g(x)$, et $z(i) - e(i)$ en fonction de $g(i)$ sur le même système d'axes permettent de faire apparaître des erreurs communes et des tendances dans les données, comme on l'a vu en présentant le modèle relationnel de Gompertz.

Étape 6 : Ajustement du modèle par sélection des points à utiliser

Au départ, tous les points doivent être inclus dans le modèle, sauf si la parité moyenne dans un groupe d'âge est supérieure à celle dans le groupe d'âge suivant. Dans ce cas, le gompit n'est pas défini et le modèle ne peut pas être ajusté en utilisant ce point (une telle situation ne peut pas survenir dans une cohorte réelle, mais elle peut être observée à cause d'erreurs dans les données ou dans une cohorte synthétique au cours d'une période où la fécondité évolue rapidement).

Si les données de parité et de fécondité sont mutuellement cohérentes, les représentations graphiques des $z() - e()$ en fonction de $g()$ doivent être alignées. Les P -points et les F -points qui font s'écarter la représentation d'une droite doivent être exclus du modèle. Une régression par les moindres carrés ordinaires est utilisée pour ajuster des

droites sur les P -points et les F -points et pour identifier, séquentiellement, les points qui ne se situent pas nettement sur une droite. Il s'agit de chercher la combinaison la plus large de P - et F -points (à peu près) alignés, et d'utiliser ceux-ci pour ajuster le modèle.

On retient ou on exclut des points en respectant les instructions suivantes :

- Une série de points contigus doit être incluse dans le modèle. Seuls des points extrêmes peuvent être exclus. (La raison est que chaque point sur le graphique est le résultat d'un calcul de ratio entre les valeurs de données adjacentes. Si l'analyse conduit à considérer une donnée comme non fiable en tant que dénominateur, il n'est pas logique de l'accepter comme numérateur du ratio suivant.)
- Il vaut mieux éliminer des P -points que des F -points, car les données de parité moyenne sont davantage sujettes à des erreurs liées à l'âge que les données de fécondité.
- Les P -points qui s'écartent de l'alignement construit sur les autres P -points, ainsi que les F -points qui s'écartent de l'alignement construit sur les autres F -points, doivent être éliminés dès le début du processus d'ajustement.
- Il vaut mieux éliminer les P - et les F -points aux âges avancés plutôt qu'aux jeunes âges car les données y sont généralement moins fiables et présentent moins de cohérence entre descendance et fécondité récente. Les données relatives aux femmes de moins de 20 ans font cependant exception, car elles reposent sur des nombres faibles d'événements, ce qui rend souvent peu fiables les estimations des parités moyennes ou de la fécondité cumulée.
- Un ajustement réalisé sur davantage de points est préférable à celui réalisé sur moins de points, même s'il est légèrement moins bon. La feuille de calcul produit l'écart quadratique moyen (*root mean squared error* : *RMSE*)

$$RMSE = \sqrt{\frac{\sum \left((z() - e()) - \left(\alpha + (\beta - 1)^2 \frac{c}{2} + \beta g() \right) \right)^2}{n}}$$

à partir des points utilisés pour ajuster le modèle. Cette statistique peut aider à déterminer le nombre optimal de points sur lesquels réaliser l'ajustement, s'il y a un doute dans le choix à faire entre deux modèles concurrents. Dans ce cas, on peut choisir celui qui a le plus faible écart quadratique.

Étape 7 : Évaluer la qualité des paramètres estimés

Les valeurs de α et β , qui représentent la droite de meilleur ajustement joignant les P -points et les F -points restants, nécessitent un contrôle pour vérifier qu'elles ne s'éloignent pas de leurs valeurs centrales de manière telle que le choix du standard serait mis en cause. Un bon ajustement est obtenu si $-0,3 < \alpha < 0,3$, et si $0,8 < \beta < 1,25$.

Si les paramètres se situent en dehors de ces intervalles, une au moins des séries de données sous-jacentes pose problème ou le standard est inadéquat. Il convient alors d'essayer un autre standard (voir plus loin) ou de modifier la sélection des points avant de poursuivre. Si les paramètres restent en dehors des intervalles ci-dessus, il faut considérer que la méthode est inappropriée.

Étape 8 : Taux de fécondité par âge et indice synthétique de fécondité ajustés

Une fois estimés les deux paramètres du modèle, ceux-ci peuvent être appliqués au standard des parités pour fournir des valeurs ajustées, $Y(i) = \alpha + \beta.Y^s(i)$. Celles-ci sont ensuite ramenées à des mesures de la proportion cumulée de fécondité réalisée par groupe d'âge i , grâce à la transformation anti-gompit. Les anti-gompits fondés sur les distributions de parité indiquent la proportion de fécondité réalisée dans le groupe d'âge. En divisant la parité observée dans chaque groupe d'âge par ces proportions, on obtient une série d'estimations de l'indice synthétique de fécondité. La moyenne de ces valeurs calculée sur le sous-ensemble des groupes d'âge qui ont été utilisés pour estimer α et β donne l'estimation ajustée de l'indice synthétique de fécondité, \hat{T} . En appliquant les mêmes α et β aux gompits standards pour les âges qui bornent les groupes d'âge conventionnels (c'est-à-dire 20, 25... 50 ans), en appliquant la transformation anti-gompit, et en multipliant par \hat{T} , on obtient une distribution de fécondité cumulée aux différents âges. En faisant la différence entre les estimations successives de la fécondité cumulée et en la divisant par cinq, on obtient un schéma de fécondité ajusté pour les groupes d'âge conventionnels (15-19, 20-24 etc.), même si le classement des données initiales englobait un décalage d'une demi-année.

EXEMPLE

L'exemple utilise les données recueillies lors des deux recensements du Kenya, réalisés à dix ans d'écart en 1989 et 1999. Les deux recensements incluent des questions sur la descen-

dance et sur les naissances au cours de l'année passée. La méthode est mise en application dans le dossier Excel associé disponibles sur le site web.

Étape 1 : Calcul des parités moyennes déclarées

Une correction d'El-Badry a été appliquée aux données du recensement de 1989. De toute évidence, les résultats du recensement de 1999 ont été corrigés : il n'y a pas de données manquantes sur la parité. Les parités moyennes tirées des deux recensements figurent aux deux premières colonnes du tableau 10.1. La descendance des femmes âgées a reculé d'environ 0,6 enfant au cours de la décennie. Mais l'accroissement de la descendance des adolescentes est un peu surprenant.

Étape 2 : Calcul des parités moyennes pour une cohorte fictive

La période intercensitaire est de 10 ans (entre 1989 et 1999). Nous suivons donc la procédure décrite à l'étape 2(b) pour calculer les parités moyennes des cohortes, qui figurent dans la dernière colonne du tableau 10.1.

TABLEAU 10.1 Parités moyennes par groupe d'âge, Kenya, recensements de 1989 et 1999

| Groupe d'âge | 1989 | 1999 | Parité de la cohorte fictive $P(i,s)$ |
|--------------|--------|--------|---------------------------------------|
| 15-19 | 0,2416 | 0,2848 | 0,2848 |
| 20-24 | 1,5247 | 1,3640 | 1,3640 |
| 25-29 | 3,2138 | 2,6073 | 2,6505 |
| 30-34 | 4,7602 | 4,1432 | 3,9825 |
| 35-39 | 6,2390 | 5,3867 | 4,8234 |
| 40-44 | 7,1204 | 6,3818 | 5,6041 |
| 45-49 | 7,5103 | 6,9143 | 5,4987 |

Comme indiqué à cette étape, $\Delta P(1) = P(1,2) = 0,2848$ et $\Delta P(2) = P(2,2) = 1,3640$, alors que

$$\begin{aligned} P(5,s) &= \Delta P(1) + \Delta P(3) + \Delta P(5) \\ &= 0,2848 + (2,6073 - 0,2416) + (5,3867 - 3,2138) \\ &= 4,8234 \end{aligned}$$

On voit aisément que d'importantes omissions ont dû affecter les parités aux âges élevés puisque la parité dans la cohorte fictive à l'âge le plus élevé est inférieure à la parité des femmes dans la cohorte fictive intercensitaire âgée de 40-44 ans.

Étape 3 : Calcul des taux de fécondité actuels

Les données disponibles sont issues des déclarations par les femmes du mois et de l'année de leur dernière naissance au cours de l'année précédant chaque recensement. Comme on l'a vu dans la section sur l'évaluation de la qualité des données sur la fécondité récente, ces déclarations peuvent être converties en estimations des taux par âge et de l'indice synthétique de fécondité, en supposant que toutes les naissances déclarées le mois du recensement ont eu lieu avant la date de celui-ci et en répartissant au prorata les naissances déclarées dans le mois un an avant la date du recensement. En procédant ainsi on obtient les estimations directes des taux par âge et de l'indice synthétique de fécondité présentées au tableau 10.2. Dans la dernière colonne, l'estimation de la fécondité intercensitaire est la moyenne des taux pour 1989 et 1999 dans chaque groupe d'âge.

On notera que la qualité des déclarations de fécondité dans les deux recensements est médiocre. Ces données impliquent des niveaux de fécondité nettement plus bas que ceux tirés des parités des cohortes synthétiques ou que la valeur de l'indice synthétique de fécondité de 5,3 enfants par femme dans l'Enquête Démographique et de Santé réalisée au Kenya en 1993.

TABLEAU 10.2 Estimations directes des taux par âge et de l'indice synthétique de fécondité, Kenya, recensements de 1989 et 1999

| Groupe d'âge | 1989 | 1999 | Fécondité moyenne |
|---------------------------------|--------|--------|-------------------|
| 15-19 | 0,0679 | 0,1107 | 0,0893 |
| 20-24 | 0,2179 | 0,2381 | 0,2280 |
| 25-29 | 0,2309 | 0,2124 | 0,2217 |
| 30-34 | 0,1908 | 0,1728 | 0,1818 |
| 35-39 | 0,1458 | 0,1193 | 0,1326 |
| 40-44 | 0,0764 | 0,0583 | 0,0673 |
| 45-49 | 0,0351 | 0,0203 | 0,0277 |
| Indice synthétique de fécondité | 4,82 | 4,66 | 4,74 |

Étape 4 : Choix du standard de fécondité à appliquer dans le modèle

Par défaut, le standard de fécondité est celui de Booth, légèrement modifié par Zaba (1981). Il n'y a pas d'autre standard de fécondité féminine validé.

Étape 5 : Évaluer la représentation graphique des P-points et des F-points

Nous commençons en ajustant des modèles sur l'ensemble de P-points et des F-points. Les résultats figurent sur la feuille des **Graphiques diagnostiques** du dossier Excel associé disponibles sur <http://demographicestimation.iussp.org/fr>.

Étape 6 : Ajustement du modèle par sélection des points à utiliser

En suivant les instructions rappelées plus haut, nous éliminons des points les uns après les autres du modèle afin d'obtenir une plus grande congruence des P-points et des F-points. Le meilleur ajustement est réalisé avec les P-points de 20-39 ans et les F-points de 20-44 ans (Figure 10.1).

Étape 7 : Évaluer la qualité des paramètres estimés

Dans cette application, les valeurs ajustées de α (– 0,0286) et β (1,0042) sont bien à l'intérieur de l'intervalle des valeurs admissibles.

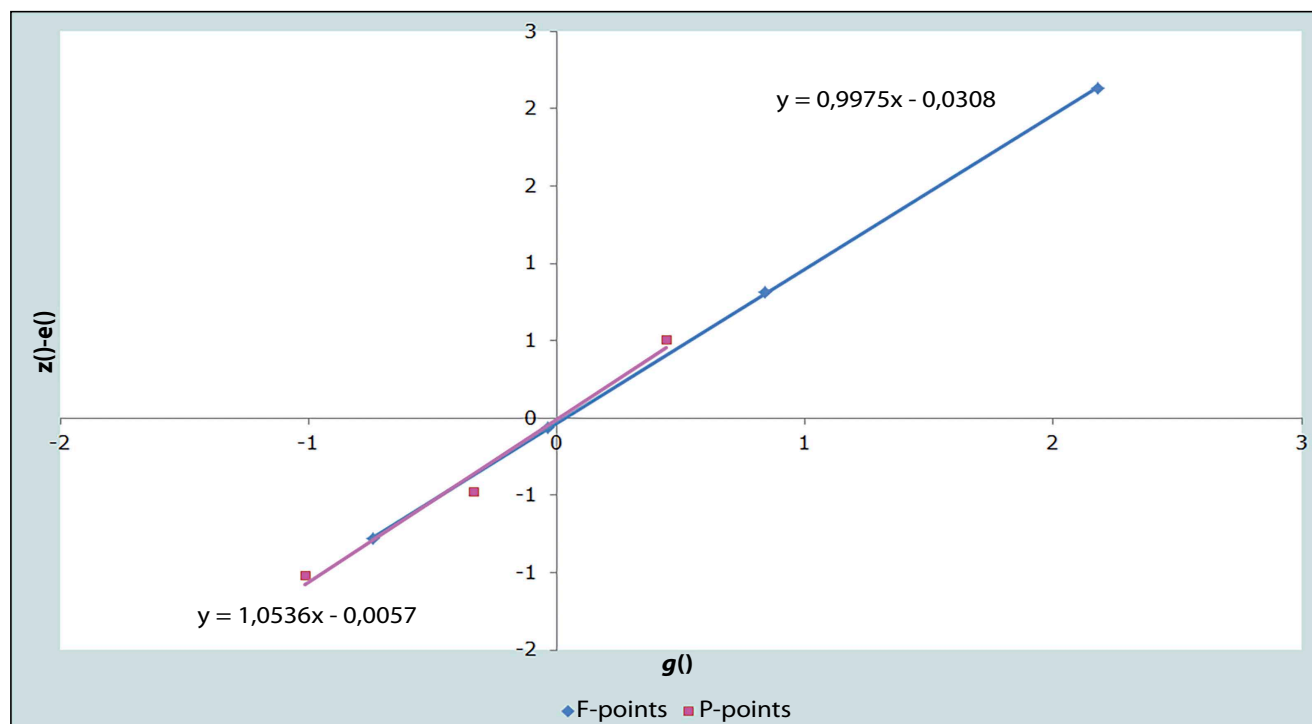
Étape 8 : Taux de fécondité par âge et indice synthétique de fécondité ajustés

L'indice synthétique de fécondité tiré du modèle ajusté est de 5,56 enfants par femme (Tableau 10.3) ; Il vaut approximativement pour août 1994, le modèle ayant tenu compte du décalage dû au classement des âges des mères. Ce niveau est à peu près cohérent avec l'estimation de 5,3 enfants par femme tirée de l'EDS du Kenya en 1993, et avec les estimations obtenues par application de la méthode relationnelle de Gompertz séparément à chaque série de données.

TABLEAU 10.3 Taux de fécondité par âge estimés sur la base des accroissements de parité dans une cohorte fictive, Kenya 1989-1999

| Groupe d'âge | Taux de fécondité par âge |
|---------------------------------|---------------------------|
| 15-19 | 0,139 |
| 20-24 | 0,267 |
| 25-29 | 0,261 |
| 30-34 | 0,213 |
| 35-39 | 0,153 |
| 40-44 | 0,070 |
| 45-49 | 0,009 |
| Indice synthétique de fécondité | 5,56 |

FIGURE 10.1 Représentation de $z() - e()$ en fonction de $g()$ après élimination de certains points, cohortes synthétiques fondées sur les données des recensements de 1989 et 1999, Kenya



DESCRIPTION DÉTAILLÉE DE LA MÉTHODE

La méthode décrite ici est une variante du modèle relationnel de Gompertz qui, au lieu d'utiliser des données de parité et de fécondité recueillies en un point du temps, élabore une distribution de fécondité « moyenne » à partir des déclarations de descendance et de fécondité récente en deux points du temps. L'exposé mathématique du modèle relationnel de Gompertz a été fait précédemment.

AUTRES LECTURES ET RÉFÉRENCES

La méthode a d'abord été décrite par Zlotnik et Hill (1981) puis reprise aux pages 41-46 du *Manuel X* (Division de la Population des NU 1984). La présente mise à jour reste conforme à l'original, à ceci près qu'elle est présentée comme une variante du modèle relationnel de Gompertz où

les parités utilisées sont les parités intercensitaires tirées des deux enquêtes, et les taux de fécondité sont les estimations pour la période séparant les deux recensements.

Division de la Population des NU. 1984. *Manuel X. Techniques indirectes d'estimation démographique*. New York : Nations Unies, Département des affaires économiques et sociales internationales, ST/ESA/SER.A/81. http://unstats.un.org/unsd/demographic/standmeth/handbooks/Manuel_X-fr.pdf

Zaba B. 1981. *Use of the Relational Gompertz Model in Analysing Fertility Data Collected in Retrospective Surveys*. Centre for Population Studies Research Paper 81-2. London: Centre for Population Studies, London School of Hygiene & Tropical Medicine.

Zlotnik H and KH Hill. 1981. "The use of hypothetical cohorts in estimating demographic parameters under conditions of changing fertility and mortality", *Demography* **18**(1):103-122. doi: <http://dx.doi.org/10.2307/2061052>

CHAPITRE I I ESTIMATION DE LA FÉCONDITÉ À PARTIR DES ACCROISSEMENTS DE PARITÉ DES COHORTES

Tom A Moultrie

DESCRIPTION DE LA MÉTHODE

Les tableaux sur le nombre d'enfants déjà nés par groupe quinquennal d'âge classique des femmes pour un seul recensement ou une seule enquête donnent une information abondante sur la fécondité passée des femmes. Malheureusement, si la fécondité a évolué, les parités moyennes des femmes des différents groupes d'âge ne permettent pas d'estimer le profil par âge de la fécondité, ni par cohorte, ni par période.

Si l'information sur le nombre d'enfants déjà nés existe dans deux recensements ou enquêtes à cinq ou dix ans d'écart environ, la variation du nombre moyen d'enfants déjà nés dans une cohorte de femmes reflète leur fécondité au cours de la période intercensitaire. Il devient alors possible d'estimer la distribution de la fécondité par âge pour cette période. Arretx (1973) a développé une méthode pour utiliser une telle information quand les deux sources sont séparées de 10 ans. Le *Manuel X* (Division de la Population des NU 1984) présente une variante d'une démarche proposée par Coale et Trussell recourant au quotient P/F. Nous proposons une amélioration par rapport à la présentation du *Manuel X*, en nous appuyant sur le modèle relationnel de Gompertz.

La méthode consiste à estimer les taux moyens de fécondité par âge de la période intercensitaire, en construisant les parités moyennes d'une cohorte fictive intercensitaire. Une distribution de fécondité cumulée est alors calculée par interpolation à partir de ces parités, et les taux de fécondité par âge sont obtenus à partir de la fécondité cumulée grâce à une suite de soustractions.

La méthode est conçue pour des situations où on peut calculer des parités moyennes par groupe d'âge des femmes en deux points du temps à peu près séparés de dix ans. Si l'intervalle entre les deux recensements est de cinq ans, les

femmes d'un groupe quinquennal d'âge à la seconde date sont les survivantes des femmes dans le groupe quinquennal cinq ans plus jeune à la première date. L'accroissement de parité moyenne de la cohorte entre le premier et le second recensement reflète sa fécondité entre les deux recensements, si on peut supposer que la descendance des femmes qui sont décédées ou qui ont migré entre temps n'était pas, en moyenne, systématiquement différente de celle des femmes qui sont restées. En cumulant les accroissements de parité, on peut estimer les parités moyennes d'une cohorte fictive qui aurait, au long de sa vie féconde supposée, les taux de fécondité par âge de la période intercensitaire. Si cette période est de 10 ans, un groupe quinquennal à la seconde date est composé des survivantes des femmes du groupe quinquennal dix ans plus jeunes à la première date. Dans ce cas, il est encore possible de calculer l'accroissement de parité de chaque cohorte pour construire les parités moyennes d'une cohorte fictive. La méthode s'applique aux données provenant entièrement ou partiellement d'enquêtes nationales sur des échantillons représentatifs aussi bien qu'à des données de recensement, car bien que les individus enquêtés d'une cohorte donnée ne soient pas les mêmes aux deux dates, leurs parités moyennes sont représentatives des parités de la population féminine échantillonnée.

Il n'est pas nécessaire que les deux séries de données se réfèrent à deux points séparés par cinq ou dix ans exactement. A moins que la fécondité soit en train d'évoluer très rapidement, des intervalles de quatre ou onze ans fourniront par exemple des estimations raisonnables. Dans ces cas, on ne suit plus une cohorte d'une enquête à l'autre, mais cela n'est pas très important puisque la parité moyenne d'un groupe d'âge ne change pas rapidement d'une année à l'autre.

Bien que la force de la méthode réside dans sa robustesse en cas de modifications de la fécondité, la technique que nous présentons ici peut aussi permettre d'estimer les taux de fécondité par âge lorsqu'on utilise les données d'un seul recensement ou d'une seule enquête, à condition que la fécondité n'ait pas changé au cours de la vie reproductive des femmes concernées.

HYPOTHÈSES

La plupart des hypothèses sont celles associées au modèle relationnel de Gompertz, à savoir

- Le standard de fécondité choisi pour la procédure d'ajustement reflète bien la forme de la distribution de la fécondité dans la population.
- Les variations de la fécondité ont été progressives et régulière et elles ont affecté tous les groupes d'âge à peu près de la même manière.
- Les jeunes femmes entre 20 et 30 ans ont déclaré correctement leurs parités.

On suppose en outre que la mortalité et la migration n'ont eu aucun effet sur les distributions effectives par parité ; c'est-à-dire qu'on suppose que la parité moyenne des femmes qui sont décédées ou ont migré entre les deux enquêtes ne diffère pas significativement de la parité moyenne aux mêmes âges des femmes survivantes et présentes en fin de période.

TRAVAIL PRÉPARATOIRE ET RECHERCHES PRÉLIMINAIRES

Avant d'entreprendre l'analyse des niveaux de la fécondité par cette méthode, les chercheurs doivent s'interroger sur la qualité des données au moins dans les domaines suivants :

- la structure par âge et sexe de la population ; et
- les parités moyennes, après application de la correction d'El-Badry si nécessaire.

PRÉCAUTIONS ET MISES EN GARDE

On gardera présentes à l'esprit les mises en garde concernant l'utilisation des informations sur les enfants déjà nés dans les estimations de fécondité. Même dans les pays où les données sont par ailleurs bonnes, les femmes âgées ont tendance à omettre de déclarer certains de leurs enfants, sans doute ceux qui sont décédés ou ceux qui ont quitté le foyer parental. En conséquence, il arrive fréquemment que les parités moyennes ne s'accroissent pas à un rythme vraisemblable ou même qu'elles décroissent au-delà de 35 ou 40 ans. Les taux de fécondité calculés à partir de parités affectées par

de telles omissions tendent à sous-estimer la fécondité aux âges élevés. Les effets peuvent ne pas être très visibles si l'erreur est relativement mineure. Les estimations de fécondité fondées sur les parités moyennes des femmes âgées doivent donc être interprétées avec prudence, en particulier si elles font apparaître une fécondité relativement faible par rapport à celle estimée à partir des déclarations des jeunes femmes. Les parités moyennes d'une cohorte fictive sont, en outre, très sensibles aux variations dans la déclaration des parités d'un recensement à l'autre, et leur calcul constitue en conséquence un intéressant test de cohérence des données initiales.

Quand il existe des données supplémentaires sur la fécondité récente, on doit préférer à la méthode décrite ici la procédure qui compare les taux de fécondité intercensitaires cumulés aux parités moyennes d'une cohorte fictive en utilisant un modèle relationnel de Gompertz synthétique ; cette seconde méthode est en effet moins sensible à l'omission d'enfants déjà nés dans les déclarations des femmes âgées.

APPLICATION DE LA MÉTHODE

Les deux premières étapes sont les mêmes que celles de la méthode relationnelle synthétique de Gompertz.

Étape 1 : Calcul des parités moyennes déclarées à partir de chaque recensement

Calculer les parités moyennes ${}_5P_x(t_1)$ et ${}_5P_x(t_2)$ des femmes de chaque groupe d'âge $[x, x + 5)$ aux deux recensements (t_1 et t_2), pour $x = 15, 20 \dots 45$ si cela n'a pas déjà été fait lors des recherches préliminaires ou à l'occasion de l'application de la correction d'El-Badry. Pour faciliter l'exposé, nous désignons par $P(i, t) = {}_5P_x(t)$ la parité moyenne dans le groupe d'âge i au temps t où $i = (x/5 - 2)$. Les parités moyennes tirées du premier recensement (ou enquête) sont notées $P(i, 1)$ et celles tirées du second $P(i, 2)$.

Étape 2 : Calcul des parités moyennes pour une cohorte fictive

Le mode de calcul des parités dépend de la longueur de l'intervalle entre les deux recensements.

a) Intervalle de cinq ans

Si l'intervalle entre les deux séries de données est de cinq ans, les survivantes du groupe d'âge i au premier recensement appartiennent au groupe d'âge $i + 1$ au deuxième, et l'accroissement de parité entre les deux recensements est égal

à $P(i+1,2) - P(i,1)$. Ces accroissements peuvent être calculés pour chaque groupe d'âge, et les parités de la cohorte fictive sont alors obtenues par cumul des accroissements successifs. Ainsi, si on désigne par $\Delta P(i+1)$ l'accroissement de parité pour la cohorte du groupe d'âge i aux premier recensement, et par $P(i,s)$ la parité du groupe d'âge i dans la cohorte fictive (où renvoie à « synthétique »), on a $\Delta P(i+1) = P(i+1,2) - P(i,1)$ pour $i = 1, \dots, 6$, et donc $P(i,s) = \sum_{j=1}^i \Delta P(j)$. On considère que l'accroissement de parité $\Delta P(i+1)$ pour le groupe d'âge le plus jeune ($i = 0$) est égal à $P(1,2)$, en supposant que $P(0,1)$, parité moyenne des femmes de 10-14 ans au premier recensement, est nulle. Si la fécondité évolue rapidement, cette valeur de $\Delta P(1)$ reflète donc des taux du moment plus proches de l'enquête que du milieu de l'intervalle, donnant un poids un peu exagéré à l'évolution de la fécondité.

b) Intervalle de dix ans

Si la période entre les recensements ou les enquêtes est de 10 ans, les survivantes de la cohorte appartenant au groupe d'âge i lors de la première enquête seront dans le groupe d'âge $(i+2)$ lors de la seconde. Les parités de la cohorte fictive sont alors obtenues en cumulant deux séries parallèles d'accroissements de parité. Une fois encore, pour les groupes d'âge les plus jeunes, on considère que $\Delta P(1)$ est égal à $P(1,2)$ et que $\Delta P(2)$ est égal à $P(2,2)$. Les autres accroissements de parité sont égaux à $\Delta P(i+2) = P(i+2,2) - P(i,1)$ pour $i = 1, \dots, 5$.

Les parités de la cohorte fictive pour les groupes d'âge ayant un numéro pair sont obtenues en cumulant les accroissements de parité pour les groupes d'âge ayant un numéro pair, alors que ceux pour les groupes d'âge ayant un numéro impair sont obtenus en cumulant les accroissements de parité pour les groupes d'âge ayant un numéro impair. Donc

$$P(1,s) = \Delta P(1) = P(1,2)$$

$$P(2,s) = \Delta P(2) = P(2,2)$$

$$P(3,s) = \Delta P(1) + \Delta P(3)$$

$$P(4,s) = \Delta P(2) + \Delta P(4)$$

$$P(5,s) = \Delta P(1) + \Delta P(3) + \Delta P(5)$$

$$P(6,s) = \Delta P(2) + \Delta P(4) + \Delta P(6)$$

$$P(7,s) = \Delta P(1) + \Delta P(3) + \Delta P(5) + \Delta P(7)$$

Les étapes suivantes reprennent celles déjà décrites lors du recours au modèle relationnel de Gompertz, mais en ajustant une droite sur les données de parité seulement.

Étape 3 : Ajustement d'un modèle relationnel de Gompertz

En cas de cohérence interne des données de parité, les points représentant $z(i) - e(i)$ en fonction de $g(i)$ sont alignés sur une droite. On doit exclure du modèle les P -points dont la position s'écarte de la droite. Une régression linéaire ordinaire (par les moindres carrés) permet d'ajuster des droites sur les P -points et d'identifier, les uns après les autres, ceux qui ne s'ajustent pas nettement sur une droite. L'objectif est de rechercher la combinaison du plus grand nombre de P -points qui s'alignent (à peu près) sur la même droite, et de les utiliser pour ajuster le modèle.

Les points sont retenus dans l'ajustement selon les règles suivantes :

- Une série de points contigus doit être incluse dans le modèle. Seuls des points extrêmes peuvent être exclus. (La raison est que chaque point sur le graphique est le résultat d'un calcul de ratio entre les valeurs de données adjacentes. Si l'analyse conduit à considérer une donnée comme non fiable en tant que dénominateur, il n'est pas logique de l'accepter comme numérateur du ratio suivant.)
- Il vaut mieux éliminer les P -points aux âges avancés plutôt qu'aux jeunes âges car les données y sont généralement moins fiables et présentent moins de cohérence entre descendance et fécondité récente.
- Un ajustement réalisé sur un nombre élevée de points est préférable à celui réalisé sur un plus faible nombre moins de points, même s'il est légèrement moins bon.

Étape 4 : Evaluer la qualité des paramètres estimés

Les valeurs de α et β , qui représentent la droite de meilleur ajustement joignant les P -points et les F -points restants, nécessitent un contrôle pour vérifier qu'elles ne s'éloignent pas de leurs valeurs centrales de telle manière que le choix du standard serait mis en cause. Un bon ajustement est obtenu si $-0,3 < \alpha < 0,3$, et si $0,8 < \beta < 1,25$.

Si les paramètres se situent en dehors de ces intervalles, une au moins des séries de données sous-jacentes pose problème ou le standard est inadéquat. Il convient alors d'essayer un autre standard ou de modifier la sélection des points avant de procéder plus avant. Si les paramètres restent en dehors aux intervalles ci-dessus, il faut considérer que la méthode est inappropriée.

Étape 5 : Taux de fécondité par âge et indice synthétique de fécondité ajustés

Une fois estimés les deux paramètres du modèle, ceux-ci peuvent être appliqués au standard des parités pour fournir des valeurs ajustées $Y(i) = \alpha + \beta.Y^s(i)$. Celles-ci sont ensuite ramenées à des mesures de la proportion cumulée de fécondité réalisée par groupe d'âge i , grâce à la transformation anti-gompit. Les anti-gompits fondés sur les distributions de parité indiquent la proportion de fécondité réalisée dans le groupe d'âge. En divisant la parité observée dans chaque groupe d'âge par ces proportions, on obtient une série d'estimations de l'indice synthétique de fécondité. La moyenne de ces valeurs calculée sur le sous-ensemble des groupes d'âge qui ont été utilisés pour estimer α et β donne l'estimation ajustée de l'indice synthétique de fécondité, \hat{T} . En appliquant les mêmes α et β aux gompits standards pour les âges qui bornent les groupes d'âge conventionnels (c'est-à-dire 20, 25 ...50 ans), en appliquant la transformation anti-gompit, et en multipliant par \hat{T} , on obtient une distribution de la fécondité cumulée aux divers âges. En faisant la différence entre les estimations successives de la fécondité cumulée et en la divisant par cinq, on obtient une distribution de fécondité ajustée pour les groupes d'âge conventionnels (15-19, 20-24 etc.).

On considère que ces taux de fécondité par âge s'appliquent au point médian de la période entre les deux enquêtes.

EXEMPLE

Nous utilisons les données sur les parités moyennes tirées des recensements du Kenya en 1989 et 1999, comme dans l'exemple sur le modèle relationnel synthétique de Gompertz. Mais dans cette application on suppose qu'on ne dispose d'informations que sur les parités moyennes, pas sur la fécondité récente. Le processus d'ajustement du modèle relationnel de Gompertz aux données de parité est, pour l'essentiel, le même que pour le modèle relationnel de Gompertz de base. L'exposé insiste donc sur les différences entre les deux procédures. La méthode est mise en œuvre dans un dossier Excel associé disponibles à <http://demographicestimation.iussp.org/fr>.

Étape 1 : Calcul des parités moyennes déclarées à partir de chaque recensement

Une correction d'El-Badry a été appliquée aux données du recensement de 1989. De toute évidence, les résultats du recensement de 1999 ont été corrigés : il n'y a pas de don-

nées manquantes sur la parité. Les parités moyennes tirées des deux recensements figurent aux deux premières colonnes du tableau 1. Les descendances des femmes âgées ont reculé d'environ 0,6 enfant au cours de la décennie. Mais l'accroissement de la descendance des adolescentes est un peu surprenant.

Étape 2 : Calcul des parités moyennes pour une cohorte fictive

La période intercensitaire est de 10 ans (entre 1989 et 1999). Nous suivons donc la procédure décrite à l'étape 2(b) pour calculer les parités moyennes des cohortes, qui figurent dans la dernière colonne du tableau 11.1.

TABLEAU 11.1 Parités moyennes par groupe d'âge, Kenya, recensements de 1989 et 1999

| Groupe d'âge | 1989 | 1999 | Parité de la cohorte fictive $P(i,s)$ |
|--------------|--------|--------|---------------------------------------|
| 15-19 | 0,2416 | 0,2848 | 0,2848 |
| 20-24 | 1,5247 | 1,3640 | 1,3640 |
| 25-29 | 3,2138 | 2,6073 | 2,6505 |
| 30-34 | 4,7602 | 4,1432 | 3,9825 |
| 35-39 | 6,2390 | 5,3867 | 4,8234 |
| 40-44 | 7,1204 | 6,3818 | 5,6041 |
| 45-49 | 7,5103 | 6,9143 | 5,4987 |

Comme indiqué à l'étape 2(b), $\Delta P(1) = P(1,2) = 0,2848$ et $\Delta P(2) = P(2,2) = 1,3640$, alors que

$$\begin{aligned} P(5,s) &= \Delta P(1) + \Delta P(3) + \Delta P(5) \\ &= 0,2848 + (2,6073 - 0,2416) + (5,3867 - 3,2138) \\ &= 4,8234 \end{aligned}$$

On voit aisément que d'importantes omissions ont dû affecter les parités aux âges élevés, puisque la parité dans la cohorte fictive à l'âge le plus élevé est inférieure à la parité des femmes dans la cohorte fictive intercensitaire âgée de 40-44 ans.

La définition de l'âge de la mère n'a aucune incidence sur la méthode. Les parités moyennes sont – par définition – celles qui prévalent à la date de l'enquête ou du recensement.

Étape 3 : Ajustement d'un modèle relationnel de Gompertz

On utilise les données de la cohorte fictive à la dernière colonne du tableau 11.1 pour estimer la fécondité au moyen du modèle relationnel de Gompertz. Les points de données (P -points) fondés sur les parités moyennes sont éliminés l'un après l'autre jusqu'à ce que les points restants soient alignés

avec les parités (transformées) du schéma de fécondité standard. Les points ajustés sont représentés sur la figure 11.1.

Seuls cinq points de parité peuvent être représentés puisque la parité supposée pour le groupe 45-49 ans est moindre que pour le groupe 40-44 ans (5,4987 contre 5,6041), ce qui signifie que le gompit du rapport entre ces deux points n'est pas défini. L'examen de ces points fait apparaître une sous-déclaration de la fécondité aux âges qui servent à déterminer le dernier point représenté. L'élimination de ce point entraîne une nette réduction de l'écart quadratique moyen ; le modèle est donc ajusté sur les quatre points restants.

Étape 4 : Évaluer la qualité des paramètres estimés

Il en résulte des valeurs de α et β égales respectivement à $-0,0084$ and $1,0071$. Le schéma de fécondité est donc proche de celui qui sous-tend le standard modifié par Zaba.

Étape 5 : Taux de fécondité par âge et indice synthétique de fécondité ajustés

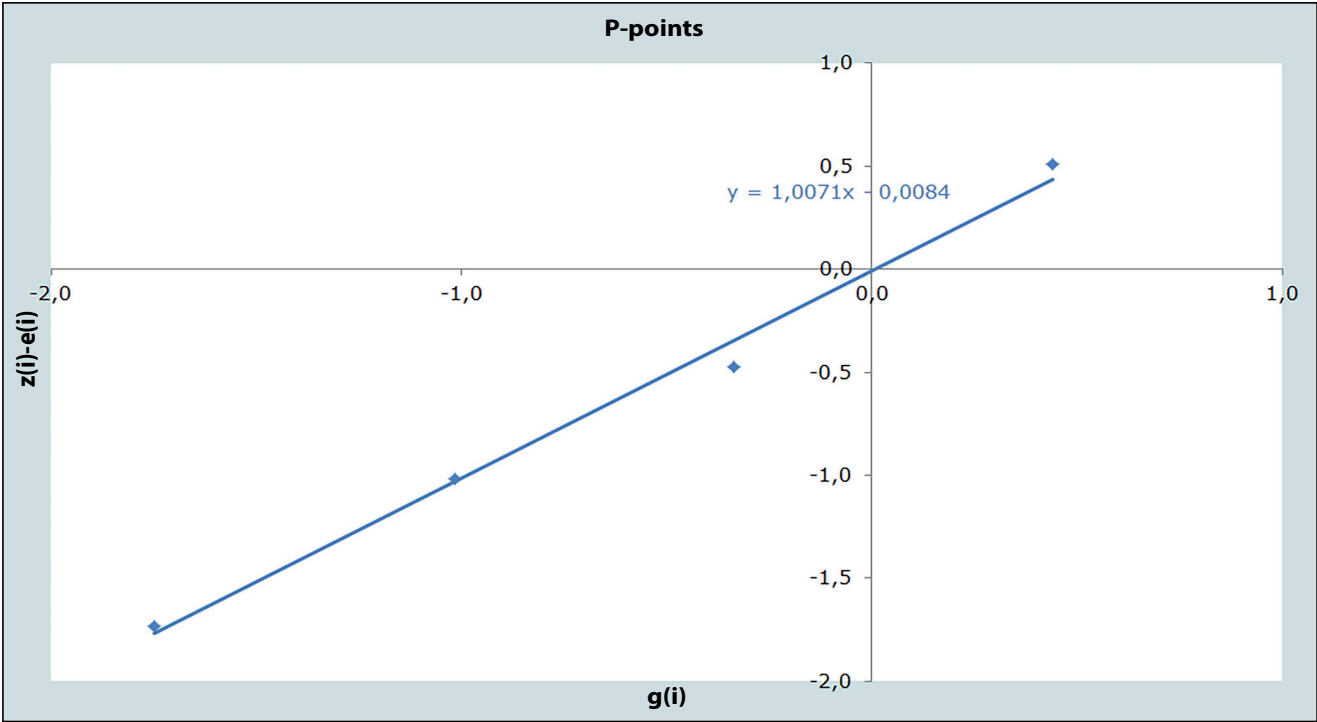
On applique ces paramètres aux gompits des parités du standard, en utilisant le modèle relationnel linéaire

$Y(i) = \alpha + \beta Y^s(i)$. En prenant les anti-gompits (colonne 4 du tableau 11.2) et en divisant par ceux-ci les parités observées aux âges choisis pour l'ajustement du modèle, on obtient une série de cinq estimations de l'indice synthétique de fécondité (comprises entre 5,4 et 5,7 enfants par femme). Leur moyenne (\hat{T}) est de 5,54 enfants par femme.

TABLEAU 11.2 Calcul de l'indice synthétique de fécondité estimé (T-chapeau), Kenya, recensements de 1989 et 1999

| Âge (i) | $Y^s(i)$ | $Y(i)$ ajusté | $\exp(-\exp(-Y(i)))$ | Cumulant effectif |
|---------|----------|---------------|----------------------|-------------------|
| 0 | - 2,0961 | - 2,1194 | 0,0002 | 0,0013 |
| 1 | - 1,0833 | - 1,0994 | 0,0497 | 0,2754 |
| 2 | - 0,3124 | - 0,3230 | 0,2513 | 1,3930 |
| 3 | 0,3541 | 0,3482 | 0,4936 | 2,7368 |
| 4 | 1,0579 | 1,0570 | 0,7065 | 3,9166 |
| 5 | 1,9561 | 1,9615 | 0,8688 | 4,8167 |
| 6 | 3,4225 | 3,4384 | 0,9684 | 5,3688 |
| 7 | 6,0922 | 6,1270 | 0,9978 | 5,5320 |

FIGURE 11.1 P-points ajustés, cohortes fictives, données des recensements du Kenya 1989 et 1999



On applique les estimations ajustées de α et β aux gompits du standard, $Y^s(x)$, dans chaque groupe d'âge pour obtenir les gompits ajustés. On en déduit les anti-gompits qui, multipliés par \hat{T} , donnent la distribution cumulée de fécondité modifiée, $FM(x)$. La différence entre deux cumuls consécutifs, divisée par cinq donne la série finale des taux de fécondité par âge à la dernière colonne du tableau 11.3.

TABLEAU 11.3 Calcul de la distribution finale de fécondité ajustée, Kenya, recensements de 1989 et 1999

| Âge (x) | $Y^s(x)$ | $Y(x)$ ajusté | $\exp(-\exp$ $(-Y(i)))$ | $FM(x)$ | $fm(x)$ |
|---------------------------------|----------|------------------|----------------------------|---------|---------|
| 15 | -1,7731 | -1,7262 | 0,0036 | 0,0212 | 0,0042 |
| 20 | -0,6913 | -0,7318 | 0,1251 | 0,7318 | 0,1421 |
| 25 | 0,0256 | -0,0727 | 0,3411 | 1,9957 | 0,2528 |
| 30 | 0,7000 | 0,5472 | 0,5607 | 3,2801 | 0,2569 |
| 35 | 1,4787 | 1,2630 | 0,7537 | 4,4090 | 0,2258 |
| 40 | 2,6260 | 2,3176 | 0,9062 | 5,3013 | 0,1785 |
| 45 | 4,8097 | 4,3249 | 0,9869 | 5,7732 | 0,0944 |
| 50 | 13,8155 | 12,6034 | 1,0000 | 5,8501 | 0,0154 |
| Indice synthétique de fécondité | | | | | 5,53 |

L'indice synthétique de fécondité estimé est de 5,53 enfants par femme ; il s'applique à mi-chemin entre les deux recensements. Dans cet exemple, les taux de fécondité par âge tirés des parités de la cohorte fictive peuvent être comparés à ceux obtenus par l'application du modèle relationnel synthétique de Gompertz (indice synthétique = 5,56 enfants par femme). La proximité des deux séries de résultats est rassurante.

Il faut toutefois se souvenir que les résultats peuvent être sérieusement biaisés si les mères omettent de déclarer des enfants déjà nés, et en particulier si l'ampleur de ces omissions varie d'une enquête à l'autre.

DESCRIPTION DÉTAILLÉE DE LA MÉTHODE

La méthode décrite ici est une variante du modèle relationnel de Gompertz, mais, au lieu d'utiliser des données de parité et de fécondité recueillies en un point du temps, elle élabore un schéma de fécondité 'moyen' fondé sur les descendance déclarées en deux points du temps et utilise celles-ci – seules – pour déterminer une distribution de fécondité. L'exposé mathématique complet du modèle relationnel a déjà été présenté.

VARIANTES DE LA MÉTHODE

Une option de la feuille de calcul permet de choisir une période intercensitaire égale à zéro. Ceci permet de calculer un indice synthétique de fécondité à partir d'une seule série de données de parité. Pour que les estimations produites soient plausibles, il ne suffit pas que les parités moyennes aient été déclarées correctement, il faut aussi que la fécondité se soit maintenue constante sur une longue période avant le recensement ou l'enquête.

AUTRES LECTURES ET RÉFÉRENCES

L'essentiel de l'exposé de la méthode est donné par Arretx (1973) et mis à jour dans le *Manuel X* (Division de la Population des NU 1984). La version donnée dans le *Manuel X* s'appuie sur la méthode du quotient P/F pour convertir les accroissements de parité en taux de fécondité ; la méthode présentée ici recourt à la méthode relationnelle de Gompertz, qui est plus souple.

Arretx C. 1973. "Fertility estimates derived from information on children ever born using data from censuses," in *International Population Conference, Liège 1973*. Vol. 2. Liège: International Union for the Scientific Study of Population, pp. 247-261.

Division de la Population des NU. 1984. *Manuel X. Techniques indirectes d'estimation démographique*. New York : Nations Unies, Département des affaires économiques et sociales internationales, ST/ESA/SER.A/81. http://unstats.un.org/unsd/demographic/standmeth/handbooks/Manuel_X-fr.pdf

Méthodes fondées
sur des données d'enquêtes

CHAPITRE 12 ESTIMATION DIRECTE DE LA FÉCONDITÉ À PARTIR DE DONNÉES D'ENQUÊTES CONTENANT DES HISTOIRES GÉNÉSIQUES

Tom A Moultrie

DESCRIPTION DE LA MÉTHODE

L'estimation directe de la fécondité (taux par âge et indice synthétique) à partir de données d'enquêtes contenant des histoires génésiques est relativement simple. Si les données sont recueillies soigneusement à l'aide d'un instrument validé (comme celui utilisé dans les Enquêtes Démographiques et de Santé), elles peuvent fournir des estimations de fécondité fiables et précises. Elles sont cependant souvent soumises à des déformations, en particulier dues au report des naissances vers des périodes plus anciennes afin d'éviter de répondre à des questions supplémentaires, par exemple sur la santé des enfants ou l'anthropométrie (Cleland 1996). Ces problèmes ont été à nouveau mis en évidence récemment par Schoumaker (2010, 2011). La mauvaise datation des naissances et leur omission peuvent entraîner une sous-estimation de la fécondité (en particulier dans la période de trois à cinq ans avant l'enquête).

Il existe deux façons d'estimer la fécondité directement à partir des données des histoires génésiques détaillées. La première – utilisée par les EDS dans leurs rapports officiels – consiste à produire une estimation couvrant une période d'un ou de trois ans avant l'enquête. (La période de trois ans est souvent préférée pour éviter des fluctuations dans les estimations dues au nombre relativement faible de naissances annuelles dans les EDS). La méthode est décrite en détail dans le *Guide des statistiques des EDS* (en anglais) (Rutstein and Rojas 2003). Elle présente deux inconvénients. D'une part, si la phase de collecte des données est étalée sur une longue période, il devient impossible de dater précisément la mesure de la fécondité. D'autre part, le calcul des taux de fécondité est complexe du fait de la nécessité de se référer à la date de l'enquête, tout en travaillant par groupes d'âges quinquennaux et par période triennale d'observation.

Nous décrivons ici une démarche plus simple qui consiste à produire des estimations de la fécondité par année d'âge et par année de calendrier. Ces estimations peuvent très facilement être agrégées pour fournir des estimations pour des groupes d'âge plus larges et pour des périodes pluriannuelles.

Comme pour la méthode EDS, les traitements initiaux doivent être réalisés sur les données individuelles. Il est donc généralement raisonnable d'estimer la fécondité directement à partir des histoires génésiques en utilisant la fonction de durée de vie intégrée dans un programme d'analyse statistique comme *Stata*. Une routine utile pour effectuer ces calculs dans *Stata* a été publiée par Schoumaker (2013). Les calculs sont toutefois suffisamment faciles pour qu'on puisse les mener à partir de simples tableaux croisés des données. Nous expliquons ici comment procéder.

DONNÉES NÉCESSAIRES ET HYPOTHÈSES

Données nécessaires

Deux séries de données sont nécessaires, toutes deux produites couramment lors du traitement d'une enquête ayant recueilli des histoires génésiques détaillées. La première est une série dont l'unité d'analyse est la femme – c'est-à-dire qu'il y a un enregistrement par femme. Ces données sont nécessaires pour estimer le dénominateur des taux de fécondité. La seconde série de données a l'enfant comme unité d'analyse – c'est-à-dire qu'il y a un enregistrement par enfant – mais des informations de base sur la mère (essentiellement sa date de naissance) sont aussi incluses dans chaque enregistrement.

Pour estimer la fécondité, les informations suivantes doivent figurer dans les données.

a) Série des femmes

1. Le mois et l'année de naissance de chaque femme, éventuellement en appliquant la procédure CMC (*century-month code*) de numérotation des mois à partir du début du siècle des EDS.
2. Le mois et l'année de l'interview.
3. Toutes les variables nécessaires pour ajuster les données en tenant compte du plan d'échantillonnage et des pondérations.
4. Les variables importantes en fonction desquelles on souhaite établir des différentiels de fécondité, en sachant que les variables à la date de l'enquête ne s'appliquaient peut-être pas au moment où ont eu lieu les événements d'intérêt (les naissances récentes).

b) Série des enfants

1. La date de naissance de l'enfant – mois et année – éventuellement convertie en code CMC.
2. La date de naissance de la mère – mois et année – éventuellement convertie en code CMC.
3. Toutes les variables nécessaires pour ajuster les données en tenant compte du plan d'échantillonnage et des pondérations.
4. Les mêmes covariables en fonction desquelles établir des différentiels de fécondité.

PRÉCAUTIONS ET MISES EN GARDE

- Les taux de fécondité par année d'âge calculés à partir d'enquêtes de taille relativement modeste fournissent des indications sur la qualité des données, mais ils sont presque toujours trop erratiques pour être utilisés directement. Leur agrégation en groupes quinquennaux (ainsi, éventuellement, qu'un lissage des taux au moyen d'un modèle relationnel de Gompertz) est presque toujours nécessaire.
- De la même façon, les taux par année de calendrier tirés des données d'enquête peuvent ne pas être fiables. Les données de plusieurs années doivent être combinées pour produire une estimation plus fiable. Dans l'idéal, on doit cependant éviter de regrouper plus de trois années pour éviter d'aplatir les tendances de la fécondité.
- Les taux estimés peuvent être affectés par l'omission et la mauvaise datation des naissances déclarées.
- Les taux estimés ne seront pas les mêmes que ceux produits par la méthode EDS. D'une part, l'estimation de la période d'exposition au risque est légèrement différente (la méthode EDS utilise des mois complets, nous des

demi-mois). D'autre part, il y a aussi une différence sur la période de référence des taux qui peut aller jusqu'à 11 mois. On peut cependant calculer les taux pour des années allant de juillet à juin (et donc centrées sur le 1^{er} janvier, ou en fait pour toute période de 12 mois) en manipulant convenablement le numérateur et le dénominateur.

APPLICATION DE LA MÉTHODE

Nous définissons les termes suivants :

- M_B^c le mois de naissance de l'enfant
- Y_B^c l'année de naissance de l'enfant
- M_B^m le mois de naissance de la mère
- Y_B^m l'année de naissance de la mère
- M_I le mois où la mère est interviewée
- Y_I l'année où la mère est interviewée
- $B(x, t)$ le nombre total de naissances des mères âgées de x à la naissance de leur enfant l'année t
- $E(x, t)$ les personnes-années d'exposition au risque des femmes âgées de x l'année t .

Les taux sont calculés selon les étapes suivantes. Pour éviter de devoir faire des hypothèses sur l'exposition au risque au cours du mois de l'interview, nous laissons de côté l'exposition au risque et les naissances survenant au cours de ce mois.

Nous présentons le cas général où toutes les femmes ne sont pas interviewées la même année civile. Le processus peut être simplifié lorsque toutes les femmes sont interviewées la même année.

Étape 1 : Établir un tableau des naissances de chaque année par âge de la mère à la naissance de l'enfant

Cette étape produit le numérateur des taux de fécondité : les naissances des enfants par année civile et par âge de la mère à la naissance.

En principe, l'élaboration de ce tableau est relativement simple, mais des précautions doivent être prises pour déterminer correctement l'âge de la mère à la naissance de son enfant, lorsque mère et enfant ont le même mois de naissance. Si, comme c'est généralement le cas, le jour de naissance n'est pas connu, on doit décider aléatoirement que le jour de naissance de la mère tombe avant ou après le jour de naissance de l'enfant. Ceci peut être fait en créant une variable binaire, b , et en utilisant un générateur de nombres aléatoires, mais cette procédure a des conséquences sur la cohérence et la répliquabilité de la recherche. À la place, b peut être généré à partir d'une variable supposément uni-

forme qui n'a pas de rapport avec notre objet, par exemple le jour du mois où la mère a été interviewée. Nous définissons donc $b = 1$ si le jour de l'interview est supérieur à 15 et $b = 0$ si le jour du mois est le 15 ou moins.

L'âge (révolu) de la mère à la naissance d'un enfant donné, x , est donné par

$$x = \text{int} \left(\frac{12(Y_B^c - Y_B^m) + (M_B^c - M_B^m - b)}{12} \right)$$

où $\text{int}()$ représente la partie entière du terme entre parenthèses.

Extraire un tableau faisant apparaître le nombre total de naissances, $B(x, t)$, dans chaque cellule définie par les combinaisons de Y_B^c et x , en pondérant les données si nécessaire et en s'assurant d'exclure les naissances survenues dans le mois où la mère a été interviewée.

Étape 2 : Calculer l'âge de chaque femme au début de l'année où elle a été interviewée

Travaillant sur la série de données des femmes (c'est-à-dire un enregistrement par femme), on commence par calculer l'âge des femmes au 1^{er} janvier de l'année de l'interview, x_I , en supposant que les naissances des mères sont distribuées uniformément le long de chaque mois :

$$x_I = \text{int} \left(\left(Y_I^m - Y_B^m - 1 \right) + \frac{(12 - M_B^m + 0,5)}{12} \right) \quad (1)$$

Il s'ensuit que l'âge de la mère au 1^{er} janvier de toute autre année, t , ($t \leq Y_I$) est égal à $x_I - (Y_I - t)$.

Étape 3 : Calculer l'exposition au risque de chaque femme dans l'année de son interview

Au cours de l'année où elle est interviewée, une femme n'est exposée au risque d'une naissance qu'une fraction du temps (c'est-à-dire, la partie de l'année qui précède l'interview). Dans ce cas, le calcul de l'exposition au risque est fonction du fait qu'elle a été interviewée avant ou après l'anniversaire fêté cette année-là. Si son mois de naissance précède le mois de l'interview, elle sera exposée au risque d'avoir une naissance à l'âge x_I pendant $E(x_I, Y_I) = \frac{M_B^m - 0,5}{12}$ d'année, et pendant $E(x_I + 1, Y_I) = \frac{M_I - M_B^m - 0,5}{12}$ d'année à l'âge $x_I + 1$. Si son mois de naissance est le

même que, ou postérieur à, celui de l'interview, son exposition au risque d'une naissance l'année de l'interview sera de $E(x_I, Y_I) = \frac{M_I - 1}{12}$ d'année à l'âge x_I et de

$$E(x_I + 1, Y_I) = 0 \text{ année à l'âge } x_I + 1.$$

Notez qu'au cours de la dernière année civile complète, l'exposition agrégée par femme est de 1 an, alors qu'elle est de $(M_I - 1)/12$ d'année, l'année de l'interview, quelle que soit la position relative du mois de la naissance et du mois de l'interview.

Les variables donnant l'exposition de chaque femme aux âges x_I et $x_I + 1$ au cours de l'année de l'interview doivent être calculées puis agrégées (et pondérées si nécessaire), afin de donner un tableau d'exposition agrégée par âge l'année de l'interview.

Étape 4a : Calculer l'exposition au risque de chaque femme au cours de la dernière année complète avant l'interview.

Au cours de la dernière année civile complète précédant son interview, l'année $t = Y_I - 1$, une femme est âgée de $x_I - 1$ avant son anniversaire, et x_I par la suite. Avec la même hypothèse de répartition uniforme des naissances au long de l'année, la fraction d'année entre le 1^{er} janvier et l'anniversaire de la femme est égale à $E(x_I - 1, Y_I - 1) = \frac{M_B^m - 0,5}{12}$.

Le restant de l'année, la femme est âgée de x_I et l'exposition est égale à :

$$E(x_I, Y_I - 1) = 1 - E(x_I - 1, Y_I - 1) = 1 - \frac{M_B^m - 0,5}{12}.$$

Sur la base de ces deux formules, les variables donnant les expositions de chaque femme aux âges x_I et $x_I + 1$ l'année $Y_I - 1$ doivent être calculées puis agrégées (après pondération si nécessaire) afin de produire un tableau de l'exposition agrégée par âge cette année.

Étape 4b : Calculer l'exposition au risque pour les années complètes antérieures

Les histoires génésiques sont recueillies rétrospectivement auprès de l'ensemble des femmes, chacun d'entre elles fournissant des informations sur toute la période où elle a été exposée au risque de fécondité. Certaines peuvent avoir déménagé ou changé de caractéristiques sur d'autres points au cours de la période, mais comme des histoires résidentielles et économiques sont rarement recueillies dans les enquêtes de fécondité, il est généralement impossible d'en tenir compte dans le calcul des taux

de fécondité. L'interprétation de certains résultats comme la fécondité par lieu de résidence peut en conséquence manquer de clarté.

En revanche, comme les dates de naissance sont fixées une fois pour toutes et que la population féminine étudiée reste constante au fil du temps, l'exposition agrégée des femmes atteignant l'âge x une année où l'exposition de toutes les femmes est complète, v , sera aussi égale à l'exposition de la cohorte les années antérieures, c'est-à-dire

$$E(x, v-1) = E(x-1, v-2) = \dots = E(x-k, v-k-1) \quad (2)$$

Étape 5 : Calculer les taux de fécondité par âge

Pour calculer l'exposition totale à chaque âge chaque année, $E(x, t)$, on cumule les résultats obtenus aux étapes 3 et 4 pour chaque âge et pour chaque année (complète et incomplète). Si le travail de terrain s'est déroulé sur deux années civiles, $Y_I - 1$ se réfère à deux années différentes, tout comme Y_I . L'exposition totale pour la dernière année civile où elle peut être calculée ne couvre que l'exposition partielle des femmes interviewées la seconde année du travail de terrain, alors que l'exposition totale de l'année immédiatement antérieure couvre l'exposition partielle des femmes interviewées la première année du travail de terrain et la pleine exposition au cours de cette année des femmes interviewées la seconde année du travail de terrain.

Les taux de fécondité à l'âge x l'année t sont égaux à

$$f_x(t) = \frac{B_x(t)}{E_x(t)}$$

Les taux de fécondité par groupe quinquennal d'âge conventionnel sont obtenus en cumulant les naissances des femmes au sein de chaque groupe d'âge, et en divisant par le cumul des expositions dans ce groupe d'âge. Ainsi si $i = (x/5) - 2$ pour $x = 15, 20, \dots, 45$, on a $f(1) = {}_5f_{15}$; $f(2) = {}_5f_{20}$; ... $f(7) = {}_5f_{45}$ et

$$f(i, t) = \frac{\sum_{a=5i+10}^{5i+14} B_a(t)}{\sum_{a=5i+10}^{5i+14} E_a(t)}$$

Pour combiner les données de plusieurs années, on doit cumuler les numérateurs et les dénominateurs séparément avant d'en faire le rapport pour obtenir le taux :

$$f(i, (t_1, t_2)) = \frac{\sum_{z=t_1}^{t_2} \sum_{a=5i+10}^{5i+14} B_a(z)}{\sum_{z=t_1}^{t_2} \sum_{a=5i+10}^{5i+14} E_a(z)}$$

EXEMPLE

Nous utilisons les données de l'EDS du Malawi en 2004. Le travail de terrain a été réalisé entre octobre 2004 et février 2005.

Étape 1 : Établir un tableau des naissances de chaque année par âge de la mère à la naissance de l'enfant

Après allocation aléatoire de l'âge de la mère à la naissance de son enfant dans les cas où l'une et l'autre ont le même mois de naissance, le tableau croisé de l'année de naissance des enfants par âge de la mère à la naissance de son enfant est tel qu'il apparaît au tableau 12.1. On constate qu'il y a eu un important décalage ou une forte omission des naissances en 2001 et 2002, puisque le nombre de naissances déclaré pour ces années est inférieur d'environ 20 % à celui déclaré en 2003. Les naissances déclarées en 2004 sont moins nombreuses qu'en 2003 en partie parce que des femmes n'ont pas été exposées pendant toute l'année civile, et parce que les naissances survenues le mois de l'interview ont été exclues de l'analyse.

Étape 2 : Calculer l'âge de chaque femme au début de l'année où elle a été interviewée

L'âge des femmes au début de l'année où elles ont été interviewées est calculé d'après l'équation 1. Le tableau 12.2 donne une illustration du calcul. À la troisième ligne (cas 444 3), la femme est née en août 1984 et a été interviewée en octobre 2004. Au 1^{er} janvier 2004 elle était âgée de 19 ans (colonne 4). À la neuvième (avant-dernière) ligne (cas 528 2), la femme née en janvier 1970 et interviewée en janvier 2005 était âgée de 34 ans au 1^{er} janvier 2005.

Étape 3 : Calculer l'exposition au risque de chaque femme dans l'année de son interview

Les colonnes (5) et (6) du tableau 12.2 illustrent le calcul de l'exposition au risque de chaque femme l'année de l'interview. À la première ligne (cas 443 4) la femme a eu son 28^e anniversaire en février 2004. Sous l'hypothèse que les jours de naissance sont, en moyenne, au milieu de chaque

TABLEAU 12.1 Naissances classées par âge de la mère à la naissance depuis 2001, Malawi, EDS 2004

| Année de naissance | | | | | | Année de naissance | | | | | |
|--------------------|--------|--------|--------|--------|------|--------------------|---------|---------|---------|---------|------|
| Âge | 2001 | 2002 | 2003 | 2004 | 2005 | Âge | 2001 | 2002 | 2003 | 2004 | 2005 |
| 13 | 1,11 | 0,96 | 0,00 | 0,00 | 0,00 | 32 | 55,11 | 72,47 | 59,85 | 61,36 | 0,00 |
| 14 | 6,44 | 3,26 | 2,00 | 4,02 | 0,00 | 33 | 34,74 | 54,08 | 72,14 | 41,23 | 0,00 |
| 15 | 19,70 | 12,74 | 17,21 | 14,65 | 0,00 | 34 | 28,09 | 44,41 | 67,04 | 52,00 | 0,00 |
| 16 | 49,84 | 41,40 | 49,87 | 39,00 | 0,00 | 35 | 50,00 | 25,28 | 41,26 | 48,16 | 0,00 |
| 17 | 93,45 | 88,79 | 93,36 | 61,67 | 0,00 | 36 | 41,61 | 33,88 | 27,42 | 33,56 | 0,00 |
| 18 | 113,79 | 133,70 | 153,38 | 110,40 | 0,00 | 37 | 30,57 | 25,46 | 48,50 | 30,46 | 0,00 |
| 19 | 145,63 | 148,18 | 162,51 | 162,48 | 0,00 | 38 | 24,47 | 32,07 | 31,55 | 36,85 | 0,00 |
| 20 | 146,03 | 166,63 | 177,72 | 155,24 | 0,00 | 39 | 23,05 | 16,87 | 39,64 | 22,38 | 0,00 |
| 21 | 159,60 | 137,76 | 179,68 | 174,46 | 0,00 | 40 | 16,95 | 20,66 | 12,56 | 26,47 | 0,00 |
| 22 | 137,50 | 128,60 | 147,12 | 148,44 | 0,00 | 41 | 19,67 | 9,72 | 17,17 | 9,87 | 0,00 |
| 23 | 115,15 | 110,30 | 173,94 | 138,36 | 2,12 | 42 | 12,44 | 7,72 | 9,79 | 8,89 | 0,00 |
| 24 | 109,24 | 96,07 | 144,74 | 149,19 | 0,00 | 43 | 9,43 | 10,35 | 17,32 | 9,15 | 0,00 |
| 25 | 113,58 | 93,61 | 105,37 | 117,68 | 0,00 | 44 | 4,17 | 10,98 | 7,11 | 11,11 | 0,00 |
| 26 | 82,08 | 69,68 | 107,11 | 105,36 | 0,00 | 45 | 4,94 | 4,86 | 3,63 | 4,29 | 0,00 |
| 27 | 74,37 | 77,16 | 129,50 | 105,48 | 0,00 | 46 | 4,02 | 9,07 | 14,65 | 4,96 | 0,00 |
| 28 | 66,31 | 66,14 | 73,87 | 91,96 | 0,00 | 47 | 0,00 | 0,82 | 3,96 | 2,35 | 0,00 |
| 29 | 62,92 | 63,28 | 75,42 | 80,13 | 0,00 | 48 | 0,00 | 0,00 | 2,16 | 0,00 | 0,00 |
| 30 | 55,93 | 55,44 | 76,98 | 68,16 | 0,00 | 49 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| 31 | 55,89 | 42,38 | 59,05 | 56,76 | 0,00 | TOTAL | 1967,84 | 1914,75 | 2404,58 | 2186,55 | 2,12 |

TABLEAU 12.2 Données illustrant le calcul de l'exposition au risque, Malawi, EDS 2004

| Cas n° | Date de naissance | Date d'interview | Âge au début de l'année d'interview | Exposition l'année de l'interview | | Exposition dans la dernière année complète | |
|--------|-------------------|------------------|-------------------------------------|-----------------------------------|---------------|--|---------------|
| | | | | Âge inférieur | Âge supérieur | Âge inférieur | Âge supérieur |
| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| 443 4 | Février 1976 | Octobre 2004 | 27 | 0,125 | 0,625 | 0,125 | 0,875 |
| 443 10 | Octobre 1974 | Octobre 2004 | 29 | 0,750 | 0,000 | 0,792 | 0,208 |
| 444 3 | Août 1984 | Octobre 2004 | 19 | 0,625 | 0,125 | 0,625 | 0,375 |
| 445 2 | Juin 1983 | Octobre 2004 | 20 | 0,458 | 0,292 | 0,458 | 0,542 |
| 519 7 | Mai 1989 | Janvier 2005 | 15 | 0,000 | 0,000 | 0,375 | 0,625 |
| 522 2 | Mars 1979 | Janvier 2005 | 25 | 0,000 | 0,000 | 0,208 | 0,792 |
| 526 4 | Décembre 1989 | Janvier 2005 | 15 | 0,000 | 0,000 | 0,958 | 0,042 |
| 526 7 | Septembre 1979 | Janvier 2005 | 25 | 0,000 | 0,000 | 0,708 | 0,292 |
| 528 2 | Janvier 1970 | Janvier 2005 | 34 | 0,000 | 0,000 | 0,042 | 0,958 |
| 529 2 | Octobre 1972 | Janvier 2005 | 32 | 0,000 | 0,000 | 0,792 | 0,208 |

mois, elle a passé 0,125 an (1,5/12) âgée de 27 ans en 2004, et 0,625 an (7,5 mois de mi-février à la fin septembre, le mois précédant son interview) âgée de 28 ans la même année.

La femme de la deuxième ligne (cas 443 10) a son anniversaire le mois où elle a été enquêtée. Elle a donc eu 9 mois pleins d'exposition (0,75 an) à 29 ans en 2004, et pas d'exposition par la suite.

Les femmes interviewées en janvier 2005 n'ont pas eu d'exposition l'année de leur interview, puisque nous ne prenons pas en compte l'exposition (et les naissances) au cours de ce mois.

Étape 4a : Calculer l'exposition au risque de chaque femme au cours

de la dernière année complète avant l'interview

Les colonnes (7) et (8) du tableau 12.2 illustrent le calcul de l'exposition au risque lors de la dernière année complète où les femmes ont été exposées au risque d'une naissance dans les données de l'enquête. Pour les femmes interviewées en 2004, c'est en 2003. Pour les femmes interviewées en 2005, c'est en 2004.

Dans le deuxième cas (443 10), l'exposition en 2003 – dernière année complète d'exposition – est de 9,5 mois à 28 ans et 2,5 mois à 29 ans. Comme exprimé par l'équation 2, l'année précédente, l'exposition aurait été distribuée dans les mêmes proportions entre les âges diminués d'un an : en 2002, l'exposition est de 9,5 mois à 27 ans et 2,5 mois à 28 ans.

Dans le dernier cas présenté (529 2), la femme a passé à peu près 9,5 mois (0,792 an) à 31 ans en 2004 et 2,5 mois (0,208 an) à 32 ans en 2004.

En additionnant les expositions par année d'âge et année civile calculées à l'étape 4 on obtient l'exposition au risque du tableau 12.3.

Étape 5 : Calculer les taux de fécondité par âge

Les taux de fécondité par année d'âge pour chaque année civile sont obtenus en divisant les naissances du tableau 12.1 par les populations exposées au risque du tableau 12.3. Les résultats figurent au tableau 12.4.

Les résultats varient fortement d'une année à l'autre, les estimations de l'indice synthétique de fécondité différant de plus d'un enfant par femme entre 2002 et 2003. L'estimation de l'indice synthétique en 2004, bien que calculée sur une exposition partielle cette année-là pour la plupart des femmes, est très cohérente avec l'estimation de 2003. La forme de la distribution par âge (cf. figure 12.1) est très cohérente entre les trois années, bien qu'elle soit établie

FIGURE 12.1 Taux de fécondité par année d'âge et année civile, Malawi EDS 2004



TABLEAU 12.3 Exposition agrégée par année d'âge et année civile, Malawi, EDS 2004

| Âge | 2002 | 2003 | 2004 | 2005 |
|--------------|-----------------|-----------------|-----------------|---------------|
| 11 | 0,063 | 0,000 | 0,000 | 0,000 |
| 12 | 198,291 | 0,063 | 0,000 | 0,000 |
| 13 | 468,833 | 198,291 | 0,063 | 0,000 |
| 14 | 432,083 | 468,833 | 197,506 | 0,000 |
| 15 | 490,890 | 432,083 | 409,831 | 0,049 |
| 16 | 522,245 | 490,890 | 370,078 | 0,402 |
| 17 | 597,259 | 522,245 | 431,191 | 0,216 |
| 18 | 606,502 | 597,259 | 444,050 | 0,337 |
| 19 | 594,975 | 606,502 | 528,989 | 0,622 |
| 20 | 573,166 | 594,975 | 514,654 | 0,674 |
| 21 | 480,330 | 573,166 | 521,777 | 0,354 |
| 22 | 574,521 | 480,330 | 489,303 | 1,172 |
| 23 | 486,871 | 574,521 | 422,082 | 0,166 |
| 24 | 405,933 | 486,871 | 503,468 | 0,939 |
| 25 | 405,592 | 405,933 | 416,489 | 0,729 |
| 26 | 407,569 | 405,592 | 350,520 | 0,000 |
| 27 | 346,264 | 407,569 | 354,229 | 0,425 |
| 28 | 313,426 | 346,264 | 349,949 | 0,265 |
| 29 | 286,749 | 313,426 | 300,703 | 0,337 |
| 30 | 308,209 | 286,749 | 262,300 | 0,177 |
| 31 | 252,422 | 308,209 | 252,010 | 0,000 |
| 32 | 309,337 | 252,422 | 256,686 | 0,166 |
| 33 | 267,239 | 309,337 | 217,728 | 0,000 |
| 34 | 183,176 | 267,239 | 271,954 | 0,000 |
| 35 | 185,172 | 183,176 | 226,209 | 0,868 |
| 36 | 222,879 | 185,172 | 151,012 | 0,000 |
| 37 | 217,592 | 222,879 | 166,838 | 0,000 |
| 38 | 236,389 | 217,592 | 192,603 | 0,110 |
| 39 | 177,195 | 236,389 | 194,856 | 0,363 |
| 40 | 161,461 | 177,195 | 195,769 | 0,591 |
| 41 | 142,134 | 161,461 | 155,461 | 0,000 |
| 42 | 173,338 | 142,134 | 133,356 | 0,166 |
| 43 | 168,616 | 173,338 | 126,403 | 0,000 |
| 44 | 148,788 | 168,616 | 147,170 | 0,088 |
| 45 | 140,768 | 148,788 | 143,087 | 0,088 |
| 46 | 138,297 | 140,768 | 125,995 | 0,000 |
| 47 | 72,711 | 138,297 | 124,497 | 0,000 |
| 48 | 0,606 | 72,711 | 117,910 | 1,027 |
| 49 | 0,000 | 0,606 | 53,140 | 0,000 |
| TOTAL | 1 169,89 | 11697,89 | 1 011,87 | 10,330 |

TABLEAU 12.4 Taux de fécondité par année d'âge et année civile, Malawi, EDS 2004

| Âge | 2001 | 2002 | 2003 | 2004 |
|--|-------------|-------------|-------------|-------------|
| 11 | 0,000 | 0,000 | 0,000 | 0,000 |
| 12 | 0,000 | 0,000 | 0,000 | 0,000 |
| 13 | 0,003 | 0,002 | 0,000 | 0,000 |
| 14 | 0,013 | 0,008 | 0,004 | 0,020 |
| 15 | 0,038 | 0,026 | 0,040 | 0,036 |
| 16 | 0,083 | 0,079 | 0,102 | 0,105 |
| 17 | 0,154 | 0,149 | 0,179 | 0,143 |
| 18 | 0,191 | 0,220 | 0,257 | 0,249 |
| 19 | 0,254 | 0,249 | 0,268 | 0,307 |
| 20 | 0,304 | 0,291 | 0,299 | 0,302 |
| 21 | 0,278 | 0,287 | 0,313 | 0,334 |
| 22 | 0,282 | 0,224 | 0,306 | 0,303 |
| 23 | 0,284 | 0,227 | 0,303 | 0,328 |
| 24 | 0,269 | 0,237 | 0,297 | 0,296 |
| 25 | 0,279 | 0,231 | 0,260 | 0,283 |
| 26 | 0,237 | 0,171 | 0,264 | 0,301 |
| 27 | 0,237 | 0,223 | 0,318 | 0,298 |
| 28 | 0,231 | 0,211 | 0,213 | 0,263 |
| 29 | 0,204 | 0,221 | 0,241 | 0,266 |
| 30 | 0,222 | 0,180 | 0,268 | 0,260 |
| 31 | 0,181 | 0,168 | 0,192 | 0,225 |
| 32 | 0,206 | 0,234 | 0,237 | 0,239 |
| 33 | 0,190 | 0,202 | 0,233 | 0,189 |
| 34 | 0,152 | 0,242 | 0,251 | 0,191 |
| 35 | 0,224 | 0,137 | 0,225 | 0,213 |
| 36 | 0,191 | 0,152 | 0,148 | 0,222 |
| 37 | 0,129 | 0,117 | 0,218 | 0,183 |
| 38 | 0,138 | 0,136 | 0,145 | 0,191 |
| 39 | 0,143 | 0,095 | 0,168 | 0,115 |
| 40 | 0,119 | 0,128 | 0,071 | 0,135 |
| 41 | 0,114 | 0,068 | 0,106 | 0,064 |
| 42 | 0,074 | 0,045 | 0,069 | 0,067 |
| 43 | 0,063 | 0,061 | 0,100 | 0,072 |
| 44 | 0,030 | 0,074 | 0,042 | 0,075 |
| 45 | 0,036 | 0,035 | 0,024 | 0,030 |
| 46 | 0,055 | 0,066 | 0,104 | 0,039 |
| 47 | 0,000 | 0,011 | 0,029 | 0,019 |
| 48 | 0,000 | 0,000 | 0,030 | 0,000 |
| 49 | 0,000 | 0,000 | 0,000 | 0,000 |
| Indice synthétique de fécondité | 5,61 | 5,20 | 6,32 | 6,36 |

à partir de taux par année d'âge. Ceux-ci font cependant preuve d'une grande variabilité, même après avoir regroupé les trois années de 2002 à 2004.

Le regroupement par classe d'âge quinquennale donne les résultats du tableau 12.5.

TABLEAU 12.5 Taux de fécondité par groupe d'années d'âge et année civile, Malawi EDS 2004

| Groupe d'âge | 2002 | 2003 | 2004 | 2002-4 | DHS* |
|--|-------------|-------------|-------------|-------------|-------------|
| 15-19 | 0,151 | 0,180 | 0,178 | 0,169 | 0,162 |
| 20-24 | 0,254 | 0,304 | 0,312 | 0,290 | 0,293 |
| 25-29 | 0,210 | 0,261 | 0,283 | 0,252 | 0,254 |
| 30-34 | 0,204 | 0,235 | 0,222 | 0,221 | 0,222 |
| 35-39 | 0,129 | 0,180 | 0,184 | 0,164 | 0,163 |
| 40-44 | 0,075 | 0,078 | 0,086 | 0,080 | 0,080 |
| 45-49 | 0,042 | 0,049 | 0,021 | 0,036 | 0,035 |
| Indice synthétique de fécondité | 5,32 | 6,44 | 6,43 | 6,05 | 6,05 |

* Les taux pour la période de trois années ayant précédé l'EDS de 2004 .

Source : Measure DHS StatCompiler

Les différences entre les deux dernières colonnes entre les taux de fécondité par âge obtenus ici et ceux figurant dans l'enquête EDS sont minimales. Toutefois, les taux de fécondité beaucoup plus faibles pour 2002 (et 2001, qui ne figurent pas ici) méritent qu'on prête attention à de possibles erreurs concernant les périodes de référence ainsi qu'à des reports de naissances.

RÉFÉRENCES

- Cleland J. 1996. "Demographic data collection in less developed countries", *Population Studies* **50**(3):433-450. doi: <http://dx.doi.org/10.1080/0032472031000149556>
- Rutstein S and G Rojas. 2003. *Guide to DHS Statistics*. Calverton, MD: ORC Macro.
- Schoumaker B. 2010. "Reconstructing fertility trends in sub-Saharan Africa by combining multiple surveys affected by data quality problems " Paper presented at Population Association of America 2010 Annual Meeting, Dallas, TX, April 15-17, 2010.
- Schoumaker B. 2011. "Omissions of births in DHS birth histories in sub-Saharan Africa: Measurement and determinants " Paper presented at Population Association of America 2011 Annual Meeting, Washington D.C., March 31 - April 2, 2011.
- Schoumaker B. 2013. "A Stata module for computing fertility rates and TFRs from birth histories: tfr2", *Demographic Research* **28**(Article 38):1093-1144. doi: <http://doi.org/10.4054/DemRes.2013.28.38>

CHAPITRE 13 FÉCONDITÉ PAR COHORTE ET PÉRIODE

Tom A Moultrie

DESCRIPTION DE LA MÉTHODE

La disponibilité de données démographiques détaillées et d'histoires génésiques recueillies dans des enquêtes démographiques (comme l'Enquête mondiale de fécondité dans les années 1970 et le programme en cours des Enquêtes démographiques et de santé organisé par ORC Macro) a conduit à préférer généralement les mesures directes d'estimation de la fécondité aux mesures indirectes. Toutefois, l'extension de méthodes indirectes à des situations où des données détaillées existent peut non seulement offrir le moyen de corroborer les résultats obtenus directement mais aussi fournir des informations importantes sur la qualité des données recueillies grâce aux histoires génésiques.

Une de ces extensions consiste à appliquer aux histoires génésiques la logique de la méthode du quotient *P/F* de Brass, et à conduire ainsi une étude détaillée des données de fécondité par âge, période et cohorte. La méthode permet d'obtenir des estimations de l'indice synthétique de fécondité (TF, *Total Fertility*) pour une ou deux périodes de cinq ans précédant la collecte des données. Elle permet aussi d'identifier des erreurs courantes dans les données.

DONNÉES NÉCESSAIRES ET HYPOTHÈSES

Tableaux de données nécessaires

- Nombres de femmes par groupe d'âge à la date d'enquête
- Nombres de naissances par groupe d'âge (actuel) de la mère, par période de cinq ans avant l'enquête. L'élaboration de ce tableau nécessite, pour chaque enregistrement dans une histoire génésique,
 - la date (mois et année) de l'interview ;
 - la date (mois et année) de naissance de l'enfant ; et
 - le groupe d'âge actuel de la mère.

Hypothèses

Il n'y a pas de différence de fécondité entre les femmes interrogées dans l'enquête et celles qui ne l'ont pas été parce qu'elles sont décédées ou ont émigré.

TRAVAIL PRÉPARATOIRE ET RECHERCHES PRÉLIMINAIRES

On peut appliquer la méthode en travaillant avec les procédures de traitement des dates disponibles dans la plupart des logiciels statistiques, ou (presque aussi bien) en utilisant les dates dans le format CMC des EDS (en mois à partir du début du 20^e siècle). Si les dates de naissance et d'interview n'ont pas été codées ainsi, nous recommandons de le faire pour appliquer la présente méthode. La procédure est décrite en fin d'ouvrage.

APPLICATION DE LA MÉTHODE

L'application de la méthode se fait selon les étapes suivantes.

Étape 1 : Déterminer le nombre (pondéré) de femmes par groupe d'âge à la date d'enquête

Le tableau est construit aisément. Avec les données des EDS, la variable *v013* donne le groupe d'âge des femmes, et *v005* (divisé par 10^6 si nécessaire) les pondérations de l'échantillon. On désigne par $N_i = {}_5N_x$ le nombre de femmes dans le groupe d'âge i , avec $x = 15, 20, \dots, 45$ et $i = x/5 - 2$.

Étape 2 : Déterminer les nombres de naissances par groupe d'âge et période avant l'enquête

Si on travaille avec des dates au format CMC, avec une histoire génésique complète dans un fichier avec un enregistrement par enfant, l'âge actuel de l'enfant est facilement calculé en retranchant le CMC de sa date de naissance du CMC de la date d'interview. En divisant le résultat par 60 et en prenant la partie entière du résultat on obtient un index qui permet d'allouer la date de naissance de l'enfant à l'une des périodes quinquennales antérieures à l'enquête.

Une modification minimale est nécessaire dans les cas où l'enfant est né le même mois que l'interview, exactement 5 ans, 10 ans, ... auparavant. Selon la position respective du jour d'interview et du jour de naissance, les enfants peuvent être dans l'un ou l'autre des deux groupes d'âge adjacents. Pour résoudre ce problème et éviter de placer tous ces cas dans le même groupe, les enfants des mois-frontières doivent être alloués aux groupes d'âge en fonction de la date d'interview, si elle est connue, et en supposant que les jours de naissance sont répartis uniformément chaque mois. Quand c'est possible, nous définissons $b = 1$ si le jour (au cours du mois calendaire) de l'interview $< 16 -$ l'anniversaire de l'enfant étant alors plus vraisemblablement dans la seconde moitié du mois – et $b = 0$ dans les autres cas.

Donc

$$\text{Moment de la naissance} = j = \text{int} \left(\frac{\text{DoI} - \text{DoB}^c - b}{60} \right);$$

$$j = 0, 1, 2, \dots \quad (1)$$

où DoI est la date de l'interview et DoB^c la date de naissance de l'enfant, toutes deux enregistrées en format CMC. Dans le cas des données EDS, DoI est fourni par la variable $v008$ et DoB^c par la variable $b3$. Le jour d'interview est donné par la variable $v016$.

On construit ensuite un tableau croisé (pondéré, si nécessaire, pour tenir compte du plan d'échantillonnage) du groupe d'âge de la mère à la date d'enquête par un regroupement de la variable « moment de la naissance » définie ci-dessus. La structure du tableau croisé est illustrée au tableau 13.1, où les B_{ij} représentent le nombre agrégé (pondéré) d'enfants nés il y a j années de mères appartenant au groupe d'âge i à la date d'enquête :

Notez que, lorsqu'on remonte dans le temps, les taux de fécondité des femmes les plus jeunes sont nuls pour des périodes où toutes ces femmes ont moins de 10 ans. Certaines des naissances passées ne sont pas déclarées si les histoires génésiques ne sont recueillies qu'auprès des femmes âgées de moins de 50 ans.

Étape 3 : Calculer les taux de fécondité par cohorte et période fondés sur le groupe d'âge de la mère au moment de l'enquête

Si nous désignons les groupes d'âge (ou les cohortes définies par l'âge à l'enquête) par l'indice i ($i = 1$ correspondant au groupe 15-19 ans, etc.) et les périodes quinquennales suc-

TABEAU 13.1 Structure des tableaux utilisés dans le calcul des taux de fécondité par cohorte et période

| Groupe d'âge de la cohorte à l'enquête (i) | Nombre de femmes | Naissances par période avant l'enquête (j) | | | | |
|--|------------------|--|-----------------|-------------------|-------------------|-------------------|
| | | 0-4 ($j = 0$) | 5-9 ($j = 1$) | 10-14 ($j = 2$) | 15-19 ($j = 3$) | 20-24 ($j = 4$) |
| 15-19 ($i = 1$) | N_1 | $B_{1,0}$ | $B_{1,1}$ | | | |
| 20-24 (2) | N_2 | $B_{2,0}$ | $B_{2,1}$ | $B_{2,2}$ | | |
| 25-29 (3) | N_3 | $B_{3,0}$ | $B_{3,1}$ | $B_{3,2}$ | $B_{3,3}$ | |
| 30-34 (4) | N_4 | $B_{4,0}$ | $B_{4,1}$ | $B_{4,2}$ | $B_{4,3}$ | $B_{4,4}$ |
| 35-39 (5) | N_5 | $B_{5,0}$ | $B_{5,1}$ | $B_{5,2}$ | $B_{5,3}$ | $B_{5,4}$ |
| 40-44 (6) | N_6 | $B_{6,0}$ | $B_{6,1}$ | $B_{6,2}$ | $B_{6,3}$ | $B_{6,4}$ |
| 45-49 (7) | N_7 | $B_{7,0}$ | $B_{7,1}$ | $B_{7,2}$ | $B_{7,3}$ | $B_{7,4}$ |

cessives antérieures à l'enquête par l'indice j ($j = 0$ correspondant à la période qui précède immédiatement l'enquête et se termine à la date de celle-ci), les taux de fécondité par cohorte et période sont définis par

$$f_{i,j} = \frac{1}{5} \left(\frac{B_{i,j}}{N_i} \right).$$

Le rapport est divisé par cinq car les femmes sont exposées cinq ans exactement, toutes étant vivantes au long de chacune des périodes antérieures à l'enquête.

Les taux par cohorte et période mesurent la fécondité des femmes d'une même cohorte (nées au cours de la même période) en ligne, les périodes étant en colonne, et les âges atteints équivalents diminuant le long des diagonales (tableau 13.2).

TABEAU 13.2 Taux de fécondité par cohorte et période, classés par âge de la cohorte à l'enquête

| Groupe d'âge de la cohorte à l'enquête (i) | Taux de fécondité par période avant l'enquête (j) | | | | |
|--|---|-----------------|-------------------|-------------------|-------------------|
| | 0-4 ($j = 0$) | 5-9 ($j = 1$) | 10-14 ($j = 2$) | 15-19 ($j = 3$) | 20-24 ($j = 4$) |
| 15-19 ($i = 1$) | $f_{1,0}$ | $f_{1,1}$ | | | |
| 20-24 (2) | $f_{2,0}$ | $f_{2,1}$ | $f_{2,2}$ | | |
| 25-29 (3) | $f_{3,0}$ | $f_{3,1}$ | $f_{3,2}$ | $f_{3,3}$ | |
| 30-34 (4) | $f_{4,0}$ | $f_{4,1}$ | $f_{4,2}$ | $f_{4,3}$ | $f_{4,4}$ |
| 35-39 (5) | $f_{5,0}$ | $f_{5,1}$ | $f_{5,2}$ | $f_{5,3}$ | $f_{5,4}$ |
| 40-44 (6) | $f_{6,0}$ | $f_{6,1}$ | $f_{6,2}$ | $f_{6,3}$ | $f_{6,4}$ |
| 45-49 (7) | $f_{7,0}$ | $f_{7,1}$ | $f_{7,2}$ | $f_{7,3}$ | $f_{7,4}$ |

Étape 4 : Transposer les taux de fécondité par cohorte et période

Les taux obtenus à l'étape 3 peuvent aussi être classés par âge de la mère au terme de chacune des périodes quinquennales successives. Le terme de la période quinquennale précédant l'enquête (quand $j = 0$) est la date d'enquête, et le terme de la période 5-9 ans avant l'enquête (quand $j = 1$) est le point exactement cinq ans avant l'enquête. Ce reclassement a pour conséquence la création d'une série révisée d'indices de fécondité par cohorte :

$$f^*_{k,j} = f^k_{j,j}$$

Avec ce reclassement, le tableau 13.2 ci-dessus est réorganisé en tableau 13.3 ci-dessous. Ainsi, la fécondité des femmes âgées de 30-34 ans à l'enquête au cours de la période 10-14 ans avant l'enquête (i.e. $f^*_{4,2}$) est remaniée en fécondité des femmes qui étaient âgées de 20-24 ans dix ans avant l'enquête ($f^*_{2,2}$).

TABLEAU 13.3 Matrice des taux de fécondité par cohorte et période, après redéfinition de l'âge

| Groupe d'âge de la cohorte à la fin de chaque période (k) | Taux de fécondité par période avant l'enquête (j) | | | | |
|---|---|--------------------|----------------------|----------------------|----------------------|
| | 0-4 ($j = 0$) | 5-9 ($j = 1$) | 10-14 ($j = 2$) | 15-19 ($j = 3$) | 20-24 ($j = 4$) |
| 15-19 ($k=1$) | $f^*_{1,0}$ | $f^*_{1,1}$ | $f^*_{1,2}$ | $f^*_{1,3}$ | $f^*_{1,4}$ |
| 20-24 (2) | $f^*_{2,0}$ | $f^*_{2,1}$ | $f^*_{2,2}$ | $f^*_{2,3}$ | $f^*_{2,4}$ |
| 25-29 (3) | $f^*_{3,0}$ | $f^*_{3,1}$ | $f^*_{3,2}$ | $f^*_{3,3}$ | $f^*_{3,4}$ |
| 30-34 (4) | $f^*_{4,0}$ | $f^*_{4,1}$ | $f^*_{4,2}$ | $f^*_{4,3}$ | |
| 35-39 (5) | $f^*_{5,0}$ | $f^*_{5,1}$ | $f^*_{5,2}$ | | |
| 40-44 (6) | $f^*_{6,0}$ | $f^*_{6,1}$ | | | |
| 45-49 (7) | $f^*_{7,0}$ | | | | |

À mesure qu'on remonte dans le temps, les périodes ont un nombre régulièrement croissant de valeurs manquantes pour les femmes âgées si les histoires génésiques n'ont pas été recueillies auprès des femmes de plus de 50 ans. Par exemple, $f^*_{6,3}$ désigne la fécondité qu'ont eue il y a 15-19 ans les femmes âgées de 40-44 ans quinze ans exactement avant la date d'enquête. A l'enquête, ces femmes seraient âgées de 55-59 ans et n'ont donc pas été concernées par la section des histoires génésiques d'une enquête EDS.

Étape 5 : Calculer des mesures de fécondité par cohorte

Soit $P_{k,j}$ la fécondité cumulée par cohorte (c'est-à-dire le nombre moyen d'enfants déjà nés) de 15 ans à la fin du groupe d'âge k de la cohorte de femmes âgées de k au temps j . Donc :

$$P_{k,j} = 5 \cdot \sum_{z=0}^{k-1} f^*_{k-z,z+j} = 5 \cdot \sum_{z=0}^{k-1} f_{k+j-z,z+j}$$

Étape 6 : Calculer des mesures de la fécondité par période et deux estimations de l'indice synthétique de fécondité

Les mesures de fécondité par période sont des taux cumulés de fécondité au cours d'une période donnée. Nous désignons par $F_{i,j}$ la fécondité cumulée jusqu'à l'âge i lors de la période j . Donc,

$$F_{k,j} = 5 \cdot \sum_{z=1}^k f^*_{z,j} = 5 \cdot \sum_{z=1}^k f_{z+j,j}$$

Notez que $F_{7,0}$ est une mesure de l'indice synthétique de fécondité (TF) de la période cinq ans précédant immédiatement l'enquête. On peut supposer que cette estimation s'applique (approximativement) 2,5 ans avant la date d'enquête.

La plupart du temps, $F_{7,1}$ ne peut pas être évalué directement, car il faudrait que les femmes aujourd'hui âgées de 50-54 aient été interrogées sur la fécondité qu'elles ont eue quand elles avaient 45-49 ans dans la période quinquennale se terminant cinq ans avant l'enquête. Mais la fécondité dans ce groupe d'âge est généralement très faible, et une estimation approchée de la fécondité dans la période 5-9 ans avant l'enquête peut être obtenue en faisant $TF_1 = F_{6,1} + 5 \cdot f^*_{7,0}$.

En d'autres termes, nous supposons que $f^*_{7,1} = f^*_{7,0} = f_{7,0}$. Si la fécondité décroît, l'estimation sera un peu trop basse, mais comme la fécondité est généralement très faible dans ce groupe d'âge, le biais sera sans importance.

Dans les populations où la fécondité est basse ou assez basse (moins de 3 enfants par femme), il est raisonnable de procéder de même pour la fécondité non mesurée des femmes âgées de 40-49 ans dix ans exactement avant l'enquête, car la fécondité à 40-44 ans y est suffisamment faible pour que de légères variations aient un impact négligeable sur l'indice synthétique estimé 10-14 ans avant l'enquête. Dans ce cas, nous supposons que $f^*_{7,2} = f_{7,0}$ et $f^*_{6,2} = f_{6,1}$ pour pouvoir écrire $TF_2 = F_{5,2} + 5 \cdot f^*_{7,0} + 5 \cdot f^*_{6,1}$.

Étape 7 : Calculer les quotients P/F

Nous pouvons calculer directement les quotients P/F à partir des résultats issus des étapes 5 et 6. Le quotient P/F applicable au groupe k à la période j est égal à

$$\frac{P}{F}(k, j) = \frac{P_{k,j}}{F_{k,j}} = \frac{5 \cdot \sum_{z=0}^{k-1} f_{k-z,z+j}^*}{5 \cdot \sum_{z=1}^k f_{z,j}^*} = \frac{5 \cdot \sum_{z=0}^{k-1} f_{k+j-z,z+j}}{5 \cdot \sum_{z=1}^k f_{z+j,j}}$$

Par exemple, le quotient P/F pour les femmes âgées de 25-29 ans dans la période se terminant dix ans avant l'enquête est, avec $k = 3$ et $j = 2$ dans la formule ci-dessus,

$$\begin{aligned} \frac{P}{F}(3,2) &= \frac{P_{3,2}}{F_{3,2}} = \frac{\sum_{z=0}^2 f_{3-z,2+z}^*}{\sum_{z=1}^3 f_{z,2}^*} = \frac{f_{3,2}^* + f_{2,3}^* + f_{1,4}^*}{f_{1,2}^* + f_{2,2}^* + f_{3,2}^*} \\ &= \frac{\sum_{z=0}^2 f_{5,2+z}}{\sum_{z=1}^3 f_{2+z,2}} = \frac{f_{5,2} + f_{5,3} + f_{5,4}}{f_{3,2} + f_{4,2} + f_{5,2}} \end{aligned}$$

INTERPRÉTATION ET DIAGNOSTICS

Plusieurs interprétations importantes émergent de ces résultats.

1. Estimations de la fécondité par période

On a vu, à l'étape 6, comment deux estimations de la fécondité pouvaient être obtenues, applicables environ 2,5 et 7,5 ans avant l'enquête. On peut en déduire une évolution à court terme de la fécondité.

2. Interprétation des quotients P/F ratios et calendrier de la baisse de la fécondité

Les quotients P/F obtenus à l'étape 7 peuvent nous donner des informations sur la nature et le calendrier de la baisse de la fécondité, ainsi que sur les problèmes de qualité des données. La section de ce manuel consacrée à quotient P/F de Brass décrit les traits essentiels de la méthode.

Des quotients P/F de, ou très proches de, 1 à chaque âge dans une période donnée impliquent qu'il n'y a pas eu de variation de la fécondité, les mesures de la fécondité par période et par cohorte étant à peu près égales. Une baisse de la fécondité se traduit par une augmentation continue des quotients P/F au fil des âges une période donnée, à partir

de valeurs proches de l'unité avant 25 ans. (Nous utilisons 25 ans comme limite car il est difficile pour la fécondité par cohorte et par période de différer sensiblement pour les cohortes les plus jeunes.) Ainsi, si les quotients par âge P/F sont à peu près constants dans une période avant l'enquête, j , mais qu'ils font apparaître une nette tendance en fonction de l'âge au cours de la période suivante plus proche de l'enquête, $j - 1$, on peut en déduire que la fécondité a commencé à baisser (à peu près) à la date séparant les deux périodes.

Une série de quotients P/F faibles au cours d'une période suivie ou précédée par une série de quotients beaucoup plus élevés indique un éventuel déplacement des naissances vers la période où les quotients sont bas et au détriment de la période où ils sont exagérément élevés. De même, une série de quotients P/F qui s'éloigne sans explication de la tendance générale le long d'une diagonale (c'est-à-dire pour une cohorte particulière) suggère une mauvaise déclaration des âges par les femmes ou des omissions de naissances si la tendance est observée aux âges avancés.

3. Évaluation de la qualité des données

L'examen des taux de fécondité par cohorte et période (les f^*) calculés à l'étape 4 peut contribuer à l'évaluation de la qualité des données. Par exemple, la lecture de droite à gauche le long des lignes montre comment évolue la fécondité de chaque groupe d'âge à mesure que s'approche la date d'enquête.

En l'absence de facteurs exogènes importants, on s'attend à des variations progressives et régulières, et les écarts à ce schéma peuvent refléter des erreurs sur la période de référence ou d'autres omissions. Trois types d'erreurs sur la période de référence sont courants dans les données d'histoires génésiques rétrospectives recueillies par enquête.

Le premier type d'erreur a été signalé par Brass, selon qui les femmes âgées tendent à exagérer l'âge de leurs aînés en situant leurs dates de naissance trop loin dans le passé. Il en résulte une surestimation du niveau de la fécondité des périodes anciennes et une sous-estimation de la fécondité récente du fait du transfert des naissances de périodes relativement récentes vers des périodes plus anciennes, exagérant ainsi la baisse apparente de la fécondité. Les « effets de Brass » sont caractérisés par des taux de fécondité par cohorte et période nettement plus élevés dans tous les groupes d'âge que ceux des cohortes un peu plus récentes aux mêmes âges. Ce décalage des naissances vers le passé a aussi pour consé-

quence une sous-estimation de la baisse de la fécondité dans les périodes récentes.

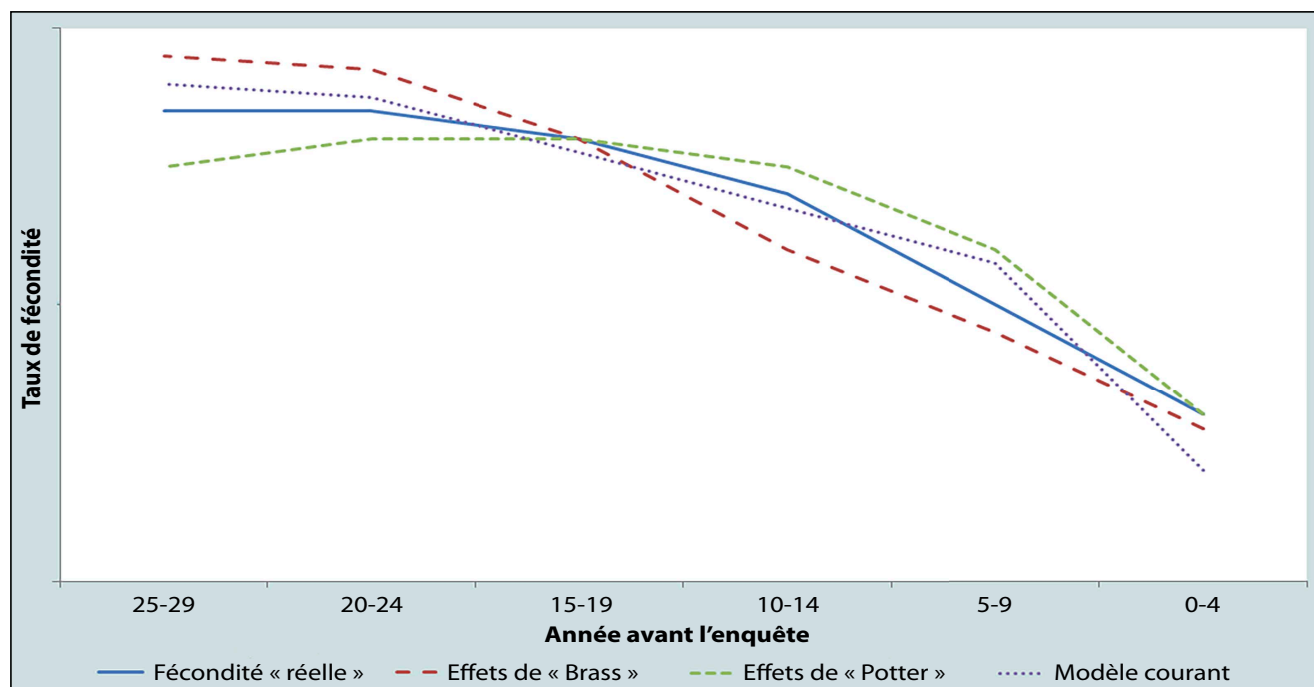
Le deuxième type d'erreur sur la période de référence a été identifié par Potter (1977). Selon lui, les femmes tendent à rapprocher les naissances anciennes de la date d'interview, mais à déclarer correctement les événements récents. Il en résulte une sous-estimation du niveau de la fécondité des périodes anciennes, des taux de fécondité récents correctement estimés et une surestimation de ceux de la période intermédiaire. Le modèle de Potter s'appuie sur deux propositions. La première : « la date d'un événement est mémorisée d'autant moins précisément qu'elle est ancienne. La seconde : si une histoire génésique est recueillie grâce à des questions sur les naissances vivantes dans l'ordre où elles sont survenues, la date qu'une femme attache à un événement autre que le premier est influencée par l'information qu'elle a déjà donnée sur l'événement précédent » (Potter 1977 : 341). Les « effets de Potter » jouent davantage quand les histoires génésiques sont recueillies dans l'ordre où les naissances se sont produites plutôt que dans l'ordre inverse, de la plus récente vers la plus ancienne.

Un troisième type d'erreur résulte de l'omission systématique des enfants nés juste avant l'enquête, ou de leur décalage vers une période plus ancienne. Il s'agit généralement pour un enquêteur d'éviter de devoir poser des questions détaillées supplémentaires (par exemple dans un questionnaire anthropométrique) sur les plus jeunes enfants (le plus souvent ceux de moins de cinq ans). Ces erreurs ont été bien démontrées par Cleland (1996) et Schoumaker (2010, 2011). Du fait de ces omissions ou de ces décalages, la baisse de la fécondité juste avant l'enquête se trouve exagérée et les quotients P/F de cette période très récente font apparaître une baisse de la fécondité largement accentuée. Une partie de cette baisse peut être réelle, mais les analystes doivent être vigilants sur les effets éventuels de ce type d'omission ou de décalage.

L'impact de ces trois effets sur les évolutions de la fécondité est illustré graphiquement sur la figure 13.1.

La courbe intitulée « fécondité réelle » retrace l'évolution de la fécondité (totale ou par âge) au fil du temps dans une population hypothétique. Les effets de « Brass » donnent l'impression d'une fécondité plus élevée pour le passé lointain et une baisse plus lente environ dans les 10 années

FIGURE 13.1 Représentation graphique des effets de Brass et Potter sur les erreurs de déclaration de la fécondité



précédant l'enquête. Les effets de « Potter » entraînent une surestimation systématique de la fécondité au cours de 5-15 ans précédant l'enquête, donnant l'impression fautive d'une baisse récente plus rapide de la fécondité. Le « modèle typique » indique la nature des distorsions fréquentes dans les données d'histoires génésiques. La fécondité de la période la plus récente est généralement trop basse, du fait de l'omission (dans les recensements) ou du décalage des naissances récentes pour éviter les modules complémentaires d'anthropométrie, etc. (dans les enquêtes), alors que la fécondité du passé très ancien est souvent surestimée (par le jeu des effets de « Brass ») et se marque par un excès apparent des naissances chez les très jeunes femmes dans de nombreuses histoires génésiques.

EXEMPLE

Notre exemple s'appuie sur les données d'histoires génésiques rétrospectives recueillies lors de l'EDS du Malawi en 2004. Les tableaux ont été pondérés grâce aux poids d'échantillonnage fournis avec les données, ce qui explique les décimales dans les tableaux d'effectifs. La méthode est mise en application dans un dossier Excel associé disponibles à <http://demographicestimation.iussp.org/fr>.

Étapes 1 et 2 : Extraction des données

Les tableaux tirés des fichiers de données de l'EDS sont illustrés par le tableau 13.4. Les naissances déclarées par période

antérieure à l'enquête ont été ajustées pour tenir compte approximativement du problème de frontière évoqué lors de l'établissement de l'équation 1.

Étape 3 : Calculer les taux de fécondité par cohorte et période classés par groupe d'âge de la mère au moment de l'enquête

Le tableau 13.5 présente les taux de fécondité par cohorte et période tirés des données du tableau 13.4. Par exemple, le taux de fécondité par cohorte associé aux naissances survenues 5-9 ans avant l'enquête chez les femmes âgées de 20-24 ans à l'enquête est égal à $(1/5) \cdot (981,8 / 2\ 869,7) = 0,068$.

Étape 4 : Transposer les taux de fécondité par cohorte et période

Les taux du tableau 13.5 sont transposés de sorte que les lignes représentent les âges atteints équivalents à la fin de chacune des périodes en colonne. Par exemple, une femme âgée de 30-34 ans à l'enquête était âgée de 25-29 ans à la fin de la période 5-9 ans avant l'enquête (et de 20-24 ans à la fin de la période 10-14 ans avant l'enquête), et ces taux de fécondité par cohorte (0,291 et 0,270 au tableau 13.5) sont maintenant classés à 25-29 et 20-24 ans respectivement (tableau 13.6).

Pour faciliter l'identification des tendances de la fécondité et des problèmes ou des déficiences dans les données, on a représenté graphiquement (figure 13.2) les taux de fécondité

TABLEAU 13.4 Nombre de femmes par groupe d'âge à l'enquête, et nombre de naissance de ces femmes par période de naissance, Malawi, EDS 2004

| Groupe d'âge de la cohorte à l'enquête (<i>i</i>) | Années de naissance approximatives de la cohorte | Nombre de femmes | Naissances par période avant l'enquête (<i>j</i>) | | | | | | |
|---|--|------------------|---|---------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| | | | 0-4 (<i>j</i> = 0) | 5-9 (<i>j</i> = 1) | 10-14 (<i>j</i> = 2) | 15-19 (<i>j</i> = 3) | 20-24 (<i>j</i> = 4) | 25-29 (<i>j</i> = 5) | 30-34 (<i>j</i> = 6) |
| 15-19 (<i>i</i> = 1) | 1985-1989 | 2 392,0 | 713,7 | 5,2 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 |
| 20-24 (2) | 1980-1984 | 2 869,7 | 3 638,8 | 981,8 | 28,7 | 0,0 | 0,0 | 0,0 | 0,0 |
| 25-29 (3) | 1975-1979 | 2 157,4 | 2 952,3 | 2 693,6 | 859,1 | 13,6 | 0,0 | 0,0 | 0,0 |
| 30-34 (4) | 1970-1974 | 1 478,0 | 1 734,4 | 2 152,7 | 1 996,7 | 595,5 | 21,9 | 0,0 | 0,0 |
| 35-39 (5) | 1965-1969 | 1 116,8 | 1 139,6 | 1 462,6 | 1 815,5 | 1 386,4 | 508,4 | 18,1 | 0,0 |
| 40-44 (6) | 1960-1964 | 935,0 | 569,3 | 923,5 | 1 372,6 | 1 456,2 | 1 267,8 | 386,4 | 13,4 |
| 45-49 (7) | 1955-1959 | 749,1 | 235,1 | 558,8 | 952,6 | 1 024,9 | 1 128,3 | 953,3 | 311,6 |

| Groupe d'âge de la cohorte à l'enquête (<i>i</i>) | Période avant l'enquête (<i>j</i>) | | | | | | |
|---|--------------------------------------|------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| | 0-4 (<i>j</i> = 0) | 5-9 (<i>j</i> = 1) | 10-14 (<i>j</i> = 2) | 15-19 (<i>j</i> = 3) | 20-24 (<i>j</i> = 4) | 25-29 (<i>j</i> = 5) | 30-34 (<i>j</i> = 6) |
| 15-19 (<i>i</i> = 1) | 0,060 | 0,000 | | | | | |
| 20-24 (2) | 0,254 | 0,068 | 0,002 | | | | |
| 25-29 (3) | 0,274 | 0,250 | 0,080 | 0,001 | | | |
| 30-34 (4) | 0,235 | 0,291 | 0,270 | 0,081 | 0,003 | | |
| 35-39 (5) | 0,204 | 0,262 | 0,325 | 0,248 | 0,091 | 0,003 | |
| 40-44 (6) | 0,122 | 0,198 | 0,294 | 0,311 | 0,271 | 0,083 | 0,003 |
| 45-49 (7) | 0,063 | 0,149 | 0,254 | 0,274 | 0,301 | 0,255 | 0,083 |

TABLERAU 13.5 Taux de fécondité par cohorte et période, classés par groupe d'âge à l'enquête, Malawi, EDS 2004

| Groupe d'âge de la cohorte à la fin de chaque période avant l'enquête, (<i>k</i>) | Période avant l'enquête (<i>j</i>) | | | | | | |
|---|--------------------------------------|------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| | 0-4 (<i>j</i> = 0) | 5-9 (<i>j</i> = 1) | 10-14 (<i>j</i> = 2) | 15-19 (<i>j</i> = 3) | 20-24 (<i>j</i> = 4) | 25-29 (<i>j</i> = 5) | 30-34 (<i>j</i> = 6) |
| 15-19 (<i>k</i> = 1) | 0,060 | 0,068 | 0,080 | 0,081 | 0,091 | 0,083 | 0,083 |
| 20-24 (2) | 0,254 | 0,250 | 0,270 | 0,248 | 0,271 | 0,255 | |
| 25-29 (3) | 0,274 | 0,291 | 0,325 | 0,311 | 0,301 | | |
| 30-34 (4) | 0,235 | 0,262 | 0,294 | 0,274 | | | |
| 35-39 (5) | 0,204 | 0,198 | 0,254 | | | | |
| 40-44 (6) | 0,122 | 0,149 | | | | | |
| 45-49 (7) | 0,063 | | | | | | |

TABLERAU 13.6 Taux de fécondité par cohorte et période, classés par âge de la mère à la fin de chaque période avant l'enquête, Malawi, EDS 2004

par cohorte et période du tableau 13.6, par cohorte de naissance et âge atteint. Des omissions apparaissent systématiquement dans la fécondité des périodes les plus anciennes : les taux de fécondité par cohorte et période des femmes les plus âgées (la cohorte 1955-59) aux jeunes âges atteints (20-24 ans) sont plus faibles que ceux des cohortes un peu plus jeunes. Ces données révèlent néanmoins des signes d'un début de baisse de la fécondité au Malawi, les taux de fécondité dans les cohortes les plus jeunes (nées après 1980) étant plus faibles que ceux des cohortes plus âgées.

La dissociation de ces effets nécessiterait des recherches plus approfondies.

Étape 5 : Calculer des mesures de fécondité par cohorte

La fécondité par cohorte cumulée jusqu'à un âge donné est calculée en additionnant les taux par cohorte le long de la diagonale du tableau 13.6 et en multipliant par 5

(tableau 13.7). Par exemple, la fécondité cumulée des femmes âgées de 25-29 ans à la fin de la période 5-9 ans avant l'enquête est égale à $5 \cdot (0,291 + 0,270 + 0,081) = 3,210$.

Étape 6 : Calculer des mesures de la fécondité par période et deux estimations de l'indice synthétique de fécondité

La fécondité cumulée par période jusqu'à un âge donné est calculée en additionnant les taux de fécondité par cohorte et période dans une colonne (une période) donnée du tableau 13.6 jusqu'à cet âge, et en multipliant par 5 (tableau 13.8). Par exemple, la fécondité cumulée jusqu'à l'âge de 30 ans au cours de la période 5-9 ans avant l'enquête est égale à $5 \cdot (0,068 + 0,250 + 0,291) = 3,047$.

On calcule deux estimations de l'indice synthétique de fécondité. La première est celle de la période 0-4 ans avant l'enquête ; elle est calculée directement à partir des don-

FIGURE 13.2 Taux de fécondité par cohorte et période, Malawi EDS 2004

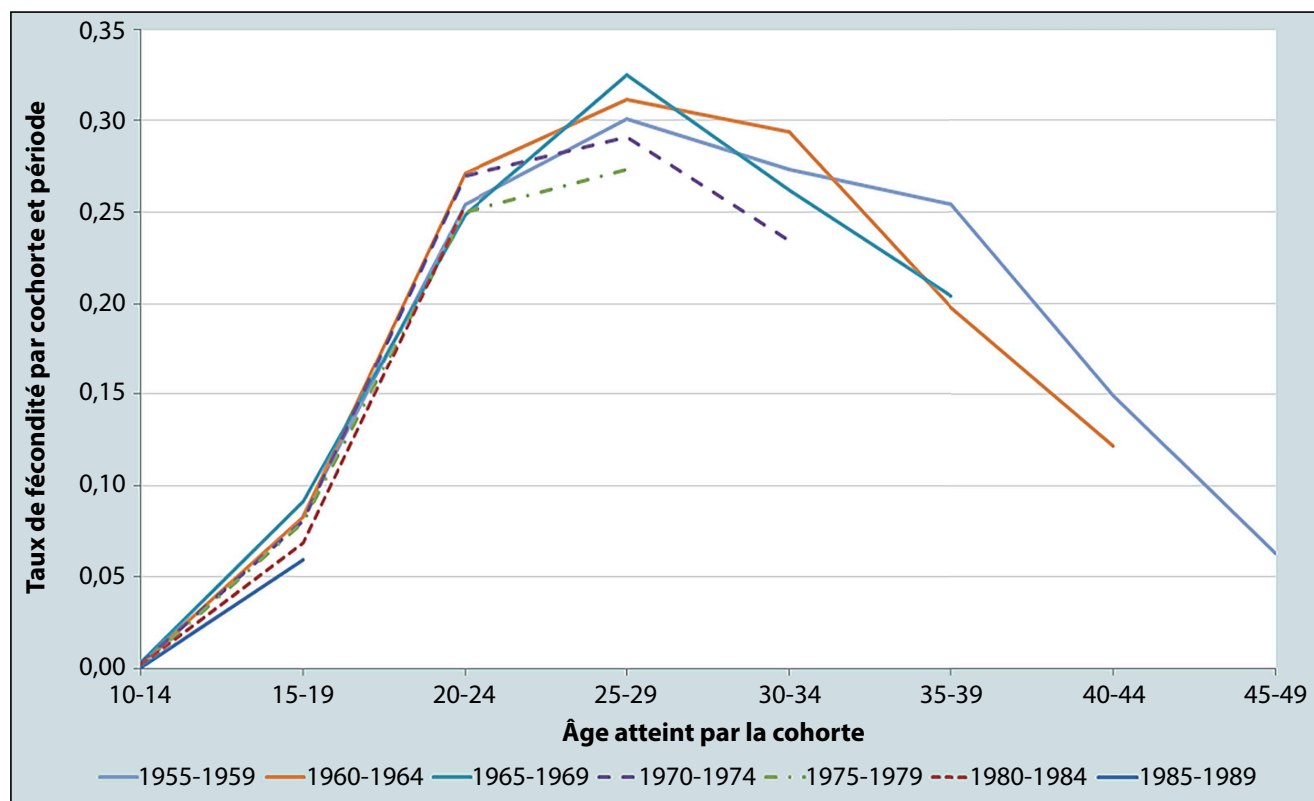


TABLEAU 13.7 Fécondité cumulée des cohortes à la fin de chaque période (P), Malawi, EDS 2004

| Groupe d'âge de la cohorte à la fin de chaque période avant l'enquête, (k) | Période avant l'enquête (j) | | | | | | |
|--|-----------------------------|-------------|---------------|---------------|---------------|---------------|---------------|
| | 0-4 (j = 0) | 5-9 (j = 1) | 10-14 (j = 2) | 15-19 (j = 3) | 20-24 (j = 4) | 25-29 (j = 5) | 30-34 (j = 6) |
| 15-19 (k = 1) | 0,298 | 0,342 | 0,398 | 0,403 | 0,455 | 0,413 | 0,416 |
| 20-24 (2) | 1,610 | 1,647 | 1,754 | 1,697 | 1,769 | 1,689 | |
| 25-29 (3) | 3,015 | 3,210 | 3,322 | 3,327 | 3,195 | | |
| 30-34 (4) | 4,384 | 4,632 | 4,795 | 4,563 | | | |
| 35-39 (5) | 5,652 | 5,783 | 5,835 | | | | |
| 40-44 (6) | 6,391 | 6,581 | | | | | |
| 45-49 (7) | 6,895 | | | | | | |

nées (6,1 enfants par femme). La fécondité de la période 5-9 ans avant l'enquête est obtenue en faisant $6,091 + 5 \cdot (0,063) = 6,406$ enfants par femme.

La date médiane des interviews de cette EDS étant en novembre 2004, on peut considérer que les deux estimations s'appliquent à mai 2002 et mai 1997 respectivement.

| Groupe d'âge de la cohorte à la fin de chaque période avant l'enquête, (<i>k</i>) | Période avant l'enquête (<i>j</i>) | | | | | | |
|---|--------------------------------------|------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| | 0-4 (<i>j</i> = 0) | 5-9 (<i>j</i> = 1) | 10-14 (<i>j</i> = 2) | 15-19 (<i>j</i> = 3) | 20-24 (<i>j</i> = 4) | 25-29 (<i>j</i> = 5) | 30-34 (<i>j</i> = 6) |
| 15-19 (<i>k</i> = 1) | 0,298 | 0,342 | 0,398 | 0,403 | 0,455 | 0,413 | 0,416 |
| 20-24 (2) | 1,566 | 1,591 | 1,749 | 1,644 | 1,811 | 1,686 | |
| 25-29 (3) | 2,935 | 3,045 | 3,375 | 3,202 | 3,317 | | |
| 30-34 (4) | 4,108 | 4,357 | 4,843 | 4,570 | | | |
| 35-39 (5) | 5,129 | 5,345 | 6,115 | | | | |
| 40-44 (6) | 5,738 | 6,091 | | | | | |
| 45-49 (7) | 6,051 | 6,406 | | | | | |

TABLERAU 13.8 Fécondité cumulée par période (F), Malawi, EDS 2004

| Groupe d'âge de la cohorte à la fin de chaque période avant l'enquête, (<i>k</i>) | Période avant l'enquête (<i>j</i>) | | | | | | |
|---|--------------------------------------|------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| | 0-4 (<i>j</i> = 0) | 5-9 (<i>j</i> = 1) | 10-14 (<i>j</i> = 2) | 15-19 (<i>j</i> = 3) | 20-24 (<i>j</i> = 4) | 25-29 (<i>j</i> = 5) | 30-34 (<i>j</i> = 6) |
| 15-19 (<i>k</i> = 1) | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 |
| 20-24 (2) | 1,028 | 1,035 | 1,003 | 1,032 | 0,977 | 1,002 | |
| 25-29 (3) | 1,027 | 1,054 | 0,984 | 1,039 | 0,963 | | |
| 30-34 (4) | 1,067 | 1,063 | 0,990 | 0,998 | | | |
| 35-39 (5) | 1,102 | 1,082 | 0,954 | | | | |
| 40-44 (6) | 1,114 | 1,080 | | | | | |
| 45-49 (7) | 1,139 | | | | | | |

TABLERAU 13.9 Quotients P/F, Malawi, EDS 2004

Ces résultats suggèrent une baisse de l'indice synthétique de fécondité de 0,4 enfant par femme entre les deux périodes, mais un décalage et une omission des naissances les plus récentes peuvent donner une impression exagérée de baisse.

Étape 7 : Calculer les quotients P/F

Les quotients P/F sont calculés en divisant les cellules équivalentes dans les tableaux 13.7 et 8 (tableau 13.9).

Les accroissements réguliers des quotients P/F au fil des âges au cours des deux périodes les plus récentes indiquent clairement qu'une baisse de la fécondité est en cours. Rien de semblable n'apparaît dans les quotients de la période 10-14 ans

avant l'enquête. La baisse de la fécondité au Malawi doit donc avoir débuté environ dix ans avant l'enquête, vers 1994.

Les quotients P/F montrent qu'il y a eu des omissions et des décalages de naissances dans certaines cohortes et à certaines périodes. Par exemple, les données sur les naissances survenues 10-14 ans avant l'enquête donnent des quotients P/F inférieurs à l'unité. En outre, le quotient relatif aux naissances 10-14 ans avant l'enquête chez les femmes âgées de 20-24 ans (1,003) s'écarte clairement des quotients des femmes du même âge 5-9 ans et 15-19 ans avant l'enquête. Ce pourrait être la conséquence d'une hausse de la fécondité au cours de la période, mais cela semble peu vraisemblable. Plus probablement – puisque les quotients sont trop bas au

cours de cette période – des naissances ont été décalées vers cette période (effets de Brass ou de Potter éventuels), gonflant donc les estimations de F et réduisant les quotients. Les femmes âgées de 45-49 ans semblent avoir omis certaines de

leurs naissances dans toutes les périodes 5-25 ans avant l'enquête, les quotients P/F pour cette cohorte sont plus faibles que ceux de la cohorte des femmes âgées de 40-44 ans à l'enquête.

AUTRES LECTURES ET RÉFÉRENCES

Le calcul des quotients P/F à partir des données d'enquête a été exposé au début des années 1980 par Hobcraft et autres (Goldman and Hobcraft 1982 ; Hobcraft, Goldman and Chidambaram 1982). Hobcraft, Goldman and Chidambaram (1982), dans leur présentation de la méthode, analysent les taux de fécondité par cohorte et période par durée de mariage (et âge au mariage) et par durée depuis la première naissance (et âge à la première naissance).

Au-delà des sources déjà citées, la méthode a été utilisée dans un certain nombre d'analyses des données de l'Enquête mondiale de fécondité et des EDS. Il y a eu par exemple des applications au Lesotho (Timæus and Balasubramanian 1984), au Zimbabwe (Muhwava and Timæus 1996), en Afrique de l'ouest (Onuoha and Timæus 1995) et au Népal (Collumbien, Timæus and Acharya 2001). Hinde and Mturi (2000) ont appliqué la méthode aux données tanzaniennes en utilisant la durée de mariage.

Cleland J. 1996. "Demographic data collection in less developed countries", *Population Studies* **50**(3):433-450. doi: <http://dx.doi.org/10.1080/0032472031000149556>

Collumbien M, IM Timæus and L Acharya. 2001. "Fertility decline in Nepal," in Sathar, ZA and JF Philips (eds). *Fertility Transition in South Asia*. Oxford: Oxford University Press, pp. 99-120.

Goldman N and J Hobcraft. 1982. *Birth Histories*. WFS Comparative Studies 17. Voorburg, Netherlands: International Statistical Institute.

Hinde A and AJ Mturi. 2000. "Recent trends in Tanzanian fertility", *Population Studies* **54**(2):177-191. doi: <http://dx.doi.org/10.1080/713779080>

Hobcraft JN, N Goldman and VC Chidambaram. 1982. "Advances in the P/F ratio method for the analysis of birth histories", *Population Studies* **36**(2):291-316. doi: <http://dx.doi.org/10.2307/2174202>

Muhwava W and IM Timæus. 1996. *Fertility Decline in Zimbabwe*. Centre for Population Studies Research Paper 96-1. London: London School of Hygiene & Tropical Medicine.

Onuoha NC and IM Timæus. 1995. "Has a fertility transition begun in West Africa?", *Journal of International Development* **7**(1):93-116. doi: <http://dx.doi.org/10.1002/jid.3380070107>

Potter JE. 1977. "Problems in using birth-history analysis to estimate trends in fertility", *Population Studies* **31**(2):335-364. doi: <http://dx.doi.org/10.2307/2173921>

Schoumaker B. 2010. "Reconstructing fertility trends in sub-Saharan Africa by combining multiple surveys affected by data quality problems" Paper presented at Population Association of America 2010 Annual Meeting, Dallas, TX, April 15-17, 2010.

Schoumaker B. 2011. "Omissions of births in DHS birth histories in sub-Saharan Africa: Measurement and determinants" Paper presented at Population Association of America 2011 Annual Meeting, Washington D.C., March 31 - April 2, 2011.

Timæus IM and K Balasubramanian. 1984. *Evaluation of the Lesotho Fertility Survey 1977*. WFS Scientific Reports, 58. Voorburg, Netherlands: International Statistical Institute.

Méthodes fondées
sur les données d'état civil

CHAPITRE 14 COMPARAISON DES PARITÉS PAR COHORTE ET DES DONNÉES D'ÉTAT CIVILS

Tom A Moultrie and Basia Zaba

DESCRIPTION DE LA MÉTHODE

Le modèle relationnel synthétique de Gompertz présente une méthode pour comparer les parités moyennes avec leurs équivalents calculés à partir des taux de fécondité par période sans avoir à supposer que la fécondité est constante. L'aspect le plus important de cette méthode est que les parités moyennes sont calculées pour une période plutôt que pour une série de cohortes. Elle nécessite cependant la disponibilité de données sur le nombre d'enfants déjà nés à deux dates, à cinq ou dix ans d'écart. S'il n'existe qu'une source de données sur le nombre d'enfants déjà nés, ou si la période entre deux enquêtes n'est pas (au moins approximativement) de cinq ou dix ans, on doit recourir à une procédure alternative qui ne requiert pas l'hypothèse de fécondité constante.

C'est le cas de la méthode présentée ici. Celle-ci a des avantages théoriques spécifiques, mais son application pratique nécessite une assez longue série de données annuelles sur les naissances classées par âge de la mère. Elle est principalement utilisée avec des données sur les naissances enregistrées à l'état civil, généralement la seule source d'information sur les naissances par année sur une période de 20 ans. Mais si des distributions de fécondité sont connues grâce à d'autres sources par intervalles quinquennaux (par exemple des histoires génésiques complètes recueillies à plusieurs reprises), elles peuvent aussi être utilisées. La description faite ici se réfère à des données d'un enregistrement à l'état civil. En général, si on utilise des parités jusqu'à l'âge x , il est nécessaire de disposer de données de fécondité pour au moins les $x-15$ années précédentes, afin de faire les comparaisons pertinentes.

La méthode s'appuie sur la nature longitudinale des parités moyennes déclarées et les compare avec leurs équivalents (descendances) tirés des taux de fécondité enregistrés appar-

tenant aux cohortes adéquates. Si on considère des femmes âgées de 30-34 ans lors d'un recensement, elles étaient âgées de 20-24 ans dix ans auparavant et de 10-14 ans vingt ans avant le recensement. Si la fécondité débute à 15 ans, les enfants déjà nés déclarés par des femmes âgées jusqu'à 35 ans au moment du recensement reflètent l'expérience cumulée de ces femmes en matière de fécondité au cours des vingt années passées. Si on peut supposer que la mortalité et la migration sont indépendantes de la fécondité de ces femmes et si des taux de fécondité peuvent être calculés pour ces vingt années, des parités moyennes équivalentes (descendances) peuvent être calculées pour chaque cohorte et comparées à la parité moyenne des femmes au moment du recensement.

La difficulté lors de l'application de ce principe général est qu'une cohorte représentée par un groupe d'âge quinquennal conventionnel au moment du recensement n'a pas appartenu à un groupe d'âge quinquennal conventionnel au cours de chacune des années antérieures. Ainsi, la population âgée de 30-34 ans au moment du recensement était âgée de 29-33 ans l'année précédente, de 28-32 ans deux années plus tôt, et ainsi de suite. Si les naissances enregistrées à l'état civil sont classées par année d'âge de la mère, le problème n'est pas important, car on peut calculer des taux de fécondité par année d'âge pour chaque année et les regrouper ensuite relativement facilement par cohorte. L'élaboration des tableaux et les calculs risquent cependant d'être longs et l'attraction de certains âges peut avoir un effet non négligeable sur les taux de fécondité. Il est donc commode de disposer d'une méthode qui puisse être appliquée à des taux par groupe quinquennal conventionnel.

Il est important de noter que la méthode peut être utilisée pour évaluer la complétude de l'enregistrement des naissances à l'état civil.

Dans sa formulation initiale, la méthode posait de nombreux problèmes d'interpolation décrits par le *Manuel X* (Division de la Population des NU 1984 : 46-56). Sa reformulation ici comme variante du modèle relationnel de Gompertz permet de les contourner.

DONNÉES NÉCESSAIRES

Les données nécessaires pour cette méthode sont :

- Le nombre d'enfants déjà nés, ou les parités moyennes, par groupe d'âge quinquennal de la mère, dans un recensement récent.
- Les naissances enregistrées à l'état civil par groupe d'âge quinquennal de la mère pour chacune des 15 ou 20 années précédant le recensement.
- Le nombre de femmes dans chaque groupe d'âge au recensement, et dans un ou plusieurs recensements antérieurs, pour permettre l'estimation de la population féminine par groupe d'âge quinquennal pour chacune des 15 ou 20 années précédant le recensement récent.

HYPOTHÈSES

La plupart des hypothèses sont celles associées au modèle relationnel de Gompertz, à savoir :

- Le schéma de fécondité standard choisi pour la procédure d'ajustement reflète de façon appropriée la forme de la distribution de la fécondité dans la population.
- Les changements de la fécondité ont été réguliers et graduels et ont affecté tous les groupes d'âge dans une mesure comparable.
- Les erreurs dans les taux de fécondité avant ajustement sont proportionnellement les mêmes parmi les femmes dans les groupes d'âge centraux (20-39 ans), de sorte que la forme de la fécondité par âge des naissances récentes déclarées est à peu près correcte.

La méthode repose aussi sur l'hypothèse que les populations au dénominateur des taux de fécondité ont été corrigées du sous-enregistrement et d'autres erreurs. C'est d'autant plus important que le numérateur et le dénominateur proviennent de sources différentes.

TRAVAIL PRÉPARATOIRE ET RECHERCHES PRÉLIMINAIRES

Avant d'entreprendre l'analyse des niveaux de fécondité en recourant à cette méthode, l'analyste doit vérifier la qualité des données au moins sur les points suivants :

- structure par âge et sexe de la population ;

- naissances déclarées au cours de l'année écoulée , et
- parités moyennes et nécessité d'une correction d'El-Badry.

PRÉCAUTIONS ET MISES EN GARDE

La méthode vise à estimer les niveaux récents de la fécondité et à mesurer la complétude de l'enregistrement des naissances, en vue d'ajuster les naissances enregistrées récemment pour tenir compte des omissions. Les effets des autres erreurs, comme les variations de la couverture du recensement au fil du temps, doivent donc avoir été corrigés avant de cumuler les taux de fécondité par âge pour comparer le résultat aux parités moyennes. Quand il apparaît que la complétude du dénombrement a varié, les recensements doivent être ajustés autant que possible avant de calculer les dénominateurs des taux. Il ne s'agit pas de corriger chaque recensement du sous-dénombrement absolu ; il suffit de s'assurer que la complétude relative des recensements antérieurs est la même que celle du plus récent.

APPLICATION DE LA MÉTHODE

Étape 1 : Calculer les parités moyennes déclarées

Calculer les parités moyennes, ${}_5P_x$ des femmes de chaque groupe d'âge ($x, x + 5$), pour $x = 15, 20, 45$, si ça n'a pas déjà été fait lors des recherches préliminaires.

Étape 2 : Estimer la population féminine par groupe d'âge au milieu de chaque année précédant le dernier recensement

La procédure exacte pour estimer la série des populations féminines par groupe d'âge en milieu d'année dépend des dates des recensements disponibles. Nous présentons donc la procédure seulement en termes généraux.

On suppose que les recensements couvrent, au moins à peu près, les années pour lesquelles on doit cumuler les taux de fécondité enregistrés, car il est préférable que la procédure ci-dessous recoure uniquement à des interpolations de population entre les dates de recensement, plutôt qu'à des extrapolations pour des dates antérieures ou postérieures à celles couvertes par les recensements.

Soit t_a la date de référence de chaque recensement a , tenant compte de la fraction d'année jusques et y compris la date exacte, ou la date de référence centrale du recensement. Par exemple, le nombre représentant la date d'un recensement réalisé la nuit du 9 au 10 octobre 2001 est calculé en additionnant les jours de l'année précédant la nuit du recensement :

$$2001 + \frac{(31 + 28 + 31 + 30 + 31 + 30 + 31 + 31 + 30 + 9)}{365}$$

$$= 2001 + \frac{282}{365} = 2001,773$$

Le taux de croissance $r(i,a)$ de chaque groupe d'âge i entre le recensement a et le recensement $a + 1$ est alors obtenu en divisant la différence entre les logarithmes naturels de la population féminine du groupe d'âge i au second et au premier recensement, $N(i,t_{a+1})$ et $N(i,t_a)$ respectivement, par la longueur de la période intercensitaire en années :

$$r(i,a) = \frac{\ln(N(i,t_{a+1})) - \ln(N(i,t_a))}{t_{a+1} - t_a} = \frac{\ln\left(\frac{N(i,t_{a+1})}{N(i,t_a)}\right)}{t_{a+1} - t_a} \quad (1)$$

S'il y a trois recensements couvrant la période étudiée, on doit calculer $r(i,1)$ pour la période entre le premier et le deuxième recensement, et $r(i,2)$ pour la période entre le deuxième et le troisième recensement.

La population au milieu de chaque année entre t_a et $t_a + 1$ pour chaque groupe d'âge peut alors être estimée à partir de la population initiale :

$$N(i,t) = N(i,t_a) \exp(r(i,a)(t + 0.5 - t_a)) \quad (2)$$

pour $t_a < t < t_a + 1$.

S'il est nécessaire d'extrapoler à des dates avant le premier recensement ou après le dernier, le taux de croissance de l'intervalle intercensitaire le plus proche doit être utilisé.

Les $N(i,t)$ étant des population estimées en milieu d'année au groupe d'âge i au temps t , ils représentent les dénominateurs approchés pour le calcul des taux de fécondité par âge cette année.

Étape 3 : Calculer les taux de fécondité par âge à partir des naissances enregistrées au cours des années antérieures au recensement

On doit disposer de taux de fécondité par âge pour un nombre d'années égal à l'âge le plus élevé pour lequel on souhaite utiliser les parités moins 15 ans. Par exemple, si on utilise des parités jusqu'à 35 ans (c'est-à-dire le groupe 30-34 ans), il est nécessaire de disposer de taux de fécondité remontant à 20 ans en arrière. Le taux pour le groupe d'âge

i et l'année civile t , $f(i,t)$, est égal à $f(i,t) = \frac{B(i,t)}{N(i,t)}$ où

$B(i,t)$ est le nombre de naissances enregistrées, étant survénues chez les femmes du groupe d'âge i l'année t .

Si on ne dispose pas des naissances par âge de la mère pour quelques-unes des vingt années nécessaires, l'application de la méthode n'en sera guère affectée si les taux des années manquantes peuvent être interpolés à partir de ceux des années voisines. Par exemple, si on ne dispose des taux de fécondité que pour les seize dernières années sur les vingt nécessaires pour une analyse de la complétude de l'enregistrement pour les femmes de moins de 35 ans, on peut utiliser les taux de l'année disponible la plus ancienne pour les quatre années précédentes sans grand risque d'introduire des erreurs importantes, car les valeurs imputées qui seront ensuite utilisées dans l'analyse concernent les femmes alors âgées de 15-19 ans, dont la fécondité est généralement relativement faible. En revanche, il faut éviter d'extrapoler les données à partir du passé afin d'imputer des valeurs pour les années les plus récentes car l'imputation concernerait davantage de groupes d'âge, y compris les âges de forte fécondité, où les variations au fil du temps risquent de se produire.

Étape 4 : Cumuler la fécondité enregistrée dans les différentes cohortes féminines afin d'estimer des descendance (parités équivalentes)

Les naissances étant classées par année civile, les taux de fécondité par âge sont obtenus selon le même classement. L'addition des taux jusqu'à la fin de chaque groupe d'âge donne donc des cumuls de fécondité qui correspondent à la fin des années civiles. Il est peu vraisemblable que le recensement dont sont tirées les parités moyennes ait sa date de référence au 31 décembre d'une année, mais comme les parités moyennes d'un groupe d'âge donné évoluent lentement même quand la fécondité change rapidement, on peut considérer que les parités au recensement se réfèrent la fin d'année la plus proche de la date de référence, et les taux enregistrés peuvent être cumulés jusqu'à cette même fin d'année. Ainsi, si le recensement a lieu avant le 1^{er} juillet, on cumulera les taux enregistrés jusqu'à la fin de l'année précédant la date du recensement, alors que si le recensement a lieu au second semestre, les taux enregistrés seront cumulés jusqu'à la fin de l'année du recensement. Il n'est pas justifié de développer des procédures plus précises et plus raffinées (en recourant par exemple à des extrapolations exponentielles), compte tenu des incertitudes et des erreurs qui affectent la méthode.

Pour estimer les descendance à partir de la fécondité des périodes, on applique un modèle relationnel de Gompertz aux taux de fécondité par âge obtenus à l'étape précédente. Les taux de fécondité par âge sont d'abord cumulés chaque année jusqu'à 20, 25, 30 et 35 ans. On prend ensuite les gompits des rapports entre les estimations adjacentes de la fécondité cumulée, comme on le fait dans la version standard du modèle relationnel de Gompertz. Ces rapports correspondent aux $z(x)$ de la formulation conventionnelle du modèle.

Dans un deuxième temps, les estimations des paramètres sont tirées des valeurs de $z(x)$ calculées pour chaque année. On procède de la même manière que lors de l'ajustement d'un modèle relationnel de Gompertz sur les seuls F -points dans la formulation conventionnelle du modèle. Nous supposons à nouveau une série de 20 ans de données d'état civil à 15-34 ans. Une régression de $z(x) - e(x)$ sur $g(x)$, où $e(x)$ et $g(x)$ sont tirés du schéma de fécondité standard, à 20, 25 et 30 ans, permet ensuite d'estimer $\alpha(t)$ et $\beta(t)$. $\beta(t)$ est la pente de la régression linéaire fondée sur les trois valeurs de $z(x)$ l'année t ; l'ordonnée à l'origine est égale à $\alpha(t) + (\beta(t) - 1)^2 \cdot c/2$, où c est une constante tirée du standard de fécondité utilisé.

Troisièmement, le modèle relationnel de Gompertz sert à estimer la fécondité par année d'âge pour chaque année civile. Les valeurs de $\alpha(t)$ et $\beta(t)$ calculées lors de la précédente étape servent à déterminer la forme du schéma de fécondité, alors que le niveau est tiré des cumuls de fécondité à chaque âge, 25, 30 et 35 ans. L'estimation de l'indice synthétique de fécondité associé à la fécondité cumulée jusqu'à l'âge x l'année t est donnée par

$$TF(x, t) = \frac{F(x, t)}{\exp(-\exp(-\alpha(t) - \beta(t)Y^s(x)))} \quad (3)$$

où $F(x, t)$ est la fécondité cumulée à l'âge x l'année t et $Y^s(x)$ est le gompit à l'âge x du schéma standard de fécondité modifié par Zaba. Il en résulte trois estimations de l'indice synthétique de fécondité dont la moyenne détermine l'indice synthétique de fécondité chaque année. Il est ensuite simple d'estimer la fécondité entre les âges x et $x + 1$ au temps t , $f(x, t)$:

$$f(x, t) = TF(x, t) \cdot \left(\frac{\exp(-\exp(-\alpha(t) - \beta(t)Y^s(x+1)))}{\exp(-\exp(-\alpha(t) - \beta(t)Y^s(x)))} \right) \quad (4)$$

Quatrièmement, les descendance sont obtenues en additionnant les taux de fécondité par année d'âge et année civile pour chaque cohorte quinquennale, et en divisant par cinq (largeur de l'intervalle d'âge). Si on note $E(i, s)$ ces descendance, où $i = 1$ correspond au groupe 15-19 ans, $i = 2$ au groupe 20-24 ans, etc., s désigne la dernière année pour laquelle les estimations sont calculées, alors

$$E(i) = \sum_{j=0}^{5i+3} \sum_{m=5i+9}^{5i+13} f(m - j, s - j) \quad (5)$$

Étape 5 : Estimer la complétude de l'enregistrement des naissances

On a construit la fécondité cumulée par cohorte à partir des naissances enregistrées, $E(i)$, calculée à l'étape précédente, de sorte qu'elle soit comparable à la parité déclarée par cohorte, $P(i)$, au dernier recensement. Le rapport $E(i)/P(i)$ donne donc une mesure de la complétude moyenne de l'enregistrement des naissances dans la cohorte i . Si la complétude de l'enregistrement est restée à peu près constante pendant une quinzaine d'années, les rapports E/P auront plus ou moins les mêmes valeurs pour toutes les cohortes, et une moyenne des rapports pour les groupes 20-24, 25-29 et 30-34 ans peut être utilisée comme estimation de la complétude de l'enregistrement à l'état civil au long de la période. Son inverse peut être utilisé comme facteur d'ajustement pour les taux de fécondité par âge calculés à l'étape 3.

L'interprétation de ces rapports E/P est prise entre deux forces contradictoires. D'un côté, en général, si la complétude de l'enregistrement des naissances s'est améliorée au fil du temps, les rapports E/P seront plus élevés pour les cohortes jeunes que pour les plus âgées. Dans un tel cas, on peut ajuster la distribution de fécondité la plus récente (fondée sur les naissances enregistrées) par $P(2)/E(2)$, le rapport reflétant le niveau de complétude le plus récent. (On évitera en général d'utiliser $P(1)/E(1)$ comme facteur d'ajustement, compte tenu de la difficulté intrinsèque à approximer $E(1)$ correctement.) Cependant, lorsque les rapports E/P indiquent une amélioration de la complétude de l'enregistrement au fil du temps, il n'existe pas de base évidente pour ajuster les distributions de fécondité des périodes anciennes.

Seconde force en jeu : les rapports E/P peuvent être moins élevés pour les femmes jeunes que pour les plus âgées du fait d'un enregistrement moins fiable et moins ponctuel des naissances par les jeunes mères. Si c'est le cas, le rapport E/P aux

jeunes âges donne une mauvaise représentation de la complétude des données d'état civil. Au contraire, les estimations de complétude fondées sur les déclarations des femmes de 25-29 et 30-34 ans sont relativement peu affectées par les fortes omissions aux jeunes âges et il peut être judicieux de déterminer la complétude en faisant la moyenne de ces deux rapports E/P .

INTERPRÉTATION ET DIAGNOSTICS

Une hypothèse importante dans cette méthode est que les dénominateurs utilisés dans le calcul des taux de fécondité par âge sont précis. Comme le numérateur et le dénominateur des taux viennent de deux sources différentes, la distribution par âge des taux de fécondité du moment sera déformée si des erreurs dans la déclaration des âges jouent différemment lors de l'enregistrement des naissances et lors du dénombrement des populations. Des variations erratiques des taux de croissance par âge peuvent suggérer des problèmes de déclaration des âges, mais les effets de ces mauvaises déclarations sur les estimations finales de complétude sont très difficiles à prédire. Les dénominateurs peuvent aussi être déformés par des variations dans la complétude du dénombrement d'un recensement à l'autre et une complétude différentielle du dénombrement selon l'âge peut affecter les résultats. Les variations dans la complétude du dénombrement peuvent aussi jouer sur les parités moyennes. Si par exemple les mères sont mieux dénombrées que les femmes sans enfants, les parités moyennes s'en trouvent surévaluées.

EXEMPLE

Nous utilisons les données tirées de trois recensements réalisés au Chili en avril 1970, avril 1982 et avril 1992 respectivement. Les tableaux originaux figurent dans les Annuaire démographiques de la Division statistique des NU ; ils portent sur le nombre de femmes par groupe d'âge aux différents recensements ; les naissances déclarées par âge de la mère et année ; le nombre moyen d'enfants déjà nés au dernier recensement. La mise en œuvre de la méthode fait l'objet d'un dossier Excel disponibles sur le site web.

Étape 1 : Calculer les parités moyennes déclarées à 15-19, ..., 30-34 ans au dernier recensement
Les parités moyennes par âge au recensement de 1992 (tableau 14.1) sont tirées de l'Annuaire démographique.

TABLEAU 14.1 Parités moyennes, Chili, recensement de 1992

| Groupe d'âge | 15-19 | 20-24 | 25-29 | 30-34 |
|----------------|-------|-------|-------|-------|
| Parité moyenne | 0,14 | 0,69 | 1,37 | 2,02 |

Aucun contrôle de ces données n'est possible, en l'absence des tableaux originaux.

Étape 2 : Estimer la population féminine par groupe d'âge au milieu de chaque année précédant le dernier recensement

Les nombres de femmes dénombrées par groupe d'âge à chacun des trois recensements (tableau 14.2) sont tirés de l'Annuaire démographique.

TABLEAU 14.2 Nombres de femmes par groupe d'âge dénombrées aux recensements de 1970, 1982 et 1992, Chili

| | | Groupe d'âge | | | |
|---|---------------------------------------|--------------|---------|---------|---------|
| a | Date du recensement (t _a) | 15-19 | 20-24 | 25-29 | 30-34 |
| 1 | 22-avril-70 | 466 736 | 398 383 | 324 130 | 267 312 |
| 2 | 22-avril-82 | 652 552 | 595 598 | 479 199 | 399 344 |
| 3 | 22-avril-92 | 600 563 | 608 933 | 623 305 | 576 710 |

Les dates de référence des recensements sont : 1970,304 ; 1982,304 et 1992,306. La date de référence est légèrement différente en 1992 car l'année était bissextile. Elle est calculée de la façon suivante : 1992 + (31 + 29 + 31 + 21)/366 = 1992,306, la date du recensement se référant à la nuit du 21-22 avril.

Le taux de croissance de chaque période intercensitaire (tableau 3) est calculé à partir de l'équation 1.

TABLEAU 14.3 Taux de croissance intercensitaires par âge, Chili, recensements de 1970, 1982 et 1992

| | | Groupe d'âge | | | |
|--|-------------------------|--------------|--------|--------|--------|
| | Période intercensitaire | 15-19 | 20-24 | 25-29 | 30-34 |
| | 1970-1982 | 0,0279 | 0,0335 | 0,0326 | 0,0335 |
| | 1982-1992 | - 0,0083 | 0,0022 | 0,0263 | 0,0367 |

Par exemple le taux de croissance à 30-34 ans entre les recensements de 1982 et 1992 est égal à

$$r(3,2) = \frac{\ln\left(\frac{576710}{399344}\right)}{1992,306 - 1982,304} = 0,0367.$$

Les taux de croissance augmentent avec l'âge et diminuent au fil du temps. Ceci suggère une baisse de la fécondité ayant débuté quelques décennies plus tôt, ce qui renforce la pertinence d'une méthode analytique qui ne postule pas une fécondité constante.

Sur la base de ces taux de croissance, on peut estimer la population des femmes par groupe d'âge au milieu de chaque année en utilisant l'équation 2. Les valeurs résultantes $N(i,t)$ figurent au tableau 14.4.

TABLEAU 14.4 Populations par groupe d'âge estimées en milieu d'année, Chili

| Année | Groupe d'âge | | | |
|-------|--------------|---------|---------|---------|
| | 15-19 | 20-24 | 25-29 | 30-34 |
| 1972 | 496 255 | 428 806 | 348 169 | 287 686 |
| 1973 | 510 309 | 443 420 | 359 700 | 297 472 |
| 1974 | 524 761 | 458 532 | 371 612 | 307 591 |
| 1975 | 539 623 | 474 159 | 383 919 | 318 054 |
| 1976 | 554 906 | 490 318 | 396 633 | 328 873 |
| 1977 | 570 621 | 507 029 | 409 769 | 340 060 |
| 1978 | 586 781 | 524 308 | 423 340 | 351 628 |
| 1979 | 603 400 | 542 177 | 437 359 | 363 589 |
| 1980 | 620 488 | 560 655 | 451 844 | 375 957 |
| 1981 | 638 061 | 579 762 | 466 808 | 388 746 |
| 1982 | 651 492 | 599 521 | 482 267 | 401 969 |
| 1983 | 646 106 | 597 177 | 494 503 | 417 283 |
| 1984 | 640 765 | 598 500 | 507 674 | 432 902 |
| 1985 | 635 469 | 599 827 | 521 196 | 449 104 |
| 1986 | 630 215 | 601 156 | 535 078 | 465 913 |
| 1987 | 625 006 | 602 489 | 549 331 | 483 352 |
| 1988 | 619 839 | 603 824 | 563 962 | 501 443 |
| 1989 | 614 715 | 605 162 | 578 984 | 520 211 |
| 1990 | 609 634 | 606 503 | 594 405 | 539 681 |
| 1991 | 604 595 | 607 847 | 610 238 | 559 880 |

Par exemple, la population des femmes de 15-19 ans, mi-1990, est égale à

$$N(1,1990) = 652552 \exp(-0,0083(1990,5 - 1982,3)) \\ = 609634.$$

Étape 3 : Calculer les taux de fécondité par âge à partir des naissances enregistrées

au cours des années antérieures au recensement

Le nombre de naissances déclarées par groupe d'âge et année (tableau 14.5) est tiré de l'Annuaire démographique.

TABLEAU 14.5 Naissances déclarées par groupe d'âge de la mère et année, Chili

| Année | Groupe d'âge | | | |
|-------|--------------|--------|--------|--------|
| | 15-19 | 20-24 | 25-29 | 30-34 |
| 1972 | 39 839 | 80 430 | 64 624 | 38 937 |
| 1973 | 40 241 | 82 108 | 63 949 | 38 499 |
| 1974 | 39 884 | 79 316 | 63 477 | 37 880 |
| 1975 | 39 086 | 75 519 | 59 365 | 35 863 |
| 1976 | 37 658 | 73 889 | 57 171 | 34 129 |
| 1977 | 36 104 | 71 445 | 53 467 | 32 190 |
| 1978 | 37 138 | 73 224 | 53 725 | 31 832 |
| 1979 | 36 833 | 75 905 | 55 361 | 32 537 |
| 1980 | 38 562 | 79 724 | 59 771 | 33 769 |
| 1981 | 40 252 | 86 037 | 64 849 | 36 494 |
| 1982 | 39 298 | 86 061 | 68 029 | 38 406 |
| 1983 | 36 077 | 81 213 | 65 236 | 37 506 |
| 1984 | 37 571 | 83 960 | 67 266 | 39 105 |
| 1985 | 34 946 | 80 735 | 69 180 | 39 828 |
| 1986 | 35 925 | 83 434 | 72 876 | 42 605 |
| 1987 | 35 633 | 84 674 | 75 416 | 45 037 |
| 1988 | 37 354 | 87 484 | 80 527 | 48 290 |
| 1989 | 39 095 | 86 990 | 82 919 | 50 875 |
| 1990 | 39 543 | 85 292 | 84 336 | 52 942 |
| 1991 | 38 324 | 79 406 | 81 907 | 53 425 |

Les taux de fécondité par âge sont obtenus en divisant les naissances (tableau 14.5) par la population estimée en milieu d'année pour chaque groupe d'âge et année (tableau 14.4). Les résultats figurent au tableau 14.6.

Étape 4 : Cumuler la fécondité enregistrée

dans les différentes cohortes féminines

afin d'estimer des descendance (parités équivalentes)

Le calcul des descendance est fait selon la procédure suivante. Les taux de fécondité par âge du tableau 14.6 sont cumulés jusqu'à la limite supérieure de chaque groupe d'âge. Par exemple, la fécondité cumulée jusqu'à 25 ans en 1972 est égale à $5(0,0803 + 0,1876) = 1,3392$. La fécondité cumulée jusqu'à 30 ans en 1972 est égale à $5(0,0803 + 0,1876 + 0,1856) = 2,2673$. Les taux cumulés figurent au tableau 14.7.

TABLEAU 14.6 Taux de fécondité par groupe d'âge et année, Chili

| Année | Groupe d'âge | | | |
|-------|--------------|--------|--------|--------|
| | 15-19 | 20-24 | 25-29 | 30-34 |
| 1972 | 0,0803 | 0,1876 | 0,1856 | 0,1353 |
| 1973 | 0,0789 | 0,1852 | 0,1778 | 0,1294 |
| 1974 | 0,0760 | 0,1730 | 0,1708 | 0,1232 |
| 1975 | 0,0724 | 0,1593 | 0,1546 | 0,1128 |
| 1976 | 0,0679 | 0,1507 | 0,1441 | 0,1038 |
| 1977 | 0,0633 | 0,1409 | 0,1305 | 0,0947 |
| 1978 | 0,0633 | 0,1397 | 0,1269 | 0,0905 |
| 1979 | 0,0610 | 0,1400 | 0,1266 | 0,0895 |
| 1980 | 0,0621 | 0,1422 | 0,1323 | 0,0898 |
| 1981 | 0,0631 | 0,1484 | 0,1389 | 0,0939 |
| 1982 | 0,0603 | 0,1435 | 0,1411 | 0,0955 |
| 1983 | 0,0558 | 0,1360 | 0,1319 | 0,0899 |
| 1984 | 0,0586 | 0,1403 | 0,1325 | 0,0903 |
| 1985 | 0,0550 | 0,1346 | 0,1327 | 0,0887 |
| 1986 | 0,0570 | 0,1388 | 0,1362 | 0,0914 |
| 1987 | 0,0570 | 0,1405 | 0,1373 | 0,0932 |
| 1988 | 0,0603 | 0,1449 | 0,1428 | 0,0963 |
| 1989 | 0,0636 | 0,1437 | 0,1432 | 0,0978 |
| 1990 | 0,0649 | 0,1406 | 0,1419 | 0,0981 |
| 1991 | 0,0634 | 0,1306 | 0,1342 | 0,0954 |

On calcule, pour chaque année, le rapport de la fécondité cumulée dans un groupe d'âge à celle dans le groupe d'âge immédiatement supérieur. Dans l'exemple du paragraphe précédent, le rapport de la fécondité cumulée à 25 ans à celle cumulée à 30 ans est égal à $1,3392/2,2673 = 0,5907$. Les rapports figurent dans les trois premières colonnes du tableau 14.8.

On applique ensuite une transformation gompit (log-log négatif) aux rapports pour obtenir une valeur de $z(x)$ pour chaque âge 25, 30 et 35 ans, chaque année. Sur le même exemple, la valeur de $z(25)$ en 1972 est $-\ln(\ln(0,5907)) = 0,6415$. Les gompits sont dans les trois dernières colonnes du tableau 8.

Les valeurs de $e(x)$ et $g(x)$ sont déterminées sans décalage des âges puisque les données de fécondité proviennent d'un système d'enregistrement à l'état civil et se réfèrent donc à l'âge de la mère au moment de la naissance de l'enfant. Les valeurs sont tirées de la version du standard de fécondité de Booth modifiée par Zaba, le seul standard actuellement disponible pour les femmes qui ait été validé. (Le standard et la procédure de détermination de $e(x)$ et $g(x)$ sont décrits en détail dans la présentation de la méthode relationnelle de Gompertz). Les valeurs de $e(x)$ et $g(x)$ pour les âges néces-

TABLEAU 14.7 Fécondité cumulée jusqu'à l'âge x, par année, Chili

| Année | Groupe d'âge | | | |
|-------|--------------|--------|--------|--------|
| | 20 | 25 | 30 | 35 |
| 1972 | 0,4014 | 1,3392 | 2,2673 | 2,9440 |
| 1973 | 0,3943 | 1,3201 | 2,2091 | 2,8562 |
| 1974 | 0,3800 | 1,2449 | 2,0990 | 2,7147 |
| 1975 | 0,3622 | 1,1585 | 1,9317 | 2,4954 |
| 1976 | 0,3393 | 1,0928 | 1,8135 | 2,3324 |
| 1977 | 0,3164 | 1,0209 | 1,6733 | 2,1466 |
| 1978 | 0,3165 | 1,0147 | 1,6493 | 2,1019 |
| 1979 | 0,3052 | 1,0052 | 1,6381 | 2,0856 |
| 1980 | 0,3107 | 1,0217 | 1,6831 | 2,1322 |
| 1981 | 0,3154 | 1,0574 | 1,7520 | 2,2214 |
| 1982 | 0,3016 | 1,0193 | 1,7247 | 2,2024 |
| 1983 | 0,2792 | 0,9592 | 1,6188 | 2,0682 |
| 1984 | 0,2932 | 0,9946 | 1,6571 | 2,1087 |
| 1985 | 0,2750 | 0,9479 | 1,6116 | 2,0550 |
| 1986 | 0,2850 | 0,9790 | 1,6600 | 2,1172 |
| 1987 | 0,2851 | 0,9878 | 1,6742 | 2,1401 |
| 1988 | 0,3013 | 1,0257 | 1,7397 | 2,2212 |
| 1989 | 0,3180 | 1,0367 | 1,7528 | 2,2418 |
| 1990 | 0,3243 | 1,0275 | 1,7369 | 2,2274 |
| 1991 | 0,3169 | 0,9701 | 1,6412 | 2,1183 |

TABLEAU 14.8 Rapports des fécondités cumulées à l'âge x et leurs gompits, par année, Chili

| Année | Rapports | | | Gompits $z(x)$ | | |
|-------|----------|--------|--------|----------------|--------|--------|
| | 20 | 25 | 30 | 20 | 25 | 30 |
| 1972 | 0,2997 | 0,5907 | 0,7701 | -0,1864 | 0,6415 | 1,3425 |
| 1973 | 0,2987 | 0,5976 | 0,7734 | -0,1893 | 0,6639 | 1,3590 |
| 1974 | 0,3053 | 0,5931 | 0,7732 | -0,1711 | 0,6493 | 1,3577 |
| 1975 | 0,3126 | 0,5997 | 0,7741 | -0,1508 | 0,6709 | 1,3622 |
| 1976 | 0,3105 | 0,6026 | 0,7775 | -0,1566 | 0,6802 | 1,3798 |
| 1977 | 0,3099 | 0,6101 | 0,7795 | -0,1583 | 0,7050 | 1,3900 |
| 1978 | 0,3119 | 0,6153 | 0,7847 | -0,1529 | 0,7222 | 1,4167 |
| 1979 | 0,3036 | 0,6136 | 0,7855 | -0,1756 | 0,7167 | 1,4209 |
| 1980 | 0,3041 | 0,6070 | 0,7894 | -0,1742 | 0,6948 | 1,4417 |
| 1981 | 0,2983 | 0,6035 | 0,7887 | -0,1904 | 0,6833 | 1,4381 |
| 1982 | 0,2959 | 0,5910 | 0,7831 | -0,1971 | 0,6427 | 1,4085 |
| 1983 | 0,2911 | 0,5925 | 0,7827 | -0,2104 | 0,6475 | 1,4065 |
| 1984 | 0,2948 | 0,6002 | 0,7858 | -0,2001 | 0,6724 | 1,4228 |
| 1985 | 0,2901 | 0,5882 | 0,7842 | -0,2132 | 0,6336 | 1,4145 |
| 1986 | 0,2911 | 0,5898 | 0,7840 | -0,2102 | 0,6386 | 1,4135 |
| 1987 | 0,2886 | 0,5900 | 0,7823 | -0,2173 | 0,6393 | 1,4044 |
| 1988 | 0,2938 | 0,5896 | 0,7832 | -0,2029 | 0,6381 | 1,4092 |
| 1989 | 0,3067 | 0,5915 | 0,7819 | -0,1670 | 0,6441 | 1,4022 |
| 1990 | 0,3156 | 0,5916 | 0,7798 | -0,1425 | 0,6444 | 1,3914 |
| 1991 | 0,3267 | 0,5911 | 0,7748 | -0,1122 | 0,6429 | 1,3658 |

saies à l'ajustement d'un modèle relationnel de Gompertz aux données de fécondité observées figurent au tableau 14.9.

TABEAU 14.9 Valeurs de $e(x)$ et $g(x)$ à partir du standard modifié par Zaba (pas de décalage des âges), certains âges

| Âge x | $e(x)$ | $g(x)$ |
|---------|--------|----------|
| 20 | 1,3539 | - 1,3753 |
| 25 | 1,4127 | - 0,6748 |
| 30 | 1,2750 | 0,0393 |

On peut donc déterminer les valeurs de $z(x) - e(x)$ pour les trois âges, et les régresser sur les valeurs de $g(x)$ aux mêmes âges. Dans le modèle relationnel de Gompertz, β est la pente de l'équation de régression, et α est estimé à partir de

$$\alpha = \text{intercept} - (\beta - 1)^2 \cdot \frac{c}{2}$$

où c est une constante (= 0,95739) tirée du standard de fécondité modifié par Zaba. Il en résulte les valeurs de α et β qui figurent aux deux premières colonnes du tableau 14.10.

L'équation 3 permet d'obtenir les estimations de l'indice synthétique de fécondité associées à la fécondité cumulée jusqu'à l'âge x . Ainsi, la fécondité cumulée à 25 ans en 1972 implique un indice synthétique de

$$\frac{1,3392}{\exp\left(-\exp\left(-0,0049 - 1,1367Y^s(x)\right)\right)} = 3,5205.$$

La moyenne des trois estimations donne l'estimation finale de l'indice synthétique de chaque année. En combinant cette dernière avec les estimations de α and β , on peut déterminer une série de taux de fécondité par année d'âge pour chaque année civile grâce une fois de plus au modèle relationnel de Gompertz.

Les taux de fécondité par année d'âge pour chaque année civile sont obtenus par l'équation 4. Il en résulte une large matrice (20 années par 25 âges), dont nous ne reproduisons qu'un extrait au tableau 14.11.

Par exemple, le taux de fécondité entre 16 et 17 ans en 1990 est égal à

$$f(16,1990)$$

$$= TF(16,1990) \cdot \left(\frac{\exp\left(-\exp\left(-\alpha(1990) - \beta(1990)Y^s(17)\right)\right)}{-\exp\left(-\exp\left(-\alpha(1990) - \beta(1990)Y^s(16)\right)\right)} \right)$$

TABEAU 14.10 Paramètres Alpha et Beta d'un modèle relationnel de Gompertz ajusté à 20, 25 et 30 ans, Chili

| Année | Indice synthétique fondé sur le cumulant à l'âge | | | | | |
|-------|--|--------|--------|--------|--------|---------|
| | Alpha | Beta | 25 | 30 | 35 | Moyenne |
| 1972 | 0,0049 | 1,1367 | 3,5205 | 3,5525 | 3,5434 | 3,5388 |
| 1973 | 0,0242 | 1,1504 | 3,4057 | 3,4173 | 3,4130 | 3,4120 |
| 1974 | 0,0177 | 1,1367 | 3,2326 | 3,2701 | 3,2598 | 3,2542 |
| 1975 | 0,0270 | 1,1255 | 2,9825 | 3,0074 | 3,0004 | 2,9968 |
| 1976 | 0,0430 | 1,1420 | 2,7704 | 2,7897 | 2,7840 | 2,7814 |
| 1977 | 0,0585 | 1,1503 | 2,5509 | 2,5508 | 2,5498 | 2,5505 |
| 1978 | 0,0828 | 1,1653 | 2,4796 | 2,4782 | 2,4773 | 2,4784 |
| 1979 | 0,0844 | 1,1843 | 2,4518 | 2,4466 | 2,4462 | 2,4482 |
| 1980 | 0,0912 | 1,1982 | 2,4761 | 2,4973 | 2,4902 | 2,4879 |
| 1981 | 0,0850 | 1,2071 | 2,5762 | 2,5995 | 2,5916 | 2,5891 |
| 1982 | 0,0516 | 1,1910 | 2,5607 | 2,6054 | 2,5928 | 2,5863 |
| 1983 | 0,0520 | 1,1990 | 2,4083 | 2,4396 | 2,4300 | 2,4260 |
| 1984 | 0,0712 | 1,2032 | 2,4537 | 2,4750 | 2,4677 | 2,4654 |
| 1985 | 0,0527 | 1,2067 | 2,3781 | 2,4227 | 2,4099 | 2,4035 |
| 1986 | 0,0537 | 1,2038 | 2,4538 | 2,4963 | 2,4840 | 2,4781 |
| 1987 | 0,0479 | 1,2024 | 2,4892 | 2,5248 | 2,5141 | 2,5094 |
| 1988 | 0,0506 | 1,1956 | 2,5788 | 2,6257 | 2,6124 | 2,6057 |
| 1989 | 0,0472 | 1,1654 | 2,6166 | 2,6728 | 2,6579 | 2,6491 |
| 1990 | 0,0388 | 1,1404 | 2,6152 | 2,6780 | 2,6617 | 2,6516 |
| 1991 | 0,0179 | 1,1009 | 2,5207 | 2,5856 | 2,5689 | 2,5584 |

$$= 2,6516 \cdot \left(\frac{\exp\left(-\exp\left(-0,0388 - 1,1404Y^s(17)\right)\right)}{-\exp\left(-\exp\left(-0,0388 - 1,1404Y^s(16)\right)\right)} \right) = 0,0348$$

Enfin, les valeurs de $E(i)$ sont obtenues en appliquant l'équation 5 aux taux de fécondité du tableau 14.11. $E(1)$, descendance à 15-19 ans, est donc obtenue ainsi

$$\begin{aligned} E(1) &= \sum_{j=0}^8 \sum_{m=14}^{18} f(m-j, s-j) \\ &= f(14,1991) + f(15,1991) + \dots + f(18,1991) + \\ &\quad f(13,1990) + \dots f(17,1990) + \dots + f(6,1983) + \\ &\quad \dots f(10,1983) \\ &= 0,07394. \end{aligned}$$

La fécondité aux âges avant 10 ans est certainement nulle, mais nous l'avons incluse dans la formule pour faciliter la présentation.

TABLEAU 14.11 Taux de fécondité par année d'âge et année civile, Chili

| Année | Âge | | | | | | | | |
|-------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
| 1982 | 0,0000 | 0,0000 | 0,0000 | 0,0001 | 0,0012 | 0,0081 | 0,0290 | 0,0574 | 0,0883 |
| 1983 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0010 | 0,0072 | 0,0263 | 0,0530 | 0,0822 |
| 1984 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0011 | 0,0078 | 0,0282 | 0,0562 | 0,0864 |
| 1985 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0009 | 0,0067 | 0,0253 | 0,0517 | 0,0809 |
| 1986 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0010 | 0,0071 | 0,0265 | 0,0538 | 0,0839 |
| 1987 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0010 | 0,0071 | 0,0264 | 0,0538 | 0,0841 |
| 1988 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0011 | 0,0079 | 0,0285 | 0,0571 | 0,0884 |
| 1989 | 0,0000 | 0,0000 | 0,0000 | 0,0001 | 0,0016 | 0,0099 | 0,0325 | 0,0615 | 0,0919 |
| 1990 | 0,0000 | 0,0000 | 0,0000 | 0,0001 | 0,0021 | 0,0113 | 0,0348 | 0,0634 | 0,0927 |
| 1991 | 0,0000 | 0,0000 | 0,0000 | 0,0002 | 0,0027 | 0,0129 | 0,0363 | 0,0627 | 0,0890 |

Étape 5 : Estimer la complétude de l'enregistrement des naissances

Pour chaque cohorte (définie par l'âge à la fin de l'année terminale), la complétude de l'enregistrement des naissances est estimée comme le rapport entre la descendance, $E(i)$, calculée à partir des naissances enregistrées, et la parité moyenne déclarée $P(i)$, tirée du dernier recensement (cf. tableau 14.1). Les résultats figurent au tableau 14.12.

La complétude estimée pour la cohorte âgée de 15-19 ans fin 1991 est égale à $0,07/0,14 = 0,5281$, soit 52 %. Les estimations de la complétude de l'enregistrement des naissances pour les femmes âgées de 25-29 et 30-34 ans sont à la fois plus élevées et plus cohérentes entre elles, suggérant un niveau moyen de complétude d'environ 94 %. L'estimation pour les femmes de 15-19 ans est beaucoup plus basse et celle pour les femmes de 20-24 ans suggère un ordre de grandeur de 80 %. L'enregistrement des naissances est vraisemblablement moins complet parmi les jeunes mères et la complétude cumulée des mères âgées de 20-24 ans est elle aussi plus basse du fait que les naissances qu'elles ont eues comme adolescentes représentent une fraction importante de leur fécondité.

Les estimations de complétude fondées sur les déclarations de femmes de 25-29 et 30-34 ans sont relativement peu affectées par les fortes omissions aux jeunes âges, de sorte que la meilleure estimation de la complétude de l'enregistrement des naissances est sans doute dans ce cas la moyenne entre 0,9382 et 0,9515. L'estimation finale est donc 0,9449.

On peut donc avoir une estimation améliorée de la fécondité en 1991 en multipliant les taux de fécondité par âge de cette année par un facteur de $1/0,9449$, soit un surcroît d'environ 5,83 %. Notez cependant que la distribution ajustée de la fécondité n'est sans doute pas un bon indicateur du profil par âge de la fécondité du fait d'une omission apparente des naissances relativement plus forte par les jeunes femmes.

TABLEAU 14.12 Estimations de la complétude de l'enregistrement des naissances, Chili 1991

| Groupe d'âge | | | | | |
|---|------|--------|--------|--------|--------|
| | | 15-19 | 20-24 | 25-29 | 30-34 |
| Parités moyennes au dernier recensement | 1991 | 0,14 | 0,69 | 1,37 | 2,02 |
| (fin d'année) | | | | | |
| Descendances (fin d'année) | 1991 | 0,07 | 0,57 | 1,29 | 1,92 |
| Complétude | | 0,5281 | 0,8207 | 0,9382 | 0,9515 |
| Complétude moyenne | | | | | 0,9449 |

Les résultats du tableau 14.12 ne font pas apparaître d'amélioration de la complétude de l'enregistrement des naissances au fil du temps ; aussi le facteur d'ajustement de 1,0583 peut-il être également appliqué aux taux de fécondité par âge des années antérieures à 1991. Toutefois, comme l'analyse a été arrêtée à 30-34 ans, les contributions

des taux de fécondité enregistrés avant 1982 à la fécondité cumulée par cohorte sont faibles. Le facteur d'ajustement estimé ne peut donc pas être appliqué valablement aux taux de fécondité enregistrés avant cette date. Quand il existe des signes d'amélioration de la qualité de l'enregistrement, par exemple des estimations de complétude qui diminuent avec l'âge – ce qui n'est pas le cas au Chili – il n'est pas judicieux d'ajuster les naissances enregistrées une année donnée par un facteur correctif tiré d'une moyenne calculée sur une longue période.

Aucune de ces considérations n'explique cependant de manière adéquate les faibles estimations de complétude pour les cohortes âgées de 15-19 ans et, dans une moindre mesure, celles âgées de 20-24 ans. Certes la procédure de répartition par année d'âge de la fécondité du moment n'est pas parfaite et elle risque d'être particulièrement imprécise à 15-19 ans, mais l'éventuelle imprécision de la méthode ne peut pas expliquer l'ampleur de la différence observée. En outre, les parités moyennes ayant été calculées sans aucune correction pour non réponse, elles risquent plutôt d'être trop basses que trop hautes. Il semble donc bien que l'enregistrement des naissances au Chili soit moins complet pour les jeunes mères que pour les femmes plus âgées.

AUTRES VARIANTES

Le *Manuel X* (Division de la Population des NU 1984 : 56ss) présente une extension de la méthode, qui utilise les parités tirées de deux enquêtes espacées de cinq ou dix ans, combinées aux données d'état civil couvrant les naissances sur une période de 20 ans.

À la différence de ce qu'on vient de voir en ajustant la fécondité du moment sur la base de la fécondité cumulée par cohorte depuis le début de la période féconde, il n'est plus nécessaire de supposer que la complétude de l'enregistrement est constante à la fois par âge de la mère et par période. Toutefois, les accroissements de parité sont très sensibles aux variations dans la qualité de la déclaration du nombre d'enfants déjà nés, de sorte que les estimations de complétude de l'état civil obtenues par la méthode décrite aux pages 56 et suivantes du *Manuel X* sont également très sensibles à de telles variations, lesquelles sont généralement fortes parmi les femmes âgées.

Les différences principales entre la procédure décrite aux pages 56ss du *Manuel X* et celle présentée plus haut sont que tous les calculs dans la version variante sont limités à la

période entre les deux enquêtes et qu'on y prend en compte toutes les cohortes en âge fécond à la seconde enquête. En procédant ainsi, la méthode repose sur les accroissements de fécondité selon les enregistrements à l'état civil pendant la période entre les deux enquêtes, plutôt que sur la fécondité cumulée par les cohortes de femmes depuis le début de leur vie féconde. Il en va de même pour les accroissements de parité dans les cohortes entre les deux enquêtes. La mesure de la complétude est donc fondée sur le rapport entre la fécondité cumulée et les accroissements de parité entre les deux enquêtes.

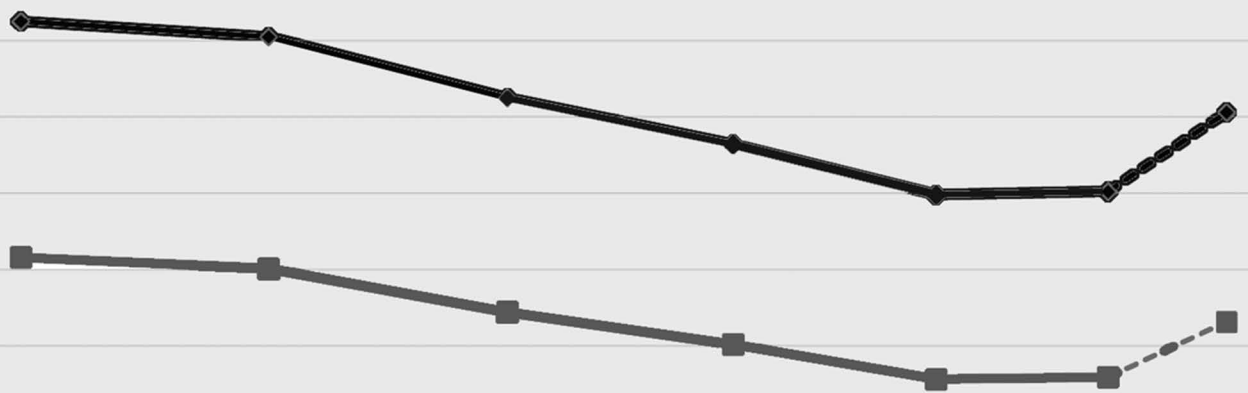
AUTRES LECTURES ET RÉFÉRENCES

La méthode a d'abord été présentée dans le *Manuel X* (Division de la Population des NU 1984 : 46-56). Elle a été peu appliquée car elle nécessite une longue série de données d'état civil. Mais elle pourrait gagner en importance à l'avenir, comme contrôle de qualité des systèmes d'enregistrement que des pays en développement essaient actuellement de mettre en place ou d'améliorer.

La seule modification importante dans la mise en œuvre de la méthode porte sur la procédure de répartition de la fécondité dans des groupes d'âge non conventionnels, qui passait originellement par des coefficients fondés sur la distribution de fécondité de Brass – une méthode longue, fastidieuse et source d'erreurs – et qui est remplacée ici par une procédure plus simple, utilisant le modèle relationnel de Gompertz. En effet, ici, nous ajustons un modèle relationnel de Gompertz séparé aux taux de fécondité par âge observés chaque année dans les données d'état civil. Les valeurs de paramètres alpha et bêta ainsi déterminées permettent de produire des estimations de l'indice synthétique de fécondité à partir de la fécondité observée jusqu'à 25, 30 et 35 ans. La moyenne de ces trois estimations est prise comme estimation de l'indice synthétique de chaque année et permet le calcul des taux de fécondité par année d'âge pour chaque année civile. À partir de là, le calcul de la fécondité cumulée par cohorte est aisé, les taux de chaque année devant simplement être additionnés.

Division de la Population des Nations Unies. 1984. *Manuel X. Techniques indirectes d'estimation démographique*. New York : Nations Unies, Département des affaires économiques et sociales internationales, ST/ESA/SER.A/81. http://unstats.un.org/unsd/demographic/standmeth/handbooks/Manuel_X-en.pdf

La mortalité des jeunes enfants



Chapitre 15 INTRODUCTION À L'ANALYSE DE LA MORTALITÉ DES JEUNES ENFANTS

Kenneth Hill

Le taux de mortalité des jeunes enfants a longtemps été considéré comme un indicateur du niveau de développement général d'un pays. Au cours des deux dernières décennies, la communauté internationale a régulièrement fixé des objectifs de réduction de la mortalité des jeunes enfants. Le Sommet mondial pour les enfants en 1990 s'était donné pour but une réduction de la probabilité de décéder entre la naissance et le 5^e anniversaire (${}_5q_0$) d'un tiers, ou à un niveau inférieur à 70 pour mille, entre 1990 et 2000. La Conférence Internationale pour la Population et le Développement de 1994 avait confirmé cette cible, mais avait aussi instauré un objectif à plus long terme : amener le niveau du ${}_5q_0$ en dessous de 45 pour mille d'ici à 2015. Le Sommet du millénaire en 2000 a ensuite adopté une série de huit objectifs de développement, dont le quatrième incluait une réduction de ${}_5q_0$ des deux tiers entre 1990 et 2015. Le rapport de la Commission de l'information et de la responsabilisation en matière de la santé de la femme et de l'enfant, établi par le Secrétaire général des Nations Unies, a récemment réaffirmé l'importance d'une information fréquente sur l'évolution de ${}_5q_0$. Il est donc clair que la mesure de la mortalité infantile et juvénile est une priorité très forte à la fois au niveau national et international. Ce chapitre présente les grandes options de cette mesure, les détails des diverses méthodes étant présentés dans les chapitres suivants.

DONNÉES POUR L'ESTIMATION DE LA MORTALITÉ AVANT 5 ANS

Les données sur la mortalité des jeunes enfants tirées de sources autres que les systèmes d'enregistrement à l'état civil (et souvent même ces dernières) émanent presque exclusivement des mères. Les données recueillies concernent soit l'ensemble des enfants (généralement nés vivants) soit un enfant ou un groupe d'enfants spécifiques. Elles sont recueillies

avec une quantité de détails variable, compte tenu de la difficulté de se souvenir de détails relatifs à des naissances anciennes. L'absence d'informations détaillées lors de la collecte des données peut être ultérieurement compensée lors de l'analyse par des hypothèses, en particulier sur les répartitions par âge passées de la mortalité des jeunes enfants et de la fécondité, mais c'est au prix d'une imprécision accrue.

Le fait que les données soient recueillies auprès des mères est important, car il pose la question du biais de sélection. Si tous les enfants (vivants ou décédés) dans une population observée sur une longue période avaient leur mère encore vivante et encore présente dans la population, chaque enfant aurait une même probabilité d'être déclaré et il n'y aurait pas de biais. Mais en réalité, les mères sont elles aussi soumises aux risques de mortalité et de migration ; en conséquence, à une date ultérieure, il ne sera plus vrai que tous les enfants dans la population visée (et seulement les enfants dans la population visée) seront déclarés. Ce biais est susceptible d'être le plus large dans des populations fortement affectées par le VIH/Sida. Dans ces populations, en l'absence de recours massif aux antirétroviraux, la corrélation est forte entre les risques de mortalité des mères et de leurs enfants, de sorte que la mortalité des jeunes enfants mesurée à partir de la déclaration des mères survivantes sous-estime la mortalité réelle. Cette question est traitée en détail dans la section consacrée aux effets du VIH sur les méthodes d'estimation de la mortalité des jeunes enfants. Plus l'intervalle est long entre la naissance d'un enfant et la collecte des données, plus est important le risque de tels biais.

Nous faisons une distinction plus bas entre, d'une part, les données recueillies lors d'interviews au sein des ménages (y compris les collectes de données visant la couverture exhaustive d'une population, comme un recensement, et celles qui recueillent les données auprès d'un échantillon représentatif

de la population, comme les Enquêtes Démographiques et de Santé), et, d'autre part, les données recueillies dans des établissements de santé, comme les maternités.

DONNÉES RECUEILLIES LORS D'INTERVIEWS AU SEIN DES MÉNAGES

Les enquêtes recueillant des données par interview auprès des ménages peuvent aller de la collecte exhaustive (recensement de population) à des échantillons restreints et ciblés. La quantité et la complexité des données susceptibles d'être recueillies par interview varient en fonction de la taille de l'enquête : en général, dans les grandes enquêtes les questionnaires sont courts et la formation des enquêteurs est limitée, alors que dans les petites enquêtes (sous réserve des contraintes financières) les questionnaires peuvent être beaucoup plus longs et les enquêteurs peuvent être formés plus soigneusement pour recueillir une information beaucoup plus détaillée. Une information plus détaillée peut permettre l'analyse d'un éventail plus large d'indicateurs, alors que des données simples tirées d'un échantillon plus large, voire de l'ensemble de la population, permettent une désagrégation beaucoup plus fine en fonction de facteurs géographiques ou socio-économiques. Les deux sources peuvent ainsi être considérées comme complémentaires plutôt que concurrentes. Dans la présente section, nous présentons les types de données des plus complexes aux plus simples. Nous décrivons ensuite les processus d'évaluation de la qualité des données et d'analyse.

L'histoire génésique complète (*en anglais, full birth history, FBH*)

Une histoire génésique complète recueille auprès des mères enquêtées des informations sur chaque naissance vivante qu'a eue la femme au cours de sa vie. Dans la plupart des cas, l'information est recueillie auprès des femmes d'âge fécond (normalement 15-49 ans). Les questions peuvent être cependant réservées aux femmes mariées ou l'ayant été, quand il est considéré comme inopportun de les poser aux femmes célibataires. L'information minimale pour chaque enfant est son nom, sa date de naissance (souvent le mois et l'année), son état de survie (vivant ou décédé), et s'il est décédé, son âge au décès. Dans les Enquêtes Démographiques et de Santé (EDS), l'âge au décès est recueilli en jours si l'enfant est décédé avant 28 jours, en mois s'il est décédé entre 28 jours et 24 mois, et en années par la suite. Pour l'analyste, la mesure en années est peu efficace car elle ne permet pas de localiser les décès précisément dans le temps.

Dans les EDS, les données sont recueillies par rang de naissance, du premier au dernier né, avec des relances entre chaque naissance pour s'assurer qu'aucun enfant n'est omis. Des informations complémentaires sont souvent recueillies pour savoir si l'enfant est issu d'une naissance multiple, connaître le sexe de l'enfant, savoir si l'enfant vit actuellement avec sa mère et, dans ce cas, quelle est sa position dans la liste des membres du ménage. Les données recueillies permettent ainsi de localiser chaque naissance et, si l'enfant est décédé, de situer son décès dans le temps ; il est ainsi possible de calculer les taux classiques rapportant les événements à la population exposée au risque et d'utiliser les méthodes analytiques standards des tables de mortalité pour estimer les risques de mortalité pour différentes catégories d'âge dans l'enfance.

Dans les EDS, l'histoire génésique complète est précédée par une histoire génésique résumée (*en anglais, summary birth history, SBH*), où chaque femme est interrogée sur le nombre de ses fils et filles vivant avec elle, le nombre de ses fils et filles vivant ailleurs et le nombre de ses fils et filles décédés. L'histoire génésique complète est ensuite recueillie et confrontée sur place avec l'histoire génésique résumée pour en assurer la cohérence. Cette confrontation étant faite sur le terrain, les EDS ne font pas apparaître la fréquence, l'ampleur et la direction des écarts entre l'histoire génésique complète et l'histoire génésique résumée d'origine.

Certaines enquêtes recueillent un historique complet des grossesses, en ajoutant toutes les grossesses qui ne se sont pas terminées par une naissance vivante. Ceci n'est pas la norme dans les EDS, bien que le questionnaire modèle inclue la collecte de données sur les fausses couches survenues dans les cinq ans précédant l'enquête. Certains avancent que l'historique des grossesses permet de recueillir de meilleures données sur les naissances vivantes que les histoires génésiques ne reprenant que les naissances vivantes, en accroissant la probabilité que les décès néonataux très précoces soient déclarés, mais on ne dispose pas de données expérimentales qui le confirment.

Comme on le verra, l'histoire génésique complète est un instrument complexe de collecte des données qui nécessite une formation approfondie des enquêteurs et un contrôle de terrain attentif. Elle ne peut être utilisée que lorsque la femme elle-même est interviewée ; ce n'est pas possible quand c'est un proche qui répond. Ce n'est pas approprié pour les opérations de collecte à grande échelle comme les recensements de population.

L'histoire génésique tronquée (*en anglais, truncated birth history, TBH*)

Certaines enquêtes ont recours à des histoires génésiques tronquées, afin de réduire les coûts du travail de terrain et l'éventuelle lassitude des enquêtées. Une histoire tronquée recueille les mêmes informations qu'une histoire génésique complète sur chaque naissance, mais elle limite le nombre de naissances pour lesquelles l'information est recueillie, en fixant soit un intervalle de temps (par exemple les naissances des cinq dernières années), soit un nombre (par exemple, les naissances les plus récentes dans la limite de trois). L'expérience suggère que c'est une forme risquée de collecte des données pour des raisons à la fois théoriques et pratiques. Une limitation temporelle réduit fortement l'information sur les risques de mortalité des enfants relativement âgés (par exemple avec une limite à cinq ans, seuls les enfants nés entre 4 et 5 ans auparavant fourniront de l'information sur la mortalité aux âges entre 4 et 5 ans, et on n'aura aucune information sur la mortalité au-dessus de 5 ans). Il se peut aussi que les enfants décédés soient intentionnellement décalés hors de la fenêtre temporelle. Une limitation sur le nombre pose la question de biais de sélection (les femmes ayant eu plus de trois naissances récemment ne déclareront que les trois plus récentes) et celle des déclarations sélectives, qui peuvent favoriser la déclaration des enfants survivants au détriment de ceux qui sont décédés.

L'histoire génésique résumée

Dans une histoire génésique résumée, la seule information recueillie porte sur les nombres agrégés d'enfants nés et d'enfants encore vivants (ou décédés). La quantité de détails varie : de seulement deux questions (nombre d'enfants nés et nombre d'enfants encore vivants) à une histoire détaillée posant des questions sur les garçons et les filles séparément et sur les enfants survivants vivant avec leur mère ou vivant ailleurs, comme on l'a vu plus haut en présentant la suite des questions dans l'histoire génésique complète des EDS. On pense, sans avoir de base empirique à notre connaissance, que la forme plus détaillée des questions améliore la qualité des données. À la différence de l'histoire génésique complète, l'histoire génésique résumée est souvent obtenue de répondants substitués. Dans un recensement par exemple, il est courant qu'une seule personne fournisse l'information relative aux différents membres du ménage. C'est parfois la femme elle-même, mais pas toujours. Dans certains cas, un homme chef de ménage fournit l'information pour toutes les femmes vivant sur place.

L'histoire génésique résumée ne peut pas servir de base au calcul direct d'indicateurs de mortalité classiques. Les événements (décès) ne sont pas explicitement situés dans le temps et on ne peut pas calculer précisément de temps d'exposition au risque. Les taux conventionnels rapportant les événements à la durée d'exposition ne peuvent pas être calculés à partir de ces données. À la place, celles-ci sont analysées à l'aide de méthodes d'estimation indirecte de la mortalité des jeunes enfants, dans lesquelles l'âge (ou parfois la durée de mariage ou la durée écoulée depuis la première naissance) d'un groupe de femmes est pris comme indicateur de la durée moyenne d'exposition des enfants au risque de décéder, des schémas-types de fécondité et de mortalité par âge étant utilisés pour convertir les proportions d'enfants décédés parmi les enfants nés d'un groupe de femmes en une fonction classique de la table de mortalité.

Survie d'une naissance récente

En pratique, l'histoire génésique résumée ne permet pas une estimation très récente de la mortalité des jeunes enfants, pour des raisons qui seront expliquées dans le chapitre sur l'estimation indirecte de la mortalité des jeunes enfants (chapitre 16). Les recensements et les enquêtes à grande échelle se sont efforcés d'obtenir des estimations plus récentes grâce à des questions sur la survie d'un enfant né récemment. (L'information sur les naissances récentes sert aussi de base à des estimations de la fécondité). Les questions couramment utilisées sont essentiellement de deux formes. Dans un cas, on demande à la femme si elle a eu une naissance vivante au cours d'une période de référence (par exemple les 12 mois précédant l'enquête) et si cet enfant est encore vivant. Dans l'autre cas, on demande à la femme la date (mois et année) de sa naissance vivante la plus récente et si l'enfant est encore en vie. Ces deux types de question permettent de calculer la proportion de survivants (d'une large majorité) d'enfants nés dans les 12 mois avant l'enquête. Si les naissances sont distribuées uniformément au cours des 12 mois, la proportion d'enfants vivants est une estimation du paramètre de la table de mortalité $1 - \left(\frac{1L_0}{l_0} \right)$. La seconde forme de question permet aussi de calculer la survie à des âges plus avancés de l'enfance, selon la procédure de Brass-Backer présentée plus loin.

En pratique, comme on le verra dans la section sur l'évaluation de la qualité des données sur la fécondité récente

(chapitre 5), le nombre de naissances déclaré au cours des 12 derniers mois est souvent peu vraisemblable (généralement trop faible), ce qui soulève des questions sur des biais de sélectivité ou de déclaration qui peuvent dépendre de la survie de l'enfant. La conversion du paramètre $1 - \left(\frac{1L_0}{l_0} \right)$

en ${}_5q_0$, la mesure communément utilisée, dépend aussi du choix du système de tables-types de mortalité.

Blacker et Brass (2005) suggèrent d'utiliser des données du second format (la date de la naissance la plus récente) afin d'estimer la mortalité des jeunes enfants à partir de la survie des naissances les plus récentes dans les deux ans précédant un recensement ou une enquête. Les auteurs montrent que, pour des schémas classiques de mortalité par âge des jeunes enfants, la proportion de décédés parmi ces naissances peut, en l'absence de biais de sélection, être convertie en une estimation du taux de mortalité infantile ${}_1q_0$ par une simple multiplication par un facteur 1,09. Mais ce facteur d'ajustement s'appuie toutefois sur l'hypothèse que les naissances les plus récentes dans les 24 derniers mois représentent bien l'ensemble des naissances des 24 derniers mois (qui incluent toutes les naissances des femmes ayant plus d'une naissance pendant la période), et pas une sélection de ces naissances sur la base de la survie des enfants. Cette hypothèse se révèle être étonnamment éloignée de la réalité. L'analyse des données tirées d'histoires génésiques complètes dans les EDS montre que, pour les femmes ayant eu plusieurs naissances au cours de la période, les naissances les plus anciennes ont beaucoup moins de chances d'avoir survécu que les naissances les plus récentes. C'est en majeure partie parce que les intervalles entre naissances après un décès infantile précoce sont systématiquement plus courts que ceux faisant suite à la survie d'un enfant. Il en résulte que les naissances dans les 24 derniers mois qui ne survivent pas ont une probabilité plus élevée d'être suivies par une naissance ultérieure (laquelle peut être survivante). Quand ceci survient, l'enfant qui est décédé n'est pas déclaré comme plus récente naissance. Dans un échantillon des EDS africaines, la proportion d'enfants décédés parmi les naissances les plus récentes des 24 derniers mois est inférieure de plus de 20 % en moyenne à la proportion d'enfants décédés parmi l'ensemble des naissances des 24 mois. Même au Bangladesh, une population où les intervalles entre naissances sont longs et où on s'attendrait à ce l'effet soit moindre, le biais atteint encore près de 16 %.

Compte tenu des préoccupations à la fois sur la qualité des données dans le cas de la survie des naissances les plus

récentes dans les 12 mois précédant une enquête et sur le biais de sélection quand on utilise des données couvrant 24 mois, nous ne recommandons pas l'analyse des informations sur la survie des naissances les plus récentes.

DONNÉES RECUEILLIES

DANS DES ÉTABLISSEMENTS DE SANTÉ

Un poste majeur dans le budget des enquêtes auprès des ménages est le coût d'accès de l'enquêteur au ménage sélectionné. Une large part de cette dépense peut être évitée si on peut se servir de la venue des enquêtées auprès de l'interviewer, par exemple à l'occasion d'une visite dans un établissement de santé. Il se peut aussi que ces établissements enregistrent les naissances et les décès qui se produisent sur place au titre de l'activité courante du système d'information sanitaire de routine. L'utilisation de telles données pour l'estimation de la mortalité des jeunes enfants pose le problème des biais de sélection, dans la mesure où on ne peut jamais être sûr que les femmes qui visitent un établissement sont représentatives de l'ensemble des mères. Pour améliorer la couverture, on mène des investigations pour voir si des agents de vulgarisation sanitaire ou d'autres personnels de santé travaillant dans les communautés peuvent recueillir des données pertinentes sur les naissances et les décès. Une telle démarche s'apparente à l'instauration d'un système d'enregistrement par échantillons (*en anglais, sample registration system*).

Deux démarches de collecte des données auprès des femmes dans les établissements de santé ont été proposées, et une d'entre elles, la technique de la naissance précédente (*en anglais, previous birth technique*), a été largement développée (Brass and Macrae 1984). Pour l'essentiel, cette méthode consiste à demander aux femmes venant dans un établissement pour accoucher si leur précédente naissance, quand elles en ont eu une, est encore en vie. Étant donné les distributions habituelles des intervalles entre naissances, la proportion de naissances précédentes décédées peut être interprétée comme une probabilité de décéder avant un âge exact de l'enfance. Les données sont extrêmement faciles et peu coûteuses à recueillir et elles peuvent donner une indication sur les tendances de la mortalité des jeunes enfants, même si le niveau peut être affecté par des biais de sélection. Ces tendances seront cependant inexactes si le biais de sélection se modifie de façon sensible au fil du temps.

La seconde démarche part du postulat que les femmes qui visitent un établissement de santé constituent un échantil-

lon biaisé de l'ensemble des mères, sans qu'il y ait d'hypothèse *a priori* sur la direction du biais. Au lieu de supposer que prédominent les femmes avec des enfants en meilleure ou en moins bonne santé que la moyenne, la méthode vise à estimer les probabilités de sélection afin de pouvoir corriger des effets du biais. Par exemple, les femmes visitant un établissement de santé pourraient être interrogées sur leur âge, le nombre de leurs enfants déjà nés et de leurs enfants encore en vie, plus un certain nombre de questions complémentaires sur leurs caractéristiques socio-économiques. Les nombres d'enfants déjà nés et d'enfants encore en vie pourraient alors être modélisés sur la structure socio-économique de l'ensemble de la population, tirée par exemple d'un recensement de population. Cette démarche n'a jamais été testée à notre connaissance.

EFFETS DU VIH SUR LES MÉTHODES D'ESTIMATION DE LA MORTALITÉ DES JEUNES ENFANTS

Toutes les méthodes d'estimation de la mortalité des jeunes enfants fondées sur les déclarations des mères sur la survie de leurs enfants sont sujettes à des biais de sélection. Les migrations et les non réponses différentielles peuvent introduire des biais, mais les estimations de la mortalité des jeunes enfants fondées sur les déclarations des femmes risquent surtout d'être affectées par les conséquences d'une épidémie de VIH/Sida généralisée. La transmission verticale du VIH de la mère à l'enfant pendant la grossesse, lors de l'accouchement ou par l'allaitement dans les premiers mois de vie accroît le risque qu'un enfant soit séropositif par 35 % en l'absence d'antirétroviraux. Plus de 60 % des enfants séropositifs vont décéder avant leur 5^e anniversaire en l'absence de traitement par les antirétroviraux (Schneider, Zwahlen and Egger 2004 ; Todd, Glynn, Marston *et al.* 2007). Comme les mères souffrent également de risques de mortalité élevés, de nombreux décès d'enfants séropositifs, en particulier ceux nés 5 ans ou plus avant l'interview, ne seront pas enregistrés. C'est pourquoi la mortalité des jeunes enfants dans son ensemble sera sous estimée, que ce soit par des estimations directes ou indirectes de la mortalité des jeunes enfants.

Effet sur les estimations directes de la mortalité des jeunes enfants

Nous ne connaissons qu'une analyse de l'ampleur du biais dans les estimations directes de la mortalité des jeunes enfants réalisée sur des données réelles plutôt que des simu-

lations. Hallett, Gregson, Kurwa *et al.* (2010) utilisent les données d'une cohorte prospective ouverte dans la province de Manicaland au Zimbabwe pour mesurer le biais résultant du décès des mères séropositives. La cohorte a été interviewée entre juillet 1998 et février 2000, puis réinterviewée après trois et cinq ans. De 1998 à 2005, la prévalence du VIH dans la population étudiée a baissé de 22 à 18 %. Au cours du passage final en 2005, une histoire génésique complète a été recueillie auprès des femmes survivantes, et le taux de mortalité avant 5 ans a été estimé pour la période 1998-2005, une période de sept ans par opposition à la période quinquennale habituelle dans les EDS. Les estimations directes ont ensuite été comparées aux vraies valeurs obtenues *a posteriori* en ajoutant l'expérience de mortalité des jeunes enfants des femmes qui sont décédées avant 2005. Le biais, mesuré en comparant les estimations tirées des mères survivantes aux estimations pour l'ensemble des mères, est de 6,7 % pour les taux de mortalité infantile et de 9,8 % pour le taux de mortalité avant 5 ans. Hallett, Gregson, Kurwa *et al.* (2010) ont aussi développé un modèle de biais, qu'ils ont appliqué au Zimbabwe et à six autres pays ayant une prévalence de VIH modérée ou élevée pour la période 1980-2015. Ils l'ont fait en utilisant les données de prévalence de l'Onusida et des estimations des EDS pour les périodes avant l'épidémie. Le modèle montre que le biais dans les estimations directes s'accroît avec (a) la durée de l'épidémie et (b) la durée entre l'estimation et l'enquête. Inversement, le biais dans les estimations directes décroît avec le niveau de la mortalité des jeunes enfants qui est attribuable à des causes autres que le VIH.

Walker, Hill and Zhao (2012) ont développé un modèle de projection simple par la méthode des composantes, en séparant les naissances entre celles dont la mère est séronégative (où les enfants sont supposés ne pas être infectés), celles dont les mères sont séropositives sans que leurs enfants soient eux-mêmes infectés et celles où les mères sont séropositives et dont les enfants sont infectés à la naissance ou ultérieurement. Les deux premiers groupes sont supposés faire face à une mortalité de base, estimée en l'absence du VIH (tirée des tables types de mortalité), alors que le troisième groupe est supposé avoir une probabilité de décéder avant 5 ans de 62 %. On fait ensuite avancer en âge les mères séropositives jusqu'à la date d'enquête en prenant en compte leur surmortalité, et le taux de mortalité avant 5 ans estimé à partir des déclarations des femmes survivantes est comparé à celui qu'on aurait observé si toutes les femmes avaient survécu jusqu'à l'enquête. Aucun

ajustement n'a été introduit pour tenir compte de la prévention ou des traitements. Comme dans l'analyse de Hallett, Gregson, Kurwa *et al.* (2010), l'ampleur du biais dépend de la prévalence du VIH et de sa trajectoire passée, du niveau de la mortalité de base (hors VIH) avant 5 ans et de l'intervalle entre la période à laquelle se réfère l'estimation et l'enquête. On ne peut donc pas proposer une procédure simple pour évaluer l'ampleur du biais. Pour disposer malgré tout d'une orientation générale, nous avons fait figurer au tableau 15.1 quelques estimations du biais pour des pays ayant recueilli des histoires génésiques vers le milieu de la dernière décennie, avant que les traitements par les antirétroviraux ne soient largement diffusés, pour des périodes variant de 1-5, 6-10 et 11-15 ans avant chaque enquête. Il est cependant important de se souvenir que le biais est fonction de la mortalité de base (non-VIH) des jeunes enfants, qui n'est pas facile à estimer, et de la prévalence du VIH qui est généralement estimée de façon erronée.

Le biais le plus fort est pour la période 6-10 ans avant l'enquête ; il dépasse 10 % si la prévalence du VIH est supérieure à 5 %. Il est important de se souvenir que l'usage des rétroviraux pour prévenir la transmission de la mère à l'enfant et pour prolonger la survie aura un effet rapide sur la réduction du biais pour la période la plus récente, mais le biais pour les périodes passées subsistera encore pendant une décennie au moins après l'introduction d'une thérapie efficace.

Effets sur l'estimation indirecte de la mortalité des jeunes enfants

Peu d'études ont été consacrées à l'impact du VIH sur les estimations indirectes de la mortalité des jeunes enfants. Le

VIH n'affecte pas seulement la précision des estimations indirectes du fait de l'association entre la mortalité des enfants et celle de leurs mères, mais aussi parce que le VIH modifie la distribution par âge de la mortalité des jeunes enfants et qu'il a des répercussions sur les procédures permettant d'estimer des schémas de fécondité à partir des rapports des descendances atteintes observés. En outre, les risques de mortalité des jeunes enfants ne peuvent plus être supposés indépendants de l'âge des mères. Mais sur un point, les estimations indirectes peuvent être moins affectées que les estimations directes par la sélection liée à la survie maternelle, parce que l'analyse est conduite par groupe d'âge. Les mères de moins de 25 ans ayant peu de risque d'être décédées du VIH/Sida, les déclarations de la survie des jeunes enfants par les femmes de 15-19 et 20-24 ans, et même 25-29 ans, seront sans doute faiblement biaisées par le VIH. Mais il est vrai par ailleurs que les groupes 15-19 et 20-24 ans sont les plus biaisés par d'autres effets de sélection.

Ward et Zaba (2008) ont estimé le biais attendu en présence de VIH affectant les estimations indirectes de la mortalité des jeunes enfants dans un contexte épidémique stable (effets constants de l'incidence et de la mortalité). Selon leur modèle, le biais pour des estimations fondées sur les femmes de moins de 30 ans n'excède pas 5 % pour une prévalence adulte de 10 % ou moins ; il dépasse à peine 10 % pour une prévalence atteignant 30 %. Ces résultats sont rassurants. Bien sûr, l'épidémie de VIH a été tout sauf stable, se développant rapidement dans de nombreux pays jusque vers 2000 et reculant depuis lors à la fois en prévalence et en impact. Mais la dynamique de l'épidémie va tendre à réduire le biais à des niveaux inférieurs à ceux estimés par le modèle de Ward et Zaba.

TABLEAU 15.1 Estimations du biais pour des taux de mortalité avant 5 ans estimés pour des périodes 1-5, 6-10 et 11-15 avant chaque enquête : divers pays d'Afrique subsaharienne

| Pays | Année EDS | Prévalence approx. du VIH 2005 (%) | Taux supposé de mortalité avant 5 ans hors VIH | Biais estimé par période avant l'enquête (%) | | |
|---------------|-----------|------------------------------------|--|--|------|-------|
| | | | | 1-5 | 6-10 | 11-15 |
| Cote d'Ivoire | 2005 | 4,6 | 125 | 4,0 | 6,6 | 3,3 |
| Kenya | 2003 | 7,1 | 75 | 8,0 | 14,1 | 6,7 |
| Lesotho | 2004 | 23,4 | 75 | 13,2 | 15,7 | 2,1 |
| Namibie | 2006-07 | 15,3 | 50 | 13,7 | 22,7 | 10,4 |
| Zambie | 2007 | 15,0 | 150 | 6,8 | 13,9 | 13,0 |
| Zimbabwe | 2005 | 18,0 | 75 | 16,6 | 31,4 | 25,6 |

Source : Walker, Hill and Zhao (2012)

Mutemaringa (2010) a comparé les estimations indirectes tirées des EDS au Zimbabwe, au Kenya, au Lesotho, au Malawi, en Namibie et en Zambie aux estimations directes dans les mêmes enquêtes. L'auteur confirme que le biais provient avant tout de la corrélation des chances de survies entre mères et enfants. Le biais dans l'estimation fondée sur les déclarations des femmes de 25-29 ans est inférieur à 5 % dans trois cas sur six ; mais il dépasse 20 % au Zimbabwe et en Namibie. Le biais dans les estimations résultant des déclarations des femmes de 30-34 et 35-39 ans dépasse généralement 20 %, voire 30 % au Kenya et en Namibie.

AUTRES LECTURES ET RÉFÉRENCES

Outre les deux articles cités plus haut, le lecteur intéressé pourra consulter les travaux de Hill (1984, 1991) et le *Manuel X* (Division de la population des Nations Unies 1984), ouvrage historique de référence sur l'estimation indirecte. Il pourra trouver dans Rajaratnam, Tran, Lopez *et al.* (2010) des développements récents sur la modélisation de la mortalité avant 5 ans, ainsi que dans Preston (1985) et Rutstein and Rojas (2003) une description de l'estimation directe de la mortalité des jeunes enfants à partir des données d'enquête.

- Blacker J and W Brass. 2005. "The estimation of infant mortality from proportions dying among births in the past 24 months", *Southern African Journal of Demography* **10**(1-2):25-42.
- Brass W and S Macrae. 1984. "Childhood mortality estimated from reports on previous births given by mothers at the time of a maternity: I. Preceding-births technique", *Asian and Pacific Census Forum* **11**(2):5-8. <http://hdl.handle.net/10125/3561>.
- Division de la Population des Nations Unies. 1984. *Manuel X. Techniques indirectes d'estimation démographique*. New York : Nations Unies, Département des affaires économiques et sociales internationales, ST/ESA/SER.A/81. http://unstats.un.org/unsd/demographic/standmeth/handbooks/Manuel_X-fr.pdf
- Hallett TB, S Gregson, F Kurwa, G Garnett *et al.* 2010. "Measuring and correcting biased child mortality statistics in countries with generalized epidemics of HIV infection", *Bulletin of the World Health Organization* **88**(10):761-788. doi: <http://dx.doi.org/10.2471/BLT.09.071779>
- Hill K. 1984. "An evaluation of indirect methods for estimating mortality," in Vallin, J, JH Pollard and L Heligman (eds). *Methodologies for the Collection and Analysis of Mortality Data*. Liège, Belgium: Ordina Editions, pp. 145-176.

Nous concluons de ces analyses que les estimations de la mortalité des jeunes enfants tirées des déclarations de femmes de 25-29 ans concernant leurs enfants déjà nés et survivants ne seront pas fortement affectées même par une épidémie généralisée de VIH. D'après le schéma du biais en fonction de la prévalence du VIH obtenu par Ward et Zaba (2008), l'estimation de la mortalité des jeunes enfants obtenue par une analyse classique d'une histoire génésique résumée devrait être corrigée à la hausse de trois points pour 1 000 pour chaque 10 points de pourcentage de prévalence du VIH.

- Hill K. 1991. "Approaches to the measurement of childhood mortality: A comparative review", *Population Index* **57**(3):368-382. doi: <http://dx.doi.org/10.2307/3643873>
- Mutemaringa T. 2011. "Impact of HIV on estimates of child mortality derived using the summary birth history (CEB/CS) method." Unpublished MPhil thesis, Cape Town: University of Cape Town.
- Preston SH. 1985. "Mortality in childhood: Lessons from the WFS," in Cleland, J and J Hobcraft (eds). *Reproductive Change in Developing Countries*. Oxford: Oxford University Press, pp. 253-272.
- Rajaratnam JK, LN Tran, AD Lopez and CJL Murray. 2010. "Measuring under-five mortality: Validation of new low-cost methods", *PLoS Medicine* **7**(4):e1000253. doi: <http://dx.doi.org/doi:10.1371/journal.pmed.1000253>
- Rutstein S and G Rojas. 2003. *Guide to DHS Statistics*. Calverton, MD: ORC Macro.
- Schneider M, M Zwahlen and M Egger. 2004. *Natural history and mortality in HIV-positive individuals living in resource-poor settings*. http://www.epidem.org/Publications/unaid%20HQ_03_463871%20final.pdf
- Todd J, JR Glynn, M Marston, T Lutalo *et al.* 2007. "Time from HIV seroconversion to death: a collaborative analysis of eight studies in six low- and middle-income countries before highly active antiretroviral therapy", *AIDS* **21**(Suppl 6):555-563. doi: <http://dx.doi.org/10.1097/01.aids.0000299411.75269.e8>
- Walker PN, K Hill and F Zhao. 2012. "Child mortality estimation: Methods used to adjust for bias due to AIDS in estimating trends in under-five mortality", *PLoS Medicine* **9**(8):e1001298. doi: <http://dx.doi.org/10.1371/journal.pmed.1001298>
- Ward P and B Zaba. 2008. "The Effect of HIV on the Estimation of Child Mortality Using the Children Surviving / Children Ever Born Technique", *Southern African Journal of Demography* **11**(1):39-73.

Méthodes fondées
sur un seul recensement

Chapter 16 ESTIMATION INDIRECTE DE LA MORTALITÉ DES JEUNES ENFANTS

Kenneth Hill

DESCRIPTION DE LA MÉTHODE

Les méthodes indirectes, dont Brass et Coale (1968) ont été les pionniers, consistent à estimer la mortalité des jeunes enfants à partir des informations sur les nombres agrégés d'enfants nés vivants et d'enfants encore en vie (ou décédés) déclarés par des femmes classées par groupe d'âge (ou par durée depuis la première naissance ou par durée de mariage). Cette information est intitulée histoire génésique résumée des naissances (*en anglais, summary birth history, SBH*). La quantité de détails varie : de seulement deux questions (nombre d'enfants déjà nés et nombre d'enfants encore vivants) à une histoire détaillée posant des questions sur les garçons et les filles séparément et sur les enfants survivants vivant avec leur mère ou vivant ailleurs, comme on l'a vu dans le chapitre 15, en présentant la suite des questions dans l'histoire génésique complète des EDS. Parmi les enfants nés de femmes classées par âge (ou par durée depuis la première naissance, ou par durée de mariage) la proportion de ceux qui sont décédés reflète le niveau de la mortalité des jeunes enfants, mais elle est affectée aussi par d'autres éléments, en particulier la répartition des âges à la maternité et le schéma par âge de la mortalité des jeunes enfants. Les jeunes mères ont généralement de jeunes enfants, qui ont été exposés au risque de décéder peu de temps et à des périodes récentes ; pour ces mères, la proportion d'enfants décédés reflète donc le risque de mortalité à un âge précoce. Au contraire, les mères plus âgées, ont un mélange d'enfants jeunes et d'autres plus âgés exposés au risque de décéder plus longtemps et dans un passé en moyenne plus ancien. Par le jeu de modèles de fécondité et de mortalité des jeunes enfants, les proportions d'enfants décédés sont converties en probabilités de décéder avant des âges exacts de l'enfance, ${}_nq_0$. Plus les femmes sont âgées, plus la valeur de n est élevée.

Si la mortalité a évolué au fil du temps, les probabilités estimées reflètent les taux de mortalité qui ont prévalu à divers âges et diverses dates. Heureusement, une méthode de « localisation dans le temps » a été développée, qui permet d'estimer à quelle période antérieure s'appliquent les probabilités de décès du moment qui ont été approximées à partir de la proportion d'enfants décédés. La longueur de ces périodes s'accroît avec l'âge des répondantes. Ainsi, si les probabilités de décéder estimées à partir des déclarations des femmes de différents groupes d'âge sont traduites en un indice de mortalité commun, ces statistiques se réfèrent à différentes dates et permettent de tracer à grand trait l'évolution de la mortalité au fil du temps.

DONNÉES NÉCESSAIRES ET HYPOTHÈSES

Tableaux des données nécessaires

- Nombre de femmes par groupe quinquennal d'âge, de durée de mariage ou de durée depuis la première naissance.
- Nombre d'enfants nés vivants des femmes par groupe quinquennal pertinent (âge, durée depuis la première naissance, durée de mariage).
- Nombre d'enfants nés vivants qui sont décédés (ou encore en vie) au moment de l'enquête, par groupe quinquennal pertinent.
- Nombre de naissances au cours de l'année précédant l'enquête, par groupe quinquennal d'âge (facultatif).

Hypothèses importantes

- Les schémas par âge de la fécondité et de la mortalité des jeunes enfants dans la population sont adéquatement représentés par ceux utilisés dans le modèle lors du développement de la méthode.

Dans aucune période, la mortalité des enfants ne varie par groupe quinquennal d'âge des mères.

Il n'y a pas de corrélation entre les risques de mortalité des enfants et la survie des mères (par mortalité ou migration) dans la population (voir les effets du VIH sur les méthodes d'estimation de la mortalité des jeunes enfants (chapitre 15)).

Tous les changements récents dans la mortalité des enfants ont été graduels et sont allés dans la même direction.

Les nombres moyens d'enfants nés vivant par âge (ou par durée de mariage ou durée depuis la première naissance) à un moment donné reflètent bien les schémas de fécondité dans les cohortes correspondantes.

TRAVAUX PRÉPARATOIRES ET RECHERCHES PRÉLIMINAIRES

L'évaluation de la qualité des données pour une histoire génésique résumée des naissances relève d'un sous-ensemble des analyses décrites dans la section sur l'estimation directe de la mortalité des jeunes enfants pour l'analyse d'une histoire génésique complète dans le chapitre 17. Comme une histoire génésique résumée ne contient pas d'information sur les dates de chaque naissance, l'évaluation se limite à l'analyse des agrégats, tabulés par groupe d'âge de la mère (ou durée de mariage ou durée depuis la première naissance, si une telle information est disponible) ; les tabulations par année d'âge peuvent aussi être révélatrices s'il y a une importante attraction sur certains âges.

Comme toujours, les évaluations s'appuient sur la vraisemblance interne des données d'une part et sur la cohérence externe avec d'autres ensembles de données relatives à la population, d'autre part.

Au titre des évaluations internes, un premier contrôle doit porter sur le nombre moyen d'enfants nés vivants dans chaque groupe de femmes. (Notez que le dénominateur approprié pour ces calculs est l'ensemble des femmes, pas le nombre de mères ou le nombre de femmes déjà mariées). À moins que la fécondité ne connaisse une tendance à la hausse, le nombre moyen d'enfants nés vivants (appelé aussi descendance atteinte) doit s'accroître au fil des groupes quinquennaux d'âge. Un second contrôle doit porter sur le nombre moyen d'enfants décédés pour chaque groupe quinquennal. À moins que la mortalité des jeunes enfants ou la fécondité ne connaissent une tendance à la hausse, le nombre moyen d'enfants décédés doit aussi s'accroître avec l'âge. Si on dispose de l'information sur le sexe des enfants nés vivants, les

rapports de masculinité des naissances doivent être calculés. Comme pour les histoires génésiques complètes, toute tendance pour ces rapports à s'écarter de 100 à 106 garçons pour 100 filles, ou à s'accroître avec l'âge (ou la durée du mariage ou la durée depuis la première naissance) doit être considérée comme un signal d'alarme, à moins que la population soit connue pour sa pratique d'un avortement sélectif par sexe. Parmi les contrôles externes, les comparaisons par cohorte des nombres moyens d'enfants nés vivants et d'enfants décédés sont souvent révélatrices ; elles sont décrites pour les histoires génésiques complètes dans la section sur l'estimation directe de la mortalité des jeunes enfants.

PRÉCAUTIONS ET MISES EN GARDE

- La correction des erreurs de données doit être menée avec attention. Les femmes avec des données manquantes sur les nombres d'enfants nés vivants, ou les nombres d'enfants décédés (ou survivants), ou les deux, doivent être exclues de l'analyse. Mais les femmes sans enfants doivent être incluses.
- Les imputations doivent également être faites avec prudence. Dans les grandes enquêtes, l'information sur le nombre d'enfants nés vivant manque le plus souvent chez des femmes sans enfants (voir la correction d'El-Badry (chapitre 4), bien qu'il est à noter ici que la correction n'est pas nécessaire car elle affecterait chaque parité moyenne dans la même proportion). La méthode d'imputation « *hot deck* », où la donnée manquante est remplacée par la valeur observée pour un répondant choisi au hasard, peut entraîner de sérieux biais.
- L'hypothèse selon laquelle la mortalité des enfants ne varie pas en fonction du groupe quinquennal des mères est généralement incorrecte quand la variable temporelle est l'âge. Les enfants de jeunes mères font systématiquement face à une mortalité supérieure à celle des enfants nés de mères âgées de plus de 25 ans. En conséquence, les estimations indirectes tirées des femmes âgées de 15-19 ans (tout particulièrement) et 20-24 ans (dans une certaine mesure) tendent à surestimer la mortalité des jeunes enfants dans la population. Des corrections pour ces effets sont rarement faites (Collumbien and Sloggett 2001). C'est en partie cette distorsion qui a conduit au développement de méthodes fondées sur la durée de mariage ou la durée depuis la première naissance. Mais ces variantes ont-elles-mêmes leurs inconvénients, que nous évoquerons plus loin.
- L'application de la méthode dans des populations où le VIH est très répandu nécessite des précautions (voir la

section sur les effets du VIH sur les méthodes d'estimation de la mortalité des jeunes enfants dans le chapitre 15).

APPLICATION DE LA MÉTHODE

Histoire génésique résumée des naissances par groupe quinquennal d'âge/durée depuis la première naissance/durée de mariage

Étape 1 : Calculer les proportions de décédés parmi les enfants déjà nés, ${}_5PD_x$

Pour chaque groupe quinquennal ($x, x + 5$) de femmes, les proportions d'enfants décédés sont calculées en divisant le nombre d'enfants décédés par le nombre d'enfants nés vivants.

Étape 2 : Calculer les nombres moyens d'enfants nés vivants des femmes dans chaque groupe quinquennal, ${}_5P_x$

Pour chaque groupe quinquennal de femmes, diviser le nombre déclaré d'enfants nés vivants ${}_5CEB_x$ (CEB pour *children ever born*) par le nombre de femmes ${}_5N_x$ dans le groupe. Noter que si la variable de temps est l'âge, le dénominateur doit être l'ensemble des femmes, quels que soient leur état matrimonial ou leur histoire féconde.

Étape 3 : Choisir une famille de tables-types de mortalité

La répartition par âge de la mortalité des jeunes enfants a une grande importance quand il s'agit de traduire une proportion d'enfants décédés en un quotient ${}_nq_0$ classique et de traduire ce quotient ${}_nq_0$ en un indice courant tel que le taux de mortalité avant 5 ans. Dans une population où ont été collectées des histoires génésiques complètes relativement récentes, la famille de tables-types peut être choisie sur cette base en faisant apparaître sur un graphique l'association existante entre ${}_4q_1$ et ${}_1q_0$ en même temps que cette même relation dans les tables-types de Coale-Demeny et celles des Nations Unies. Si ce genre de données n'existe pas, on peut choisir la table-type en s'appuyant sur les schémas de mortalité des jeunes enfants observés dans des pays voisins. Il est peu vraisemblable que les données s'ajustent parfaitement à un des modèles. L'analyste doit choisir le modèle qui représente au mieux l'éventail des observations disponibles. La même famille de tables-types doit être utilisée aux étapes 4, 5 et 6. Les étapes 4 et 5 ont différentes variantes, selon que les données qui font l'objet de l'analyse sont classées par âge de la mère (variante a), durée depuis la première naissance

de la mère (variante b), ou durée de mariage de la mère (variante c).

Données classées par âge de la mère (variante a)

Étape 4a(1) : Estimer l'âge moyen de la courbe de fécondité par âge

Cette étape n'est nécessaire que lorsque l'analyse s'appuie sur une des tables-types de mortalité des Nations Unies pour les pays en développement. L'âge moyen est calculé à partir des taux de fécondité par âges ${}_5f_x$ de la façon suivante :

$$\bar{m} = \frac{\sum_{x=15,5}^{45} {}_5f_x \cdot (x + 2)}{\sum_{x=15,5}^{45} {}_5f_x}$$

Le terme $(x + 2)$ au numérateur représente le point médian du groupe d'âge entre x et $x + 5$ au moment où ont eu lieu les naissances. Ceci suppose que les taux de fécondité par âge sont calculés à partir d'informations sur les naissances dans l'année précédant l'enquête classées par âge de la femme au moment de l'enquête (voir la section sur l'évaluation de la qualité des données sur la fécondité récente pour plus de détails). Si les taux de fécondité par âge sont calculés à partir des naissances enregistrées par âge de la mère à la naissance, le terme doit être $(x + 2,5)$. Notez que c'est seulement la forme de la courbe de fécondité qui détermine la moyenne ; il n'est donc pas nécessaire de corriger le niveau, par exemple par un modèle relationnel de Gompertz, avant de calculer l'âge moyen.

Étape 4a(2) : Estimer ${}_nq_0$ à partir de chaque ${}_5PD_x$

Une fois qu'une famille de tables-types de mortalité j a été identifiée, les paramètres appropriés $a(x, j)$, $b(x, j)$ et $c(x, j)$ (et $d(x, j)$ si une table type des Nations Unies est utilisée) sont remplacés dans l'équation suivante :

$$\frac{{}_nq_0}{{}_5PD_x} = a(x, j) + b(x, j) \times \frac{{}_5P_{15}}{{}_5P_{20}} + c(x, j) \times \frac{{}_5P_{20}}{{}_5P_{25}} + d(x, j) \times \bar{m}$$

Notez que $d(x, j)$ est nul, sauf si on utilise les tables-types des Nations Unies.

Pour chaque groupe d'âge ($x, x + 5$), ${}_nq_0$ est estimé en multipliant la partie droite de l'équation par la valeur observée de ${}_5PD_x$ (tableau 16.1).

TABLEAU 16.1 Valeurs de $a(x,j)$, $b(x,j)$, $c(x,j)$ et éventuellement $d(x,j)$ pour estimer les probabilités de décéder avant les âges exacts de l'enfance à partir des proportions de décédés parmi les enfants nés vivants, classés par âge de la mère

| Famille j | Groupe d'âge de la mère et valeur de n dans ${}_nq_0$ | | | | | | | |
|-----------------------------------|---|----------|----------|----------|----------|----------|----------|----------|
| | Coefficient | 15-19 | 20-24 | 25-29 | 30-34 | 35-39 | 40-44 | 45-49 |
| | Valeur de n dans ${}_nq_0$ | 1 | 2 | 3 | 5 | 10 | 15 | 20 |
| Princeton « Nord » | $a(x,j)$ | 1,1119 | 1,2390 | 1,1884 | 1,2046 | 1,2586 | 1,2240 | 1,1772 |
| | $b(x,j)$ | - 2,9287 | - 0,6865 | 0,0421 | 0,3037 | 0,4236 | 0,4222 | 0,3486 |
| | $c(x,j)$ | 0,8507 | - 0,2745 | - 0,5156 | - 0,5656 | - 0,5898 | - 0,5456 | - 0,4624 |
| Princeton « Sud » | $a(x,j)$ | 1,0819 | 1,2846 | 1,2223 | 1,1905 | 1,1911 | 1,1564 | 1,1307 |
| | $b(x,j)$ | - 3,0005 | - 0,6181 | 0,0851 | 0,2631 | 0,3152 | 0,3017 | 0,2596 |
| | $c(x,j)$ | 0,8689 | - 0,3024 | - 0,4704 | - 0,4487 | - 0,4291 | - 0,3958 | - 0,3538 |
| Princeton « Est » | $a(x,j)$ | 1,1461 | 1,2231 | 1,1593 | 1,1404 | 1,1540 | 1,1336 | 1,1201 |
| | $b(x,j)$ | - 2,2536 | - 0,4301 | 0,0581 | 0,1991 | 0,2511 | 0,2556 | 0,2362 |
| | $c(x,j)$ | 0,6259 | - 0,2245 | - 0,3479 | - 0,3487 | - 0,3506 | - 0,3428 | - 0,3268 |
| Princeton « Ouest » | $a(x,j)$ | 1,1415 | 1,2563 | 1,1851 | 1,1720 | 1,1865 | 1,1746 | 1,1639 |
| | $b(x,j)$ | - 2,7070 | - 0,5381 | 0,0633 | 0,2341 | 0,3080 | 0,3314 | 0,3190 |
| | $c(x,j)$ | 0,7663 | - 0,2637 | - 0,4177 | - 0,4272 | - 0,4452 | - 0,4537 | - 0,4435 |
| Nations Unies « Amérique latine » | $a(x,j)$ | 0,6892 | 1,3625 | 1,0877 | 0,7500 | 0,5605 | 0,5024 | 0,5326 |
| | $b(x,j)$ | - 1,6937 | - 0,3778 | 0,0197 | 0,0532 | 0,0222 | 0,0028 | 0,0052 |
| | $c(x,j)$ | 0,6464 | - 0,2892 | - 0,2986 | - 0,1106 | 0,0170 | 0,0048 | 0,0256 |
| | $d(x,j)$ | 0,0106 | - 0,0041 | 0,0024 | 0,0115 | 0,0171 | 0,0180 | 0,0168 |
| Nations Unies « Chili » | $a(x,j)$ | 0,8274 | 1,3129 | 1,0632 | 0,8236 | 0,6895 | 0,6098 | 0,5615 |
| | $b(x,j)$ | - 1,5854 | - 0,2457 | 0,0196 | 0,0293 | 0,0068 | - 0,0014 | 0,0040 |
| | $c(x,j)$ | 0,5949 | - 0,2329 | - 0,1996 | - 0,0684 | 0,0032 | 0,0166 | 0,0073 |
| | $d(x,j)$ | 0,0097 | - 0,0031 | 0,0021 | 0,0081 | 0,0119 | 0,0141 | 0,0159 |
| Nations Unies « Asie du Sud » | $a(x,j)$ | 0,6749 | 1,3716 | 1,0899 | 0,7694 | 0,6156 | 0,6077 | 0,6952 |
| | $b(x,j)$ | - 1,7580 | - 0,3652 | 0,0299 | 0,0548 | 0,0231 | 0,0040 | 0,0018 |
| | $c(x,j)$ | 0,6805 | - 0,2966 | - 0,2887 | - 0,0934 | 0,0298 | 0,0573 | 0,0306 |
| | $d(x,j)$ | 0,0109 | - 0,0041 | 0,0024 | 0,0108 | 0,0149 | 0,0141 | 0,0109 |
| Nations Unies « Extrême orient » | $a(x,j)$ | 0,7194 | 1,2671 | 1,0668 | 0,7833 | 0,5765 | 0,4115 | 0,3071 |
| | $b(x,j)$ | - 1,3143 | - 0,2996 | 0,0017 | 0,0307 | 0,0068 | 0,0014 | 0,0111 |
| | $c(x,j)$ | 0,5432 | - 0,2105 | - 0,2424 | - 0,1103 | - 0,0202 | 0,0083 | 0,0129 |
| | $d(x,j)$ | 0,0093 | - 0,0029 | 0,0019 | 0,0098 | 0,0165 | 0,0213 | 0,0251 |
| Nations Unies « Général » | $a(x,j)$ | 0,7210 | 1,3115 | 1,0768 | 0,7682 | 0,5769 | 0,4845 | 0,4760 |
| | $b(x,j)$ | - 1,4686 | - 0,3360 | 0,0109 | 0,0439 | 0,0176 | 0,0034 | 0,0071 |
| | $c(x,j)$ | 0,5746 | - 0,2475 | - 0,2695 | - 0,1090 | 0,0038 | 0,0036 | 0,0246 |
| | $d(x,j)$ | 0,0095 | - 0,0034 | 0,0021 | 0,0105 | 0,0165 | 0,0187 | 0,0189 |

Sources : Tables de Princeton: *Division de la Population des Nations Unies (1984)* ; Tables des Nations Unies : *UN Population Division (1991)*

Étape 5a : Estimer la date de référence $t(x)$ de chaque ${}_nq_0$ estimé

La famille de tables-types de mortalité j ayant été identifiée, les paramètres $e(x,j)$, $f(x,j)$ et $g(x,j)$ sont remplacés dans l'équation suivante :

$$t(x) = e(x, j) + f(x, j) \times \frac{{}_5P_{15}}{{}_5P_{20}} + f(x, j) \times \frac{{}_5P_{20}}{{}_5P_{25}}$$

La localisation des estimations dans le temps est obtenue en retranchant les $t(x)$ de la date du recensement ou de l'enquête (tableau 16.2).

TABLEAU 16.2 Valeurs de $e(x,j)$, $f(x,j)$ et $g(x,j)$ pour estimer la date de référence $t(x)$ pour les probabilités de décéder avant les âges exacts de l'enfance à partir des proportions d'enfants décédés parmi les enfants nés vivants par âge de la mère

| Famille j | Groupe d'âge de la mère et valeur de n dans ${}_nq_0$ | | | | | | | |
|-----------------------------------|---|----------|----------|--------|----------|-----------|-----------|-----------|
| | Coefficient | 15-19 | 20-24 | 25-29 | 30-34 | 35-39 | 40-44 | 45-49 |
| | Valeur de n dans ${}_nq_0$ | 1 | 2 | 3 | 5 | 10 | 15 | 20 |
| Princeton « Nord » | $e(x,j)$ | 1,0921 | 1,3207 | 1,5996 | 2,0779 | 2,7705 | 4,1520 | 6,9650 |
| | $f(x,j)$ | 5,4732 | 5,3751 | 2,6268 | - 1,7908 | - 7,3403 | - 12,2448 | - 13,9160 |
| | $g(x,j)$ | - 1,9672 | 0,2133 | 4,3701 | 9,4126 | 14,9352 | 19,2349 | 19,9542 |
| Princeton « Sud » | $e(x,j)$ | 1,0900 | 1,3079 | 1,5173 | 1,9399 | 2,6157 | 4,0794 | 7,1796 |
| | $f(x,j)$ | 5,4443 | 5,5568 | 2,6755 | - 2,2739 | - 8,4819 | - 13,8308 | - 15,3880 |
| | $g(x,j)$ | - 1,9721 | 0,2021 | 4,7471 | 10,3876 | 16,5153 | 21,1866 | 21,7892 |
| Princeton « Est » | $e(x,j)$ | 1,0959 | 1,2921 | 1,5021 | 1,9347 | 2,6197 | 4,1317 | 7,3657 |
| | $f(x,j)$ | 5,5864 | 5,5897 | 2,4692 | - 2,6419 | - 8,9693 | - 14,3550 | - 15,8083 |
| | $g(x,j)$ | - 1,9949 | 0,3631 | 5,0927 | 10,8533 | 17,0981 | 21,8247 | 22,3005 |
| Princeton « Ouest » | $e(x,j)$ | 1,0970 | 1,3062 | 1,5305 | 1,9991 | 2,7632 | 4,3468 | 7,5242 |
| | $f(x,j)$ | 5,5628 | 5,5677 | 2,5528 | - 2,4261 | - 8,4065 | - 13,2436 | - 14,2013 |
| | $g(x,j)$ | - 1,9956 | 0,2962 | 4,8962 | 10,4282 | 16,1787 | 20,1990 | 20,0162 |
| Nations Unies « Amérique latine » | $e(x,j)$ | 1,1703 | 1,6955 | 1,8296 | 2,1783 | 2,8836 | 4,4580 | 6,9351 |
| | $f(x,j)$ | 0,5129 | 4,1320 | 2,9020 | - 2,5688 | - 10,3282 | - 17,1809 | - 19,3871 |
| | $g(x,j)$ | - 0,3850 | - 0,1635 | 3,4707 | 9,0883 | 15,4301 | 20,4296 | 23,4007 |
| Nations Unies « Chili » | $e(x,j)$ | 1,3092 | 1,6897 | 1,8368 | 2,2036 | 2,9955 | 4,7734 | 7,4495 |
| | $f(x,j)$ | 1,9474 | 4,6176 | 2,6370 | - 3,3520 | - 11,4013 | - 17,8850 | - 19,0513 |
| | $g(x,j)$ | - 0,7982 | - 0,0173 | 4,0305 | 9,9233 | 16,3441 | 20,8883 | 23,0529 |
| Nations Unies « Asie du Sud » | $e(x,j)$ | 1,1922 | 1,7173 | 1,8631 | 2,1808 | 2,7654 | 4,1378 | 6,4885 |
| | $f(x,j)$ | 0,7940 | 4,3117 | 2,8767 | - 2,7219 | - 10,8808 | - 18,6219 | - 22,2001 |
| | $g(x,j)$ | - 0,5425 | - 0,1653 | 3,5848 | 9,3705 | 16,2255 | 22,2390 | 26,4911 |
| Nations Unies « Extrême orient » | $e(x,j)$ | 1,2779 | 1,7471 | 1,9107 | 2,3172 | 3,2087 | 5,1141 | 7,6383 |
| | $f(x,j)$ | 1,5714 | 4,2638 | 2,7285 | - 2,6259 | - 9,8891 | - 15,3263 | - 15,5739 |
| | $g(x,j)$ | - 0,6994 | - 0,0752 | 3,5881 | 9,0238 | 14,7339 | 18,2507 | 19,7669 |
| Nations Unies « Général » | $e(x,j)$ | 1,2136 | 1,7025 | 1,8360 | 2,1882 | 2,9682 | 4,6526 | 7,1425 |
| | $f(x,j)$ | 0,9740 | 4,1569 | 2,8632 | - 2,6521 | - 10,3053 | - 16,6920 | - 18,3021 |
| | $g(x,j)$ | - 0,5247 | - 0,1232 | 3,5220 | 9,1961 | 15,3161 | 19,8534 | 22,4168 |

Sources : Tables de Princeton : *Division de la Population des Nations Unies (1984)* ; Tables des Nations Unies : *UN Population Division (1991)*

Données classées par durée depuis la première naissance des femmes (variante b)

Étape 4b : Estimer ${}_nq_0$ à partir de chaque ${}_5PD_x$

Une fois qu'une famille de tables-types de mortalité j a été identifiée, les paramètres appropriés $a(x,j)$, $b(x,j)$ et $c(x,j)$ sont remplacés dans l'équation suivante :

$$\frac{{}_nq_0}{{}_5PD_x} = a(x,j) + b(x,j) \times \frac{{}_5P_{15}}{{}_5P_{20}} + c(x,j) \times \frac{{}_5P_{20}}{{}_5P_{25}}$$

Note : les coefficients et les valeurs de n dans ${}_nq_0$ ont été mis à jour par Hill à partir de ceux publiés dans Hill and Figueroa (2001)

Pour chaque groupe d'âge $(x, x+5)$, ${}_nq_0$ est estimé en multipliant la partie droite de l'équation par la valeur observée de ${}_5PD_x$ (tableau 16.3).

Étape 5b : Estimer la date de référence $t(x)$ de chaque ${}_nq_0$ estimé

La famille de tables-types de mortalité j ayant été identifiée, les paramètres $e(x,j)$, $f(x,j)$ et $g(x,j)$ sont remplacés dans l'équation suivante :

$$t(x) = e(x,j) + f(x,j) \times \frac{{}_5P_{15}}{{}_5P_{20}} + g(x,j) \times \frac{{}_5P_{20}}{{}_5P_{25}}$$

Les résultats sont présentés au tableau 16.4. Note : les coefficients et les valeurs de n dans ${}_nq_0$ ont été mis à jour par Hill à partir de ceux publiés dans Hill and Figueroa (2001).

Données classées par durée depuis le mariage de la mère (variante c)

Étape 4c : Estimer ${}_nq_0$ à partir de chaque ${}_5PD_x$

Une fois qu'une famille de tables-types de mortalité j a été identifiée, les paramètres appropriés $a(x,j)$, $b(x,j)$ et $c(x,j)$ sont remplacés dans l'équation suivante :

$$\frac{{}_nq_0}{{}_5PD_x} = a(x,j) + b(x,j) \times \frac{{}_5P_{15}}{{}_5P_{20}} + c(x,j) \times \frac{{}_5P_{20}}{{}_5P_{25}}$$

Pour chaque groupe d'âge $(x, x+5)$, ${}_nq_0$ est estimé en multipliant la partie droite de l'équation par la valeur observée de ${}_5PD_x$ (tableau 16.5).

Étape 5c : Estimer la date de référence $t(x)$ de chaque ${}_nq_0$ estimé

La famille de tables-types de mortalité j ayant été identifiée, les paramètres $e(x,j)$, $f(x,j)$ et $g(x,j)$ sont remplacés dans l'équation suivante :

$$t(x) = e(x,j) + f(x,j) \times \frac{{}_5P_{15}}{{}_5P_{20}} + g(x,j) \times \frac{{}_5P_{20}}{{}_5P_{25}}$$

Étape 6 : Convertir chaque estimation de ${}_nq_0$ en une estimation de ${}_5q_0$

Dans les applications de l'estimation indirecte de la mortalité des jeunes enfants présentées ici, chacune des probabilités de décéder avant les âges exacts de l'enfance, ${}_nq_0$, est convertie en une valeur de α , le paramètre de niveau d'un système relationnel logit de tables-types de mortalité. Le paramètre α est ensuite utilisé pour estimer la probabilité de décéder entre la naissance et le 5^e anniversaire, ${}_5q_0$:

$$\alpha = 0,5 \left(\ln \left(\frac{{}_nq_0}{1 - {}_nq_0} \right) \right) - Y^s(n)$$

où les estimations de ${}_nq_0$ proviennent de l'Étape 4 et les valeurs de $Y^s(n)$ sont des transformations logit de la table de mortalité qui sert de standard. Ainsi, on obtient une série de valeurs de α correspondant aux probabilités de décéder estimées à partir des données des différents groupes d'âge des répondantes. Pour chaque α on a :

$${}_5\hat{q}_0 = \frac{e^{2(\alpha + {}_5Y_0')}}{1 + e^{2(\alpha + {}_5Y_0')}}}$$

Pour appliquer la démarche du système logit relationnel, il est nécessaire de choisir une table de mortalité pour servir de référence. Pour appliquer la procédure indirecte d'estimation, il faut identifier un réseau de tables-types approprié pour les Étapes 4 et 5, et la table de référence doit être tirée de la même famille de tables. Le niveau précis de la mortalité au sein de la famille est moins important que la famille elle-même (la sélection appropriée permettant de supposer que β est égal à 1 dans le système logit relationnel). C'est pourquoi nous recommandons de choisir comme standard une table de mortalité avec une espérance de vie de 60 ans.

| | | Durée depuis la première naissance de la mère | | | | |
|------------------------|------------------------------|---|---------|---------|---------|---------|
| Famille j | Coefficient | 0-4 | 5-9 | 10-14 | 15-19 | 20-24 |
| | Valeur de n dans ${}_nq_0$ | 2 | 5 | 5 | 5 | 10 |
| Princeton « Nord » | $a(x,j)$ | 1,1980 | 1,2248 | 1,2076 | 1,2030 | 1,3292 |
| | $b(x,j)$ | -0,1266 | -0,1919 | -0,0105 | 0,0896 | 0,1598 |
| | $c(x,j)$ | 0,0038 | -0,0870 | -0,2911 | -0,4265 | -0,5778 |
| Princeton « Sud » | $a(x,j)$ | 1,1705 | 1,3166 | 1,2952 | 1,2836 | 1,5269 |
| | $b(x,j)$ | -0,1461 | -0,3157 | -0,0423 | 0,1308 | 0,2659 |
| | $c(x,j)$ | 0,0051 | -0,0971 | -0,4295 | -0,6496 | -0,9174 |
| Princeton « Est » | $a(x,j)$ | 1,2182 | 1,2769 | 1,2731 | 1,2585 | 1,3410 |
| | $b(x,j)$ | -0,1809 | -0,2268 | 0,0005 | 0,1216 | 0,1749 |
| | $c(x,j)$ | 0,0214 | -0,1052 | -0,3720 | -0,5013 | -0,5964 |
| Princeton « Ouest » | $a(x,j)$ | 1,2049 | 1,2573 | 1,2431 | 1,2469 | 1,4258 |
| | $b(x,j)$ | -0,1553 | -0,2266 | -0,0230 | 0,0999 | 0,1948 |
| | $c(x,j)$ | 0,0135 | -0,0944 | -0,3409 | -0,5267 | -0,7454 |

TABLEAU 16.3 Valeurs de $a(x,j)$, $b(x,j)$, $c(x,j)$ pour estimer les probabilités de décéder avant les âges exacts de l'enfance à partir des proportions d'enfants décédés parmi les enfants nés vivants classés par durée depuis la première naissance de la mère

| | | Durée depuis la première naissance de la mère | | | | |
|------------------------|------------------------------|---|-------|-------|-------|-------|
| Famille j | Coefficient | 0-4 | 5-9 | 10-14 | 15-19 | 20-24 |
| | Valeur de n dans ${}_nq_0$ | 2 | 5 | 5 | 5 | 10 |
| Princeton « Nord » | $e(x,j)$ | 1,71 | 2,16 | 0,66 | -1,96 | -3,85 |
| | $f(x,j)$ | 1,07 | 4,36 | 3,50 | -0,90 | -6,42 |
| | $g(x,j)$ | -0,35 | 0,12 | 6,65 | 17,66 | 28,94 |
| Princeton « Sud » | $e(x,j)$ | 1,68 | 2,29 | 1,19 | -1,01 | -2,68 |
| | $f(x,j)$ | 0,96 | 3,84 | 3,45 | -0,18 | -5,06 |
| | $g(x,j)$ | -0,32 | -0,01 | 5,41 | 15,03 | 25,21 |
| Princeton « Est » | $e(x,j)$ | 1,68 | 2,19 | 0,71 | -1,96 | -4,06 |
| | $f(x,j)$ | 0,99 | 4,28 | 3,63 | -0,71 | -6,35 |
| | $g(x,j)$ | -0,33 | 0,02 | 6,36 | 17,42 | 29,14 |
| Princeton « Ouest » | $e(x,j)$ | 1,70 | 2,20 | 0,86 | -1,46 | -2,97 |
| | $f(x,j)$ | 1,03 | 4,20 | 3,47 | -0,69 | -5,80 |
| | $g(x,j)$ | -0,34 | 0,06 | 6,21 | 16,49 | 26,65 |

TABLEAU 16.4 Valeurs de $e(x,j)$, $f(x,j)$ et $g(x,j)$ pour estimer la date de référence $t(x)$ pour les probabilités de décéder avant les âges exacts de l'enfance à partir des proportions d'enfants décédés parmi les enfants nés vivants classés par durée depuis la première naissance de la mère

TABLEAU 16.5 Valeurs de $a(x,j)$, $b(x,j)$, $c(x,j)$ pour estimer les probabilités de décéder avant les âges exacts de l'enfance à partir des proportions d'enfants décédés parmi les enfants nés vivants classées par durée de mariage de la mère

| | | Durée de mariage de la mère | | | | | |
|---------------------|----------|-----------------------------|---------|---------|---------|---------|---------|
| Famille j | | 0-4 | 5-9 | 10-14 | 15-19 | 20-24 | 25-29 |
| | | 2 | 3 | 5 | 10 | 15 | 20 |
| Princeton « Nord » | $a(x,j)$ | 1,2615 | 1,1957 | 1,3067 | 1,4701 | 1,5039 | 1,4798 |
| | $b(x,j)$ | -0,5340 | -0,4103 | -0,0103 | 0,1763 | 0,0039 | -0,2487 |
| | $c(x,j)$ | 0,1252 | -0,0930 | -0,4618 | -0,7268 | -0,7071 | -0,5582 |
| Princeton « Sud » | $a(x,j)$ | 1,3103 | 1,2309 | 1,2774 | 1,3493 | 1,3592 | 1,3532 |
| | $b(x,j)$ | -0,5856 | -0,3463 | 0,0336 | 0,1366 | -0,0315 | -0,1978 |
| | $c(x,j)$ | 0,1367 | -0,1073 | -0,3987 | -0,5403 | -0,4944 | -0,4099 |
| Princeton « Est » | $a(x,j)$ | 1,2299 | 1,1611 | 1,2036 | 1,2773 | 1,3014 | 1,3160 |
| | $b(x,j)$ | -0,3998 | -0,2451 | 0,0171 | 0,1015 | -0,0219 | -0,1630 |
| | $c(x,j)$ | 0,0910 | -0,0797 | -0,2992 | -0,4276 | -0,4195 | -0,3751 |
| Princeton « Ouest » | $a(x,j)$ | 1,2584 | 1,1841 | 1,2446 | 1,3353 | 1,3875 | 1,4227 |
| | $b(x,j)$ | -0,4683 | -0,3006 | 0,0131 | 0,1157 | -0,0193 | -0,1954 |
| | $c(x,j)$ | 0,1080 | -0,0892 | -0,3555 | -0,5245 | -0,5472 | -0,5127 |

Étape 7 : Identifier et interpréter les résultats

On doit maintenant évaluer la pertinence des estimations de ${}_nq_0$ obtenues pour chaque groupe d'âge, des estimations correspondantes de ${}_5q_0$, des estimations des localisations dans le temps et de l'évolution de ${}_5q_0$ au fil du temps. Représentée graphiquement, la série des ${}_5q_0$ donne une indication sur l'évolution chronologique de la mortalité des jeunes enfants. Si des données proviennent de plusieurs recensements ou enquêtes, diverses estimations de ${}_5q_0$ peuvent être comparées pour la même période, permettant ainsi une évaluation de la cohérence et de la fiabilité des données.

EXEMPLE

L'exemple s'appuie sur les données relatives aux enfants nés vivants et aux enfants survivants par âge de la mère, tirées du recensement de 2008 au Malawi. La méthode est mise en œuvre dans le fichier Excel associé.

Étape 1 : Calculer les proportions de décédés parmi les enfants nés vivants, ${}_5PD_x$

Le tableau 16.7 présente les données de base sur le nombre de femmes, le nombre d'enfants nés vivants et le nombre d'enfants survivants par groupe quinquennal d'âge de la mère tirées du recensement de 2008 du Malawi. La proportion d'enfants décédés parmi les enfants nés vivants, ${}_5PD_x$, est calculée comme le complément à l'unité du rapport entre

le nombre d'enfants survivants (${}_5CS_x$ pour *children surviving*) et le nombre d'enfants nés vivants (${}_5CEB_x$) :

$${}_5PD_x = 1 - \frac{{}_5CS_x}{{}_5CEB_x}$$

Les résultats apparaissent dans la cinquième et sixième colonnes du tableau 16.7.

Étape 2 : Calculer les nombres moyens d'enfants nés vivants des femmes dans chaque groupe quinquennal, ${}_5P_x$

Bien que le nombre moyen d'enfants nés vivants des femmes dans chaque groupe quinquennal d'âge, ${}_5P_x$, ne soit nécessaire que pour les groupes 15-19, 20-24 et 25-29 ans, nous recommandons de calculer ces valeurs pour tous les groupes d'âge afin d'évaluer la qualité de données. La moyenne est calculée simplement en divisant le nombre d'enfants nés vivants par le nombre de femmes dans le groupe d'âge :

$${}_5P_x = \frac{{}_5CEB_x}{{}_5N_x}$$

Les résultats apparaissent dans la quatrième colonne du tableau 16.7. Les rapports de descendance (ou ratios de parité) sont ensuite calculés comme suit :

TABLEAU 16.6 Valeurs de $e(x,j)$, $f(x,j)$ et $g(x,j)$ pour estimer la date de référence $t(x)$ pour les probabilités de décéder avant les âges exacts de l'enfance à partir des proportions d'enfants décédés parmi les enfants nés vivants classées par durée de mariage de la mère

| | | Durée de mariage de la mère et valeur de n dans ${}_nq_0$ | | | | |
|---------------------|----------|---|----------|----------|----------|----------|
| Famille j | | 0-4 | 5-9 | 10-14 | 15-19 | 20-24 |
| | | 2 | 5 | 5 | 5 | 15 |
| Princeton « Nord » | $a(x,j)$ | 1,1980 | 1,2248 | 1,2076 | 1,2030 | 1,3292 |
| | $b(x,j)$ | - 0,1266 | - 0,1919 | - 0,0105 | 0,0896 | 0,1598 |
| | $c(x,j)$ | 0,0038 | - 0,0870 | - 0,2911 | - 0,4265 | - 0,5778 |
| Princeton « Sud » | $a(x,j)$ | 1,1705 | 1,3166 | 1,2952 | 1,2836 | 1,5269 |
| | $b(x,j)$ | - 0,1461 | - 0,3157 | - 0,0423 | 0,1308 | 0,2659 |
| | $c(x,j)$ | 0,0051 | - 0,0971 | - 0,4295 | - 0,6496 | - 0,9174 |
| Princeton « Est » | $a(x,j)$ | 1,2182 | 1,2769 | 1,2731 | 1,2585 | 1,3410 |
| | $b(x,j)$ | - 0,1809 | - 0,2268 | 0,0005 | 0,1216 | 0,1749 |
| | $c(x,j)$ | 0,0214 | - 0,1052 | - 0,3720 | - 0,5013 | - 0,5964 |
| Princeton « Ouest » | $a(x,j)$ | 1,2049 | 1,2573 | 1,2431 | 1,2469 | 1,4258 |
| | $b(x,j)$ | - 0,1553 | - 0,2266 | - 0,0230 | 0,0999 | 0,1948 |
| | $c(x,j)$ | 0,0135 | - 0,0944 | - 0,3409 | - 0,5267 | - 0,7454 |

$$\frac{{}_5P_{15}}{{}_5P_{20}} = \frac{0,2833}{1,5316} = 0,1850$$

et

$$\frac{{}_5P_{20}}{{}_5P_{25}} = \frac{1,5316}{2,8487} = 0,5376$$

Étape 3 : Choisir une famille de tables-types de mortalité

Une série de proportion d'enfants décédés par âge de la mère ne contient pas d'information sur la répartition par âge de la mortalité des jeunes enfants dans une population. Mais dans le cadre de l'analyse de la situation particulière d'un pays, il y aura inévitablement des informations pertinentes pour guider le choix de la table-type à retenir. L'information idéale est la répartition par âge de la mortalité des jeunes enfants tirée d'un histoire génésique complète pour la même population. La comparaison porte sur la relation entre ${}_1q_0$ et ${}_4q_1$ et elle fait généralement l'objet d'un graphique, où les points observés sont superposés aux courbes montrant la même relation à différents niveaux de mortalité dans chaque réseau de tables-types.

Plusieurs enquêtes collectant des histoires génésiques complètes ont été réalisées au Malawi. Sur la figure 16.1, on a représenté simultanément des estimations directes de la mortalité obtenues pour les 0-4 ans précédant les enquêtes EDS de 1992, 2000 et 2004 (et aussi les 5-9 ans précédant l'enquête de 2000) et des informations tirées des tables-types. (On notera que seuls trois réseaux des Nations Unies

sont représentés. Ce choix se justifie parce que la répartition par âge de la mortalité avant 5 ans est très peu différente entre les modèles Général, Asie du Sud et Extrême orient). Toutes les observations font apparaître des rapports entre ${}_4q_1$ et ${}_1q_0$ plus élevés que dans tous les réseaux de tables-types. Le meilleur choix dans le cas présent est sans doute le modèle 'Nord' de Princeton ; c'est lui que nous utiliserons par la suite.

Étape 4a(1) : Estimer l'âge moyen de la courbe de fécondité par âge

Nous utilisons le réseau « Nord » des tables-types de mortalité de Princeton, mais nous estimons tout de même l'âge moyen à la maternité à partir des taux de fécondité par âge, à des fins d'illustration. Le recensement de 2008 au Malawi incluait une question aux femmes d'âge fécond sur le nombre d'enfants qu'elles avaient eu l'année précédant le recensement. Le tableau 16.8 présente les données de base et le calcul de \bar{m} est ensuite calculé en rapportant le total de la colonne (vi) par le total de la colonne (iv), $= 29,817/1,0375 = 28,74$.

Étape 4a(2) : Estimer ${}_nq_0$ à partir de chaque ${}_5PD_x$

Chaque proportion ${}_5PD_x$ est ensuite convertie en un quotient ${}_nq_0$ estimé grâce aux coefficients appropriés tirés du tableau 16.3, comme il apparaît au tableau 16.9. Ainsi pour le groupe d'âge 35-39 ans on a :

$$\begin{aligned} {}_5q_0 &= 0,1597.(1,2046 + (0,3037 \times 0,1850) \\ &\quad + (-0,5656 \times 0,5376)) = 0,1528. \end{aligned}$$

TABLEAU 16.7 Enfants nés vivants et enfants survivants, recensement du Malawi 2008

| Groupe d'âge x , $x + 4$ | Nombre de femmes | Enfants nés vivants | Enfants survivants | Nombre moyen d'enfants nés vivants | Nombre moyen d'enfants survivants | Proportion décédés, ${}_5PD_x$ |
|-------------------------------|---------------------|------------------------|-----------------------|--|---|-----------------------------------|
| 15-19 | 635 927 | 180 178 | 161 541 | 0,2833 | 0,2540 | 0,1034 |
| 20-24 | 678 071 | 1 038 556 | 919 584 | 1,5316 | 1,3562 | 0,1145 |
| 25-29 | 566 350 | 1 613 374 | 1 398 776 | 2,8487 | 2,4698 | 0,1330 |
| 30-34 | 405 602 | 1 697 566 | 1 426 516 | 4,1853 | 3,5170 | 0,1597 |
| 35-39 | 298 004 | 1 553 676 | 1 266 514 | 5,2136 | 4,2500 | 0,1848 |
| 40-44 | 221 274 | 1 335 242 | 1 043 357 | 6,0343 | 4,7152 | 0,2186 |
| 45-49 | 174 875 | 1 128 423 | 851 048 | 6,4527 | 4,8666 | 0,2458 |

FIGURE 16.1 Estimations directes de $4q_1$ et $1q_0$ tirées des EDS menées au Malawi, et relations entre les quotients dans les tables-types de Princeton et des Nations Unies

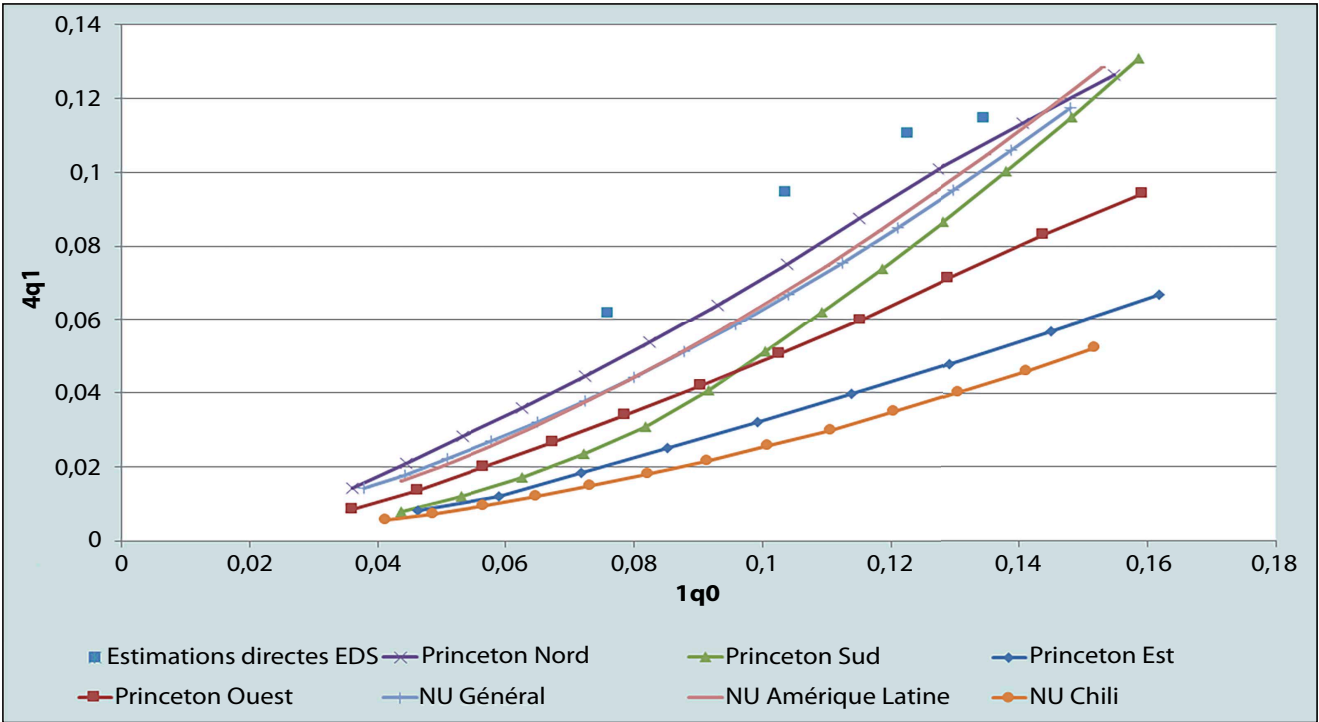


TABLEAU 16.8 Calcul de $t(x)$ à partir des données de naissances tirées du recensement de 2008 au Malawi

| Groupe d'âge x , $x + 4$ | Nombre de femmes | Naissances des 12 mois précédents | Taux de fécondité par âge, ${}_5f_x$ | Point médian du groupe d'âge $(x + 2)$ | ${}_5f_{x \cdot}(x + 2)$ |
|-------------------------------|------------------|---|---|---|--------------------------|
| 15-19 | 635 927 | 70 737 | 0,1112 | 17 | 1,891 |
| 20-24 | 678 071 | 169 406 | 0,2498 | 22 | 5,496 |
| 25-29 | 566 350 | 130 331 | 0,2301 | 27 | 6,213 |
| 30-34 | 405 602 | 79 232 | 0,1953 | 32 | 6,251 |
| 35-39 | 298 004 | 43 747 | 0,1468 | 37 | 5,432 |
| 40-44 | 221 274 | 13 956 | 0,0721 | 42 | 3,029 |
| 45-49 | 174 875 | 5 599 | 0,0320 | 47 | 1,505 |
| Total | | | 1,0375 | | 29,817 |

Étape 5 : Estimer la date de référence $t(x)$ de chaque quotient „ q_0 estimé

Pour chaque estimation, la date de référence $t(x)$ avant le recensement ou l'enquête est ensuite obtenue en utilisant les coefficients appropriés tirés du tableau 16.4, comme il appa-

raît au tableau 16.10. Ainsi pour le groupe d'âge 30-34 ans, on a :

$$t(30) = 2,0779 + (-1,7908) \times 0,1850 + 9,4126 \times 0,5377 = 6,81.$$

Le recensement a eu lieu entre le 8 et le 28 juin 2008 ; la date de référence est donc obtenue en retranchant t de 2008,46 (expression décimale du 18 juin 2008). Les résultats apparaissent dans la dernière colonne du tableau 16.10.

Étape 6 : Convertir chaque estimation de ${}_nq_0$ en une estimation de ${}_5\hat{q}_0$

La dernière étape consiste à convertir chaque quotient ${}_nq_0$ estimé en une estimation de l'indice courant ${}_5\hat{q}_0$. Ceci permettra de comparer les estimations entre les groupes d'âge. Chaque ${}_nq_0$ est converti en son logit $Y(n)$ au moyen de l'identité $Y(n) = 0,5 \cdot \ln({}_nq_0 / (1 - {}_nq_0))$. La valeur de α est ensuite obtenue en retranchant de $Y(n)$ le logit standard $Y^s(n)$, tiré de la table-type de mortalité Nord, sexes réunis, ayant une espérance de vie à la naissance de 60 ans. Chaque α est ensuite utilisé avec le standard $Y^s(5)$ de sorte à estimer ${}_5\hat{q}_0$. Ainsi pour le groupe d'âge 25-29 ans, on

a $Y(3) = 0,5 \ln(0,1222 / (1 - 0,1222)) = -0,9857$, et $\alpha = -0,9857 - (-1,1664) = 0,1806$. Donc,

$${}_5\hat{q}_0 = \frac{e^{2(0,1806 + (-1,0900))}}{1 + e^{2(0,1806 + (-1,0900))}} = 0,1396.$$

Des estimations de ${}_1\hat{q}_0$ sont tirées de façon analogue en utilisant le logit standard pour 1 an.

Sur la figure 16.2 nous avons représenté chaque estimation ${}_1\hat{q}_0$ et ${}_5\hat{q}_0$ à la date de référence correspondante. Le graphique indique que la mortalité avant 5 ans a baissé, de 185 pour 1 000 environ au début des années 1990 à 140 environ en 2005. La légère hausse apparente de la mortalité des jeunes enfants en 2007 ne doit pas être prise en compte, du fait de la surestimation vraisemblable de la mortalité à partir des déclarations des jeunes mères, comme on l'a dit plus haut.

| Groupe d'âge | Proportion d'enfants décédés | Coefficients de régression pour ${}_nq_0$ (Princeton Modèle « Nord ») | | | ${}_nq_0$ |
|--------------|------------------------------|---|---------|---------|-----------|
| | | $a(i)$ | $b(i)$ | $c(i)$ | |
| 15-19 | 0,1034 | 1,1119 | -2,9287 | 0,8507 | 0,1063 |
| 20-24 | 0,1146 | 1,2390 | -0,6865 | -0,2745 | 0,1105 |
| 25-29 | 0,1330 | 1,1884 | 0,0421 | -0,5156 | 0,1222 |
| 30-34 | 0,1597 | 1,2046 | 0,3037 | -0,5656 | 0,1528 |
| 35-39 | 0,1848 | 1,2586 | 0,4236 | -0,5898 | 0,1885 |
| 40-44 | 0,2186 | 1,2240 | 0,4222 | -0,5456 | 0,2205 |
| 45-49 | 0,2458 | 1,1772 | 0,3486 | -0,4624 | 0,2441 |

TABLEAU 16.9 Estimation du quotient ${}_nq_0$ pour chaque proportion ${}_5PD_x$

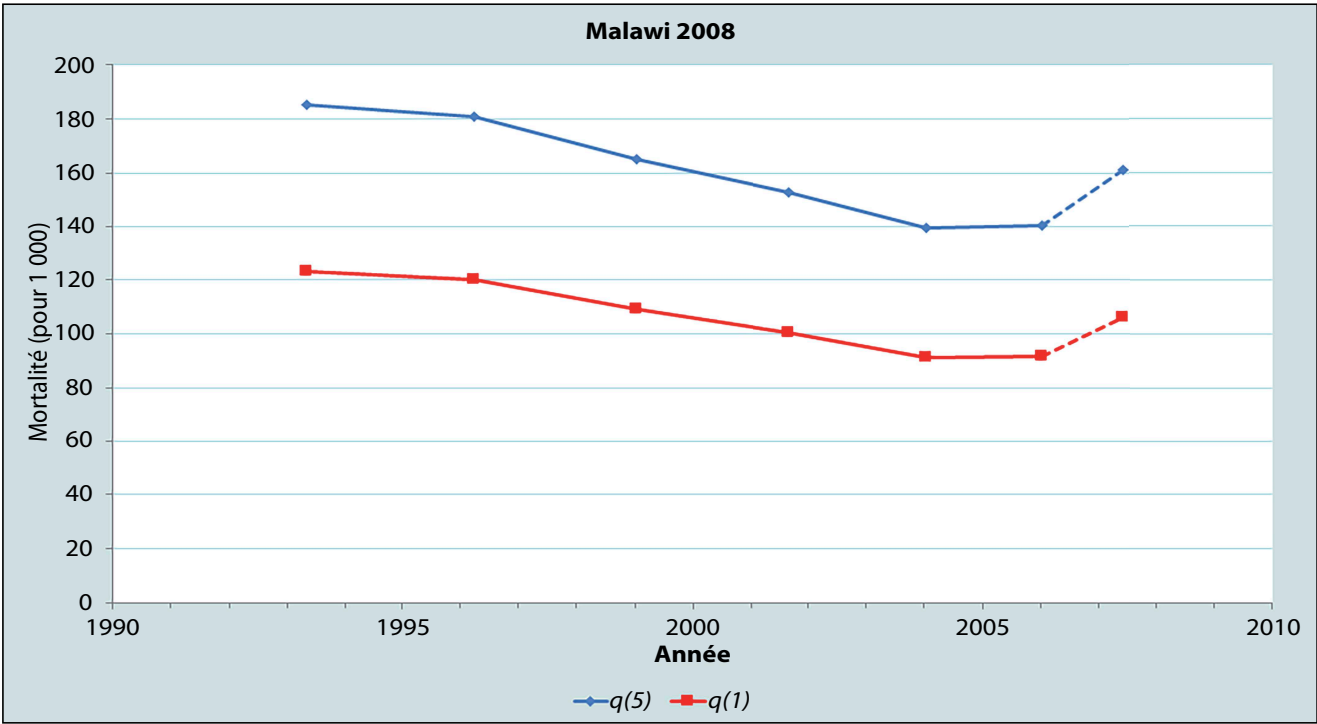
| Groupe d'âge | Coefficients de régression pour la date: Princeton modèle « Nord » | | | Date t | Date de référence (2008,46- t) |
|--------------|--|----------|---------|----------|-----------------------------------|
| | $e(i)$ | $f(i)$ | $g(i)$ | | |
| 15-19 | 1,0921 | 5,4732 | -1,9672 | 1,05 | 2007,42 |
| 20-24 | 1,3207 | 5,3751 | 0,2133 | 2,43 | 2006,03 |
| 25-29 | 1,5996 | 2,6268 | 4,3701 | 4,44 | 2004,03 |
| 30-34 | 2,0779 | -1,7908 | 9,4126 | 6,81 | 2001,66 |
| 35-39 | 2,7705 | -7,3403 | 14,9352 | 9,44 | 1999,02 |
| 40-44 | 4,1520 | -12,2448 | 19,2349 | 12,23 | 1996,24 |
| 45-49 | 6,9650 | -13,9160 | 19,9542 | 15,12 | 1993,35 |

TABLEAU 16.10 Estimation de la date de référence $t(x)$ pour chaque estimation, en années avant le recensement

TABLEAU 16.11 Estimation du paramètre α de la table de mortalité logit pour chaque estimation, et déduction d'une série de ${}_1\hat{q}_0$ et ${}_5\hat{q}_0$

| Groupe d'âge | ${}_nq_0$ | n | logit $Y(n)$ | Standard logit $Y^s(n)$ | α | ${}_1\hat{q}_0$ | ${}_5\hat{q}_0$ |
|--------------|-----------|-----|--------------|-------------------------|----------|-----------------|-----------------|
| 15-19 | 0,1063 | 1 | -1,0647 | -1,3300 | 0,2653 | 0,1063 | 0,1612 |
| 20-24 | 0,1105 | 2 | -1,0431 | -1,2273 | 0,1842 | 0,0918 | 0,1405 |
| 25-29 | 0,1222 | 3 | -0,9857 | -1,1664 | 0,1806 | 0,0912 | 0,1396 |
| 30-34 | 0,1528 | 5 | -0,8566 | -1,0900 | 0,2334 | 0,1004 | 0,1528 |
| 35-39 | 0,1885 | 10 | -0,7299 | -1,0091 | 0,2791 | 0,1089 | 0,1650 |
| 40-44 | 0,2205 | 15 | -0,6313 | -0,9664 | 0,3350 | 0,1203 | 0,1809 |
| 45-49 | 0,2441 | 20 | -0,5652 | -0,9138 | 0,3487 | 0,1232 | 0,1850 |

FIGURE 16.2 Estimation de l'évolution de la mortalité avant 5 ans et avant 1 an, Malawi, recensement de 2008



DIAGNOSTICS, ANALYSE ET INTERPRÉTATION

Contrôles et validation

Indépendamment du mode de collecte des données et de ce qu'on peut savoir sur la formation et la supervision des enquêteurs, un examen méticuleux de la qualité des données est une première étape essentielle de toute analyse. Toutes les séries de données contiennent des erreurs, qui peuvent

provenir de plusieurs sources, tels qu'un enquêteur prenant des raccourcis ou une interviewée ne connaissant la réponse correcte à une question, etc.

Un aspect propre à l'histoire génésique résumée doit être souligné pour le contrôle des données : le non-respect de l'hypothèse 2, selon laquelle les risques de mortalité des enfants au cours d'une période ne varient pas avec l'âge de la mère. Dans de nombreuses applications, il apparaît clai-

rement que l'hypothèse ne tient pas. Les risques de décès pour les enfants de mères âgées de 15-19 ans (et l'estimation indirecte de la mortalité des jeunes enfants tirée du nombre d'enfants nés vivants et du nombre d'enfants décédés) sont fréquemment plus élevés, parfois même très nettement, que la moyenne. C'est également vrai dans une moindre mesure pour les enfants des mères de 20-24 ans. Deux facteurs contribuent à ces distorsions : la distribution des enfants par rang de naissance et des facteurs socioéconomiques. Il est bien connu que les premiers nés courent des risques de décéder plus élevés que les naissances de rang plus élevé, et les enfants nés de mères jeunes sont des premiers nés dans une proportion supérieure à la moyenne. Les femmes qui ont des enfants de façon précoce ont également tendance à appartenir à des groupes socioéconomiques moins favorisés et leurs enfants sont ainsi exposés à des risques de mortalité supérieurs à la moyenne. En conséquence, les estimations de mortalité tirées des déclarations des femmes âgées de 15-19 ans doivent être traitées avec beaucoup de circonspection, ou laissées de côté.

Interprétation

Deux caractéristiques essentielles de la méthode doivent être présentes à l'esprit quand on interprète les résultats. D'abord, il n'y a pas d'information sur les dates et les âges dans les données de base. La seule information relative au nombre d'enfants nés vivants d'une femme est que ces naissances sont survenues entre le moment où elle a eu sa première naissance et son âge au moment de l'enquête. Le nombre d'enfants décédés nous en dit encore moins sur l'âge où ces décès ont pu se produire, car l'éventail des âges possibles dépend de la distribution des naissances dans le temps. Il est donc impossible de tirer des conclusions sur les fluctuations à court terme de la mortalité des jeunes enfants à partir d'une histoire génésique résumée. Les déclarations de deux femmes du même âge (ou de la même durée de mariage, ou de la même durée depuis la première naissance) ayant les mêmes nombres d'enfants déjà nés et décédés peuvent refléter des conditions de mortalité différentes. Ce qu'une histoire génésique résumée peut offrir de mieux est une indication générale sur la tendance passée en moyenne. Et même cette tendance moyenne doit être interprétée prudemment pour le passé récent car des biais de sélection affectent les déclarations des femmes âgées de 15-19 ans, et celles âgées de 20-24 ans, quoique dans une moindre mesure.

La seconde caractéristique est que l'information n'est fournie que pour des femmes survivantes résidant encore dans des ménages enquêtés, ce qui risque de représenter un ensemble sélectionné, donc biaisé. La mortalité des enfants nés dans un groupe et dont les mères ne vivent plus dans le groupe ne sera pas incluse dans les mesures. Si ces enfants ont une mortalité plus élevée que ceux nés de mères vivant encore dans le groupe, la mortalité sera sous-estimée. La forme la plus sévère de ce biais peut survenir du fait d'une prévalence élevée du VIH dans le groupe, car en l'absence d'une thérapie rétrovirale largement diffusée, cette prévalence entraînera une forte corrélation positive entre la survie de l'enfant et la survie de la mère (voir la section sur les effets du VIH sur les méthodes d'estimation de la mortalité des jeunes enfants (chapitre 15)). Mais plus généralement, une certaine corrélation positive entre la survie de la mère et celle de l'enfant est presque systématique dans toute population.

Il peut y avoir d'autres sources de biais de sélection des répondantes. Par exemple, des taux élevés d'immigration impliqueront que des femmes répondent aux questions sur la survie d'enfants nés et élevés ailleurs, alors qu'une forte émigration exclura des réponses sur des enfants qui étaient nés et ont été élevés dans le groupe. Bien qu'il soit impossible de connaître *a priori* la direction et l'ampleur de tels biais, l'analyste doit avoir leur effet potentiel présent à l'esprit. Les non-réponses sont un problème moins important pour les histoires génésiques résumées que pour les histoires génésiques complètes, car l'information est souvent fournie par des tierces personnes, pas nécessairement par la femme elle-même. Ainsi une histoire génésique résumée peut être établie pour une femme absente du groupe pour un long voyage, alors qu'une histoire génésique complète ne sera pas recueillie si la femme ne peut pas être interviewée en personne.

DESCRIPTION DÉTAILLÉE DE LA MÉTHODE

L'idée que les proportions d'enfants décédés parmi les enfants nés vivants sont des indicateurs de la mortalité des jeunes enfants a une longue histoire. Des questions sur les enfants nés vivants et les enfants survivants étaient déjà intégrées dans le recensement de 1900 aux États-Unis (Preston et Haines 1991), le recensement de Grande Bretagne en 1911 et celui du Brésil en 1940, entre autres. Mais la première méthode pour traduire ces proportions en indicateurs d'une table de mortalité classique a été proposée par Brass et Coale (1968).

Pour illustrer l'idée de base, prenez l'exemple simple (et irréaliste) d'une population dans laquelle toutes les femmes ont exactement un enfant, nés quand elles avaient exactement 25 ans, où toutes les femmes survivent de 25 à 30 ans, et où il n'y a pas de migration. Dans une enquête, la proportion d'enfants décédés parmi les enfants nés vivants parmi les femmes âgées de 30 ans exactement mesurerait précisément la probabilité que la cohorte d'enfants décède entre la naissance et le 5^e anniversaire, ${}_5q_0$. Dans une autre population où toutes les femmes auraient aussi un enfant exactement, mais à 27 ans, la proportion d'enfants décédés mesurerait précisément la probabilité que la cohorte décède avant 3 ans, ${}_3q_0$.

Ces deux exemples illustrent plusieurs points importants. Premièrement, l'âge des femmes est une approximation de l'exposition au risque de leurs enfants. Toutes choses égales par ailleurs, plus une mère est âgée, plus est longue en moyenne la période d'exposition au risque de décéder de ses enfants. Deuxièmement, l'interprétation d'une proportion d'enfants décédés en termes de mesure classique de la table de mortalité dépend de l'âge à la maternité. Troisièmement, l'équivalence entre une proportion d'enfants décédés et une mesure tirée d'une table de mortalité nécessite une absence d'effets de sélection par la mortalité ou l'émigration, ou d'effets de contamination par l'immigration. Quatrièmement, les mesures obtenues sont relatives à des cohortes (ou des moyennes sur plusieurs cohortes) plutôt qu'à des périodes.

Évidemment dans les populations réelles les enfants naissent de mères de divers âges et sont exposés à des risques de mortalité par âge qui peuvent varier au fil du temps. Les méthodes d'estimation utilisent des schéma-types de fécondité et de mortalité des jeunes enfants par âge pour modéliser les proportions d'enfants décédés parmi les enfants nés vivants, qui peuvent ensuite être rattachées aux paramètres sous-jacents de tables de mortalité. Un trait commun aux données sur la survie des enfants classés par groupe d'âge de la mère est que les proportions d'enfants décédés sont plus élevées pour les femmes âgées de 15-19 et 20-24 ans que pour celles des groupes d'âge suivants, bien qu'elles reflètent des durées moyennes plus courtes d'exposition des enfants. C'est parce que les jeunes femmes qui ont des enfants ont généralement un statut socioéconomique inférieur à la moyenne et que leurs enfants sont, plus souvent que la moyenne, des premiers nés, deux facteurs connus comme exposant les jeunes enfants à des

risques de mortalité élevés. La mortalité des enfants nés de ces jeunes mères n'est donc pas représentative de la mortalité de l'ensemble des enfants nés dans cette population. En partie pour éviter ce biais, des méthodes ont été développées où les femmes sont classées par durée de mariage (Sullivan 1972) et durée depuis la première naissance (Hill and Figueroa 2001). Ces méthodes sont aussi moins affectées par les variations de la fécondité.

La méthode a été initialement développée par Brass sans considération explicite des effets des variations de la mortalité sur les estimations, même s'il note qu'en cas de telles variations « les estimations de $q(2)$ et $q(3)$ seraient représentatives de la mortalité moyenne sur une courte période (moins d'une décennie) avant le recensement ou l'enquête » (Brass and Coale, 1968 : 116). Il est clair aujourd'hui que la mortalité des jeunes enfants a diminué dans le monde, et très rapidement dans certaines populations. À la suite du travail pionnier de Feeney (1976, 1980), des méthodes ont été développées pour estimer la « date de référence » des estimations tirées de chaque groupe d'âge ou de durée (Coale and Trussell (1977) ; Palloni and Heligman (1985) ; Hill and Figueroa (2001)). La proportion d'enfants décédés parmi les enfants nés vivants d'un groupe de femmes représente une moyenne des risques de mortalité sur l'ensemble des cohortes de naissance de ces enfants. Plus les femmes sont âgées, ou plus longue est leur exposition au risque, plus les cohortes d'enfants s'étirent dans le passé et plus s'éloigne dans le temps la période de référence couverte par l'estimation de la mortalité des jeunes enfants tirée de la proportion d'enfants décédés. Comme les risques de mortalité des jeunes enfants sont fortement concentrés sur les très jeunes âges, la date de référence n'est pas très différente en pratique du nombre moyen d'années écoulées depuis la naissance des enfants, si l'évolution de la mortalité a été relativement stable au fil du temps. Les dates de référence exactes pour les estimations de mortalité des jeunes enfants doivent en fait être conçues comme des points centraux du temps avec une distribution des décès autour d'eux. En conséquence, les estimations de mortalité des jeunes enfants tirées des histoires génésiques résumées et les variantes des méthodes indirectes de Brass ne peuvent pas servir à identifier des évolutions ou des crises de la mortalité situées en un point précis du temps. La méthode fournit une bonne description des tendances générales de la mortalité des jeunes enfants, mais lissées par rapport aux fluctuations réelles d'une année sur l'autre

qu'on observe dans presque chaque population. D'autres méthodes de mesure (comme les histoires génésiques complètes de naissances) sont nécessaires pour estimer l'impact dans le temps qu'ont pu avoir des interventions visant à améliorer la santé des enfants.

EXPOSÉ MATHÉMATIQUE (MÉTHODES DU MANUEL X ET DES TABLES-TYPES DE MORTALITÉ DES NATIONS UNIES)

La proportion d'enfants décédés parmi les enfants déjà nés des femmes d'âge exact x , $PD(x)$, est une moyenne des probabilités de décéder des cohortes pondérées par les naissances,

$$PD(x) = \frac{\int_{\alpha}^x f(y) {}_{x-y}q_0^c dy}{\int_{\alpha}^x f(y) dy}$$

où $f(y)$ est le taux de fécondité à l'âge y , ${}_{x-y}q_0^c$ est la probabilité de décéder avant l'âge $(x-y)$ pour la cohorte née $(x-y)$ années plus tôt, et α est l'âge du début de la période féconde. L'expression est identique pour la durée de mariage et la durée depuis la première naissance, à ceci près que α devient 0. Les proportions d'enfants décédés pour des groupes quinquennaux d'âge ou de durée peuvent ensuite être estimées en faisant la moyenne des estimations ponctuelle de PD sur l'ensemble du groupe quinquennal, moyenne pondérée par les distributions supposées de population Nx à chaque x . On suppose le plus souvent que la population sous-jacente peut être considérée comme stable, avec un taux de croissance compatible avec les paramètres démographiques qui sous-tendent les $PD(x)$, et que la table de mortalité aux âges adultes est la même que celle utilisée dans le calcul des $PD(x)$. Ces calculs utilisent des groupes d'âge discrets, soit par exemple, en utilisant des années d'âge,

$${}_5PD_x = \frac{\sum_{a=x}^{x+5} PD(a) \cdot {}_1L_a \cdot e^{-r(a+0,5)}}{\sum_{a=x}^{x+5} {}_1L_a \cdot e^{-r(a+0,5)}}.$$

Des schémas-types de f , q et N sont utilisés pour modéliser les valeurs de ${}_5PD_x$, qui sont rapportées aux valeurs appropriées de q par une analyse de régression utilisant les rapports de descendance comme variables indépendantes (voir les équations d'estimation plus haut).

EXTENSIONS DE LA MÉTHODE

Variantes par durée de mariage et durée depuis la première naissance

Comme on l'a dit plus haut, sans l'illustrer en détail, des variantes de la méthode originale ont été développées, classant les femmes par durée de mariage (Sullivan 1972) et durée depuis la première naissance (Hill et Figueroa 2001). Ces méthodes ont été mises au point pour éviter deux sources potentielles d'erreur dans la méthode fondée sur l'âge : les effets des variations de la fécondité (qui faussent les rapports de descendance) et la surmortalité des enfants nés de jeunes mères.

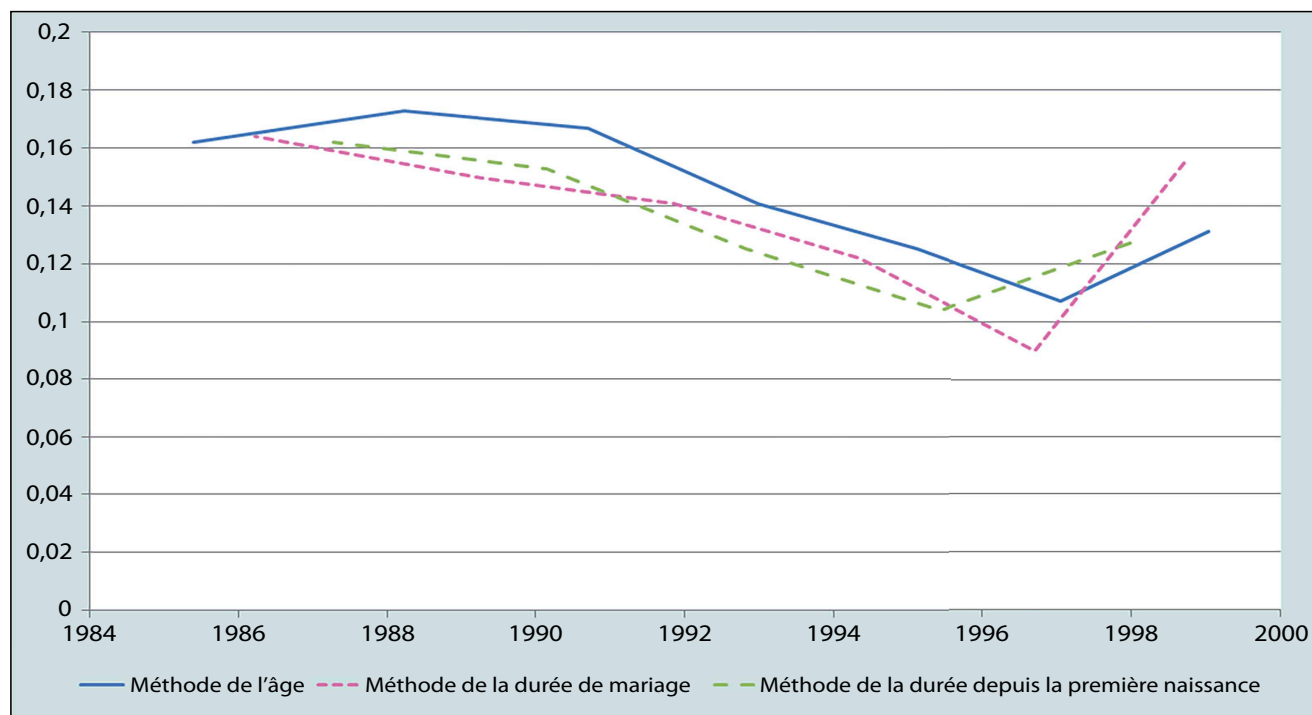
Ces variantes sont illustrées plus bas, mais ces deux améliorations ont elles-mêmes leurs limites. Premièrement, rares sont les recensements de population recueillant l'information nécessaire pour classer les femmes et leurs enfants par durée de mariage ou par durée depuis la première naissance. Deuxièmement, concernant la durée de mariage, dans de nombreux pays en développement le mariage n'est pas un préalable nécessaire au début de la vie féconde. En outre, les enfants de mères non mariées ont souvent une mortalité supérieure à celle des enfants de mère mariée. Il peut donc aussi y avoir des distorsions dans les résultats de ces variantes. Quand de telles données sont recueillies, comme dans de nombreux pays arabes où les proportions de naissances hors mariage sont faibles, elles permettent d'apprécier de façon intéressante les mérites des différentes démarches.

Certaines enquêtes, comme les MICS (Enquêtes à indicateurs multiples, conduites sous les auspices de l'Unicef) réalisées avant 2010, ont recueilli des histoires génésiques résumées et les informations nécessaires sur la date du mariage et de la première naissance. La question se pose donc de la méthodologie qu'il faut préférer.

Pour répondre, nous avons utilisé les histoires génésiques complètes recueillies lors de l'EDS 1999-2000 au Bangladesh. Les données ont été présentées dans les trois formats et analysées en utilisant le réseau « Sud » des tables-types de mortalité de Princeton. La figure 16.3 donne les estimations de ${}_5q_0$ pour chaque variante. (On notera que dans le cas étudié les données ne font pas apparaître une forte surmortalité quand l'estimation s'appuie sur le groupe d'âge le plus jeune, 15-19 ans).

Hormis pour le point le plus récent dans chaque série (soit 15-19 ans pour l'âge et 0-4 ans à la fois pour la durée de mariage et la durée depuis la première naissance), les

FIGURE 16.3 Estimations de la mortalité avant 5 ans tirées des proportions d'enfants décédés parmi les enfants nés vivants classés par groupe d'âge, groupe de durée de mariage et groupe de durée depuis la première naissance, Bangladesh EDS 1999-2000



évolutions générales retracées par les trois estimations sont très semblables. Dans les méthodes par durée de mariage et durée depuis la première naissance, les niveaux sont très proches, mais à partir du groupe d'âge 25-29 ans les estimations par l'âge sont systématiquement les plus élevées, supérieures de 20 pour 1 000 en moyenne (à 155 pour 1 000) à chacune des deux autres méthodes. Des moyennes simples des estimations pour toutes les durées de mariage ou depuis la première naissance, excepté la plus récente, donnent un résultat presque semblable : 133 pour la durée de mariage et 136 pour la durée depuis la première naissance, bien que les périodes couvertes soient légèrement différentes.

Plusieurs points sont intéressants dans cette application. Premièrement, les méthodes de durée de mariage et durée depuis la première naissance ont été développées en partie pour contourner le problème du biais de sélection dans les points fondés sur l'âge pour les femmes de 15-19 et 20-24 ans. Mais dans la présente application, ces deux méthodes font apparaître des sauts pour le point le plus récent au moins aussi importants que pour les estimations fondées sur l'âge.

Ceci pourrait s'expliquer si le biais a essentiellement pour origine le poids des premiers nés, lesquels sont tous (durée depuis la première naissance) ou presque tous (durée de mariage) concentrés dans la première catégorie, alors qu'ils sont davantage répartis entre les groupes d'âge.

Une seconde question est l'utilisation des rapports de descendance observés au moment de l'enquête pour estimer la répartition dans le temps des naissances passées. Comme on l'a vu plus haut, si la fécondité baisse (même si le calendrier de la fécondité ne se modifie pas), les descendances des jeunes femmes seront relativement basses et celles des femmes âgées seront relativement élevées, ce qui réduira les rapports des descendances au dessous de leur valeur dans les cohortes réelles. La fécondité apparaîtra ainsi plus tardive que dans la réalité (et donc la durée d'exposition des jeunes enfants au risque de décéder plus courte), ce qui conduira à surestimer la mortalité des jeunes enfants.

Le Bangladesh ayant connu une forte baisse de la fécondité entre le milieu des années 1980 et l'EDS de 1999-2000, il est intéressant d'essayer d'évaluer l'ampleur du

biais. Comme les données proviennent d’une histoire génésique complète, nous pouvons calculer les descendance des cohortes il y a 5 ans, 10 ans, 15 ans, etc., en retranchant les naissances récentes du nombre des enfants nés vivants, et calculer ensuite les rapports de descendance pour des cohortes réelles. Nous utilisons la méthodologie classique (en calculant des rapports $P(1)/P(2)$ et $P(2)/P(3)$) pour les cohortes qui ont atteint le troisième groupe d’âge ou de durée, c’est-à-dire les groupes d’âge 25-29 ans et plus et les groupes de durée 10-14 ans et plus.

La première partie du tableau 16.12 compare les rapports de descendance au moment de l’enquête et dans les cohortes. Le résultat est clair et conforme aux attentes pour l’âge, les rapports au moment de l’enquête étant nettement inférieurs aux rapports dans toutes les cohortes, lesquels sont relativement stables de cohorte en cohorte. Mais c’est moins clair pour les méthodes par durée de mariage et durée depuis la première naissance : les rapports $P(2)/P(3)$ sont généralement plus élevés que ceux au moment de l’enquête, alors que les rapports $P(1)/P(2)$ sont tous plus bas.

Pour apprécier l’importance de cet effet sur les estimations, nous pouvons utiliser les rapports des descendance pour chaque proportion d’enfants décédés parmi les enfants nés vivants (pour les groupes d’âge 25-29 ans et plus et les

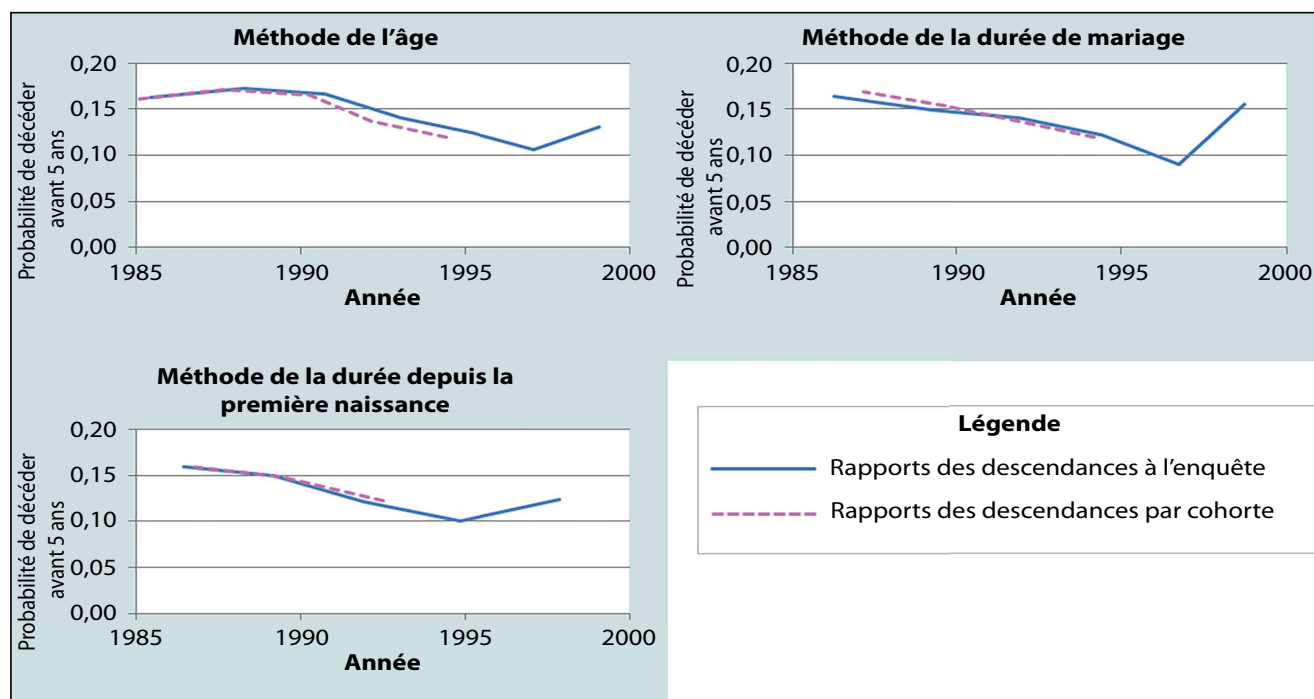
groupes de durée 10-14 ans et plus). Les trois graphiques de la figure 16.4 représentent les estimations originales et celles dans les cohortes localisées à la date de référence, pour les groupes d’âge, les groupes de durée de mariage et les groupes de durée depuis la première naissance respectivement. Le recours aux rapports de descendance dans les cohortes se traduit par des estimations moins élevées aux groupes d’âges 25-29 et 30-34 ans, mais il a peu d’effet sur les estimations obtenues par les deux autres méthodes. C’est sans doute parce que la baisse de la fécondité au Bangladesh pendant cette période était essentiellement due à une hausse de l’âge à la maternité, avec peu d’impact sur la durée du mariage ou la durée depuis la première naissance. Au total, le recours aux rapports de descendance dans les cohortes améliore la cohérence entre méthodes, mais le prix à payer est plutôt élevé : la perte des estimations les plus récentes.

L’idée générale est que la méthode fondée sur l’âge tend à surestimer la mortalité des jeunes enfants si la fécondité évolue rapidement, alors que les effets sur les deux autres méthodes sont faibles. Si les données existent, il est recommandé d’utiliser une des deux autres variantes plutôt que la méthode fondée sur l’âge, mais cela ne vaut sans doute pas la peine d’ajouter une question supplémentaire dans un recensement juste pour obtenir l’information nécessaire.

TABLEAU 16.12 Rapports des descendance $P(1)/P(2)$ et $P(2)/P(3)$ calculés au moment de l’enquête et dans les vraies cohortes, Bangladesh, 1999-2000

| Méthode de l'âge | | | | | | |
|--|---------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| Rapport des descendances | Au moment de l'enquête | Cohorte âgée de 25-29 | Cohorte âgée de 30-34 | Cohorte âgée de 35-39 | Cohorte âgée de 40-44 | Cohorte âgée de 45-49 |
| $P(1)/P(2)$ | 0,269 | 0,342 | 0,311 | 0,327 | 0,361 | 0,336 |
| $P(2)/P(3)$ | 0,544 | 0,66 | 0,639 | 0,603 | 0,632 | 0,619 |
| Méthode de la durée de mariage | | | | | | |
| Rapport des descendances | Au moment de l'enquête | Cohorte 10-14 | Cohorte 15-19 | Cohorte 20-24 | Cohorte 25-29 | |
| $P(1)/P(2)$ | 0,353 | 0,352 | 0,288 | 0,274 | 0,287 | |
| $P(2)/P(3)$ | 0,635 | 0,682 | 0,65 | 0,598 | 0,603 | |
| Méthode de la durée depuis la première naissance | | | | | | |
| Rapport des descendances | Au moment de l'enquête | Cohorte 10-14 | Cohorte 15-19 | Cohorte 20-24 | | |
| $P(1)/P(2)$ | 0,553 | 0,544 | 0,495 | 0,505 | | |
| $P(2)/P(3)$ | 0,709 | 0,775 | 0,753 | 0,723 | | |

FIGURE 16.4 Estimations indirectes de la mortalité avant 5 ans obtenues en utilisant les rapports de descendance observés à l'enquête ou ceux dans des cohortes réelles, Bangladesh, EDS 1999-2000



AUTRES LECTURES ET RÉFÉRENCES

L'estimation indirecte de la mortalité des jeunes enfants est présentée dans tous les manuels classiques d'estimation indirecte (Sloggett, Brass, Eldridge *et al.* 1994 ; Division de la Population des Nations Unies 1984).

Brass W and AJ Coale. 1968. "Methods of analysis and estimation," in Brass, W, AJ Coale, P Demeny, DF Heisel, *et al.* (eds). *The Demography of Tropical Africa*. Princeton NJ: Princeton University Press, pp. 88-139.

Brass W, AJ Coale, P Demeny, DF Heisel *et al.* (eds). 1968. *The Demography of Tropical Africa*. Princeton NJ: Princeton University Press.

Coale AJ and J Trussell. 1977. "Estimating the time to which Brass estimates apply; Annex I to Preston SH and Palloni A "Fine-tuning Brass-type mortality estimates with data on ages of surviving children", *Population Bulletin of the United Nations* **10**:87-89.

Collumbien M and A Sloggett. 2001. "Adjustment methods for bias in the indirect childhood mortality estimates," in Zaba, B and J Blacker (eds). *Brass Tacks: Essays in Medical Demography*. London: Athlone, pp. 20-42.

Division de la Population des Nations Unies. 1984. *Manuel X. Techniques indirectes d'estimation démographique*. New York : Nations Unies, Département des affaires économiques et sociales internationales, ST/ESA/SER.A/81. http://unstats.un.org/unsd/demographic/standmeth/handbooks/Manuel_X-fr.pdf

Feeney G. 1976. "Estimating infant mortality rates from child survivorship data by age of mother", *Asian and Pacific Census Newsletter* **3**(2):12-16. <http://hdl.handle.net/10125/3556>.

Feeney G. 1980. "Estimating infant mortality trends from child survivorship data", *Population Studies* **34**(1):109-128. doi: <http://dx.doi.org/10.1080/00324728.1980.10412839>

Hill K and M-E Figueroa. 2001. "Child mortality estimation by time since first birth," in Zaba, B and J Blacker (eds). *Brass Tacks: Essays in Medical Demography*. London: Athlone, pp. 9-19.

Palloni A and L Heligman. 1985. "Re-estimation of structural parameters to obtain estimates of mortality in developing countries", *Population Bulletin of the United Nations* **18**:10-33.

Preston SH and MR Haines. 1991. *Fatal Years: Child Mortality in Late Nineteenth-century America*. Princeton, NJ: Princeton University Press.

Sloggett A, W Brass, SM Eldridge, IM Timæus, P Ward and B Zaba. 1994. *Estimation of Demographic Parameters from Census Data*. Tokyo, Japan: United Nations Statistical Institute for Asia and the Pacific.

Sullivan JM. 1972. "Models for the estimation of the probability of dying between birth and exact ages of early childhood", *Population Studies* **26**(1):79-97. doi: <http://dx.doi.org/10.1080/00324728.1972.10405204>

UN Population Division. 1991. *Child Mortality in Developing Countries*. New York: United Nations, Department of Economic and Social Affairs, ST/ESA/SER.A/123.

Méthodes fondées
sur des données d'enquêtes

Chapter 17 ESTIMATION DIRECTE DE LA MORTALITÉ DES JEUNES ENFANTS À PARTIR D'HISTOIRES GÉNÉSIQUES

Kenneth Hill

Dans le présent chapitre nous traitons de l'utilisation des données d'histoires génésiques complètes (*en anglais : full birth histories, FBH*) ou tronquées (*en anglais : truncated birth histories, TBH*) pour estimer la mortalité des jeunes enfants. La caractéristique essentielle de ces données est que pour chaque naissance incluse, la date de naissance, la survie et (en cas de décès) la date du décès sont enregistrés. L'analyse des données s'appuie sur la construction de tables de mortalité. L'estimation indirecte de la mortalité des jeunes enfants, et l'estimation de la mortalité des jeunes enfants à partir de la survie d'une naissance récente, sont traitées dans les chapitres 16 et 18, respectivement.

DONNÉES NÉCESSAIRES ET HYPOTHÈSES

Données nécessaires

Pour chaque femme d'âge fécond (dans certains contextes pour des raisons culturelles, la collecte d'information ne porte que sur les femmes déjà mariées) :

- le nom de chaque enfant né vivant ;
- le mois et l'année de naissance de chaque enfant ;
- le sexe de l'enfant (facultatif) ;
- si l'enfant est encore en vie ; et
- si l'enfant est décédé, son âge au décès (dans les EDS, l'âge au décès est recueilli en jours pour les décès dans les 28 premiers jours, en mois pour les décès entre un et 23 mois, et en années au-delà).

Hypothèses importantes

- Les enfants encore en vie et les enfants décédés sont déclarés avec la même précision.
- Les dates de naissance et de décès sont déclarées avec une précision raisonnable.

- Il n'y a pas de corrélation entre les risques de mortalité des enfants et les taux de survie des mères (que ce soit du fait de la mortalité ou de la migration) dans la population.

PRÉCAUTIONS ET MISES EN GARDE

Le travail sur des données collectées directement présente des dangers provenant de deux sources. Le premier est le risque de biais de survie, puisque seules les femmes vivantes sont interrogées sur leur histoire génésique détaillée à l'origine des données. Dans les situations où il se pourrait que les mères décédées aient eu une fécondité différente de celles des femmes survivantes, ou que leurs enfants aient fait face à des risques de décéder différents, il existe un risque de biais important dans les estimations obtenues. Divers aspects du biais de survie sont présentés dans chapitre 15 dans la section traitant des effets du VIH sur l'estimation de la mortalité des jeunes enfants.

Le second danger provient du biais de troncature, si une limite d'âge supérieure s'applique aux femmes dont on recueille les histoires génésiques détaillées ; ce biais est d'autant plus significatif qu'on remonte plus loin dans le temps. Si on applique une limite d'âge de 49 ans lors de la collecte des données, l'information disponible pour la période 10 ans avant l'enquête ne porte que sur les femmes alors âgées de moins de 40 ans. La mortalité des jeunes enfants estimée à partir de telles données pour des périodes de plus en plus anciennes reposera progressivement sur les histoires génésiques de femmes de plus en plus jeunes. Il peut en résulter un biais de mesure, car cette troncature entraîne une surreprésentation des premières naissances parmi les jeunes femmes ; la mortalité des jeunes enfants ainsi mesurée risque donc d'être d'autant plus surestimée que la période est ancienne. Certains résultats montrent que cette surestimation est contrebalancée par une sous-estimation due aux erreurs de mémoire (et à l'omission sélective des enfants qui sont décédés dans un passé lointain).

ÉVALUATION DE LA QUALITÉ DES DONNÉES ET ANALYSE DES DONNÉES

Indépendamment de ce que l'on peut savoir sur le mode de collecte des données et sur la façon dont les enquêteurs ont été formés et supervisés, une appréciation méticuleuse de la qualité des données est une première étape essentielle de toute analyse. Toutes les séries de données contiennent des erreurs, qui peuvent provenir de nombreuses sources, telles que certains raccourcis pris par des interviewers ou un interviewé qui, simplement, ne connaît pas la réponse correcte à une question. Chaque section ci-dessous commence par une description des techniques d'évaluation de la qualité des données avant de passer aux méthodes d'analyse. Ces techniques d'évaluation examinent à la fois la cohérence interne d'une série de données et la cohérence externe par comparaison avec d'autres séries de données pour la même population. Notez en passant que la présence d'erreurs ne signifie pas nécessairement qu'une série de données ne doit pas être analysée, l'important est d'apprécier l'ampleur de ces erreurs et d'en tenir compte quand on interprète les résultats.

L'histoire génésique complète : évaluation de la qualité des données

La première étape d'une évaluation systématique de la qualité des données consiste à examiner la fréquence des valeurs manquantes. Dans une histoire génésique complète, des valeurs peuvent manquer pour diverses raisons. Par exemple, des ménages entiers inclus dans le plan de sondage original peuvent manquer. Il se peut en outre qu'au sein d'un ménage interviewé des femmes éligibles ne puissent pas être interrogées. Par ailleurs, des informations spécifiques peuvent manquer au sein d'une histoire génésique lorsque la femme interrogée ne connaît pas la date de naissance d'un enfant, ou qu'elle ignore si un enfant est encore en vie, ou (en cas de décès) l'âge au décès. Les proportions d'événements susceptibles d'être affectés par ces erreurs doivent être examinées. Les items manquants peuvent être imputés au cours du nettoyage des données, mais les valeurs imputées doivent être signalées. L'absence de valeurs manquantes ne doit pas être considérée comme un marqueur fort de la qualité des données ; cela peut même être considéré comme un avertissement : dans certaines enquêtes, les enquêteurs et les superviseurs sont formés pour éviter les valeurs manquantes et, dans ces cas, les données peuvent être plus ou moins fabriquées par l'interviewer.

La seconde étape dans l'évaluation de la qualité des données consiste à rechercher des irrégularités peu vraisem-

blables dans les résultats agrégés. Les irrégularités le plus souvent détectées concernent les rapports de masculinité à la naissance, dans les distributions annuelles de naissances vivantes, et les âges au décès. En l'absence d'intervention, les rapports de masculinité dans les populations humaines sont généralement compris entre 100 et 106 garçons pour 100 filles. Des rapports de masculinité à la naissance hors de cet intervalle indiquent probablement une erreur. Des rapports de masculinité qui augmentent quand on remonte vers des cohortes nées longtemps avant l'enquête sont des indicateurs particulièrement clairs d'une erreur, dans ce cas une sous-déclaration des naissances féminines anciennes.

En l'absence d'événements majeurs, les naissances évoluent normalement de manière régulière au fil des années, les variations saisonnières n'affectant pas les nombres annuels. D'éventuelles erreurs peuvent être détectées par le calcul de « rapports de naissances », définis comme $\frac{2B_t}{B_{t-1} + B_{t+1}}$ où

B_t est le nombre de naissances déclarées l'année t . Une erreur courante dans les fichiers de données des EDS a été intitulée « transfert de naissances ». Les enquêtes EDS recueillent de nombreuses données supplémentaires sur les enfants nés depuis une date précise, en général le 1^{er} janvier de l'année cinq ans avant l'enquête. Il arrive souvent que les naissances survenues cette année soient déclarées comme ayant eu lieu l'année précédente, sans doute pour réduire la charge de travail. Il en résulte un déficit des naissances l'année qui suit la date marquant le début de la période concernée, et un surplus l'année qui précède. Les rapports de naissances feront apparaître cette erreur, puisque le rapport sera bas pour l'année qui débute à la date de référence et sera élevé pour l'année qui précède. Ce transfert de naissances est très souvent plus marqué pour les enfants décédés que pour ceux encore en vie, aussi est-il bon de calculer des rapports séparés pour les enfants survivants et décédés.

Des irrégularités dans la déclaration des âges au décès peuvent être détectées de la même façon par le calcul du rapport entre les décès à l'âge x et la moyenne des décès aux âges encadrants $(x-1)$ et $(x+1)$. Dans les données des EDS, il y a généralement un excès de décès à 7 jours, dans une moindre mesure à 14 jours, et à 12 mois.

Les EDS publient ces indicateurs de qualité des données au niveau agrégé (national) dans les rapports d'enquêtes (souvent en Annexe C). Ceux qui veulent réaliser des analyses à un niveau plus fin doivent calculer ces indicateurs par eux-mêmes.

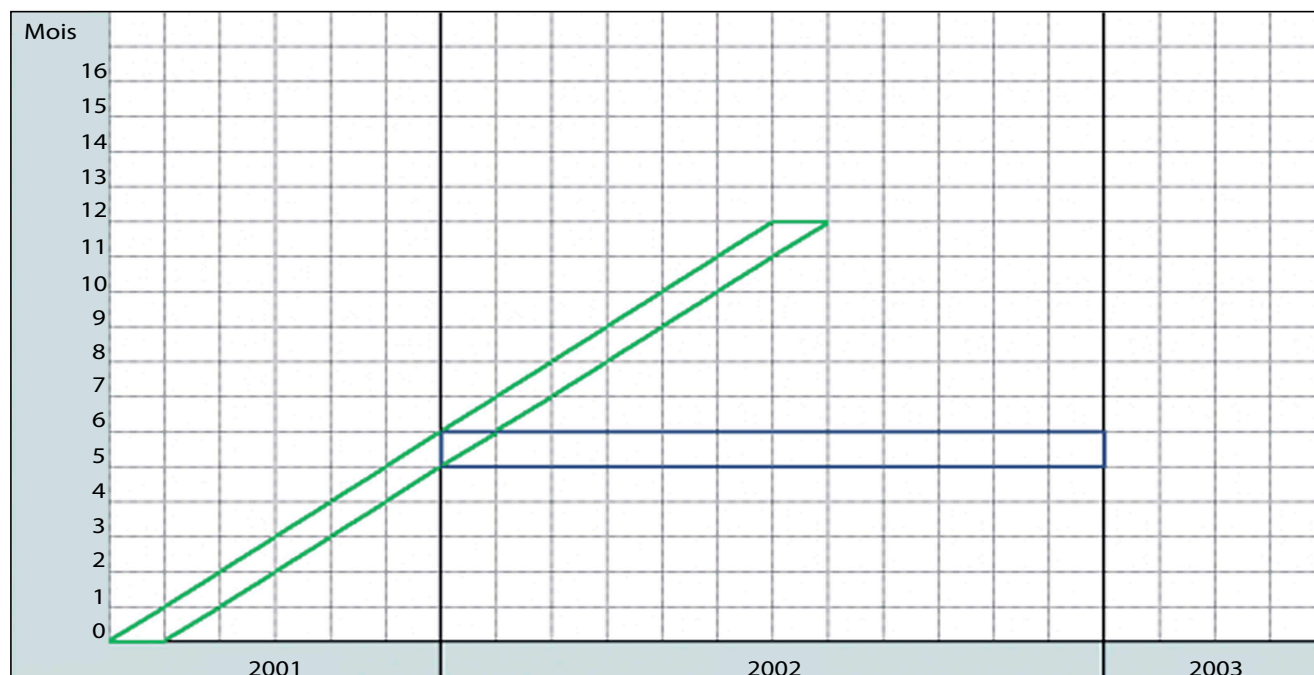
Les indicateurs ci-dessus sont des mesures de vraisemblance interne. Mais des données peuvent être vraisemblables en interne et pourtant erronées. Les données doivent être aussi évaluées par comparaison avec d'autres enquêtes pour la même population. Les comparaisons par cohorte sont particulièrement utiles, par exemple la comparaison du nombre moyen d'enfants nés vivants des femmes âgées de 30-34 ans dans une enquête au nombre moyen d'enfants nés vivants des femmes âgées de 35-39 ans dans une autre enquête menée cinq ans plus tard. Les mêmes comparaisons peuvent être faites avec les nombres moyens d'enfants décédés. Des suites de naissances par année civile peuvent aussi être comparées pour des périodes qui se chevauchent, mais en ayant à l'esprit que les naissances du passé sont tronquées de façon croissante quand les histoires génésiques ne sont obtenues qu'auprès des femmes âgées de 15-49 ans au moment de l'enquête.

L'HISTOIRE GÉNÉSIQUE COMPLÈTE DES NAISSANCES : CALCUL DES INDICATEURS DE MORTALITÉ DES JEUNES ENFANTS DANS LES COHORTES DE NAISSANCE

Les indicateurs de mortalité des jeunes enfants couramment utilisés sont exprimés comme des probabilités. Ainsi, le taux

de mortalité infantile est la probabilité de décéder avant le premier anniversaire, ${}_1q_0$ (ceci est une approximation car le taux de mortalité infantile est conventionnellement défini comme le rapport du nombre de décès infantiles au cours d'une année au nombre de naissances de cette même année, une valeur très voisine de ${}_1q_0$). Le taux de mortalité avant 5 ans est la probabilité de décéder avant le 5^e anniversaire, ${}_5q_0$. Au sens strict, les probabilités sont des mesures relatives à des cohortes réelles, même si la plupart des tables de mortalité calculent des mesures relatives à des cohortes synthétiques, pour des périodes, à partir de taux de mortalité par âge et période. Il est très simple de calculer des probabilités pour des cohortes à partir d'histoires génésiques complètes. Par exemple, le taux de mortalité infantile par cohorte pour les enfants nés 12 à 23 mois avant l'enquête est simplement le nombre de ces enfants décédés avant l'âge d'un an divisé par le nombre de naissances. De même, le taux de mortalité avant 5 ans par cohorte pour les enfants nés 5 à 9 ans avant l'enquête est le nombre d'enfants décédés avant leur 5^e anniversaire divisé par le nombre de naissances il y a 5-9 ans. À la figure 17.1, on a représenté sur un diagramme de Lexis la probabilité par âge et cohorte de décéder avant le 1^{er} anniversaire pour la cohorte née en juillet 2001 (en vert),

FIGURE 17.1 Représentation sur un diagramme de Lexis de la mortalité avant un an par âge et cohorte (vert) et de la mortalité à 5 mois par âge et période (bleu)



et la mortalité par âge et période des enfants âgés de 5 mois au cours de l'année 2002 (le rectangle bleu, reprenant un exemple qui sera utilisé plus loin dans ce chapitre).

Le tableau 17.1 présente les données pour le calcul de la probabilité par âge et cohorte de décéder avant l'âge d'un an pour les enfants nés 12-23 mois avant l'enquête et de la probabilité de décéder avant 5 ans pour la cohorte née 5 à 9 ans avant l'enquête, d'après l'EDS 2004 du Malawi.

TABLEAU 17.1 Calcul du taux de mortalité infantile et du taux de mortalité avant 5 ans, Malawi, EDS 2004

| | Période de naissance | |
|--|----------------------------|-----------------------------|
| | 12-23 mois avant l'enquête | 60-119 mois avant l'enquête |
| Naissances | 2 229 | 7 178 |
| Dont, décès avant 12 mois | 143 | |
| Dont, décès avant 5 ans | | 1 568 |
| Indicateur de mortalité | 190 | 590 |
| Estimation par cohorte pour 1 000 naissances | 64,2 | 218,4 |

Note : données pondérées ; les événements se produisant le mois de l'interview ont été exclus

Notez qu'il n'y a pas d'interprétation par période (en transversal) de ces valeurs de cohortes ; dans l'exemple du taux de mortalité avant 5 ans, la probabilité par cohorte reflète des risques de mortalité dans chacune des 10 années précédant l'enquête. Notez aussi que la probabilité de décéder avant l'âge x ne peut être calculée que pour des cohortes nées au moins x années avant l'enquête. Ces deux considérations limitent la valeur des mesures de cohorte, car dans la plupart des cas les analystes et les responsables politiques sont davantage intéressés par les mesures transversales.

L'HISTOIRE GÉNÉSIQUE COMPLÈTE :
CALCUL DES INDICATEURS DE MORTALITÉ
DES JEUNES ENFANTS PAR PÉRIODE

Les mesures par période sont estimées en utilisant le concept de cohorte synthétique. Les taux de mortalité pour des intervalles d'âge restreints et des périodes définies sont calculés sur la base d'événements et d'exposition au risque dans des rectangles du diagramme de Lexis. Les taux sont ensuite convertis en probabilités, en utilisant les relations démographiques classiques (voir, par exemple, Preston, Heuveline et Guillot (2001)) et en faisant généralement quelques

hypothèses peu contraignantes sur la distribution des décès dans chaque rectangle. Enfin, les probabilités de décéder sont appliquées successivement à une cohorte hypothétique de naissances ; il en résulte une courbe de survie $\ell(x)$ pour chaque âge x , à partir de laquelle on tire facilement des probabilités de décéder.

Les données des histoires génésiques complètes de naissances se prêtent aisément à ces calculs de tables de mortalité. Si les données sont collectées selon la pratique des EDS – en mois et année pour les dates de naissance ; en jours, mois ou années (en fonction de l'âge) pour les âges au décès – les décès peuvent être situés sans grande ambiguïté dans des rectangles âge-période du diagramme de Lexis (il restera une légère ambiguïté, du fait de l'imprécision de l'information sur la date de naissance et l'âge au décès, mais l'impact dépendra de la taille des rectangles.) Nous décrivons ici une démarche fondée sur le calcul de taux de mortalité par âge pour une seule année civile (taux par âge et période) pour la mortalité avant 5 ans. Il est facile d'étendre ensuite à d'autres périodes de temps. On suppose que les données sont reprises dans un format EDS classique, c'est-à-dire que les dates de naissance sont enregistrées en nombre de mois depuis le 1^{er} janvier 1900, et les âges au décès sont en jours, mois ou années. On doit disposer de données individuelles. L'unité d'âge est le mois. Les calculs de base sont donc des taux de mortalité par âge, par mois d'âge et année civile. Ces taux sont convertis en probabilités de décéder chaque mois. Ces probabilités sont ensuite converties en probabilité de survie, et sont combinées sur tout intervalle d'âge nécessaire (en particulier, jusqu'à 5 ans). L'essentiel du calcul consiste à assigner les décès et les expositions au risque à des segments d'âge d'un mois au cours d'une année civile.

Traitement des données

Dans une série de données sur les naissances de type EDS, on utilise quatre variables :

- 1. $b3$, date de naissance en nombre de mois depuis le 1^{er} janvier 1900 ;
- 2. $b5$, survie de l'enfant ;
- 3. $b6$, âge au décès, où le premier chiffre indique l'unité (1, jours ; 2, mois ; et 3, années) et les deuxième et troisième chiffres la valeur dans cette unité ; et
- 4. $v005$, coefficient de pondération d'échantillonnage, en millions.

Notez que la variable $b7$, âge au décès (en mois, par imputation), n'est pas utilisée. Cette variable ne se prête pas à la

procédure de calcul des taux de mortalité décrite ici car, dans les cas où l'âge au décès est enregistré en années, le mois « imputé » est en fait égal à la limite inférieure de l'intervalle (c'est-à-dire que si l'âge au décès est enregistré comme « 3 ans », l'âge imputé en mois est 36 mois). Cette variable conduirait donc à une localisation systématiquement erronée des décès dans le temps.

APPLICATION DE LA MÉTHODE

Étape 1 : Traitement de l'âge au décès et calcul de la date de naissance et de l'âge au décès estimés

Nous voulons situer les décès par mois de leur survenue. Comme nous n'avons pas une date de naissance précise (car elle est exprimée seulement en nombre de mois depuis le 1^{er} janvier 1900), et qu'en général nous n'avons par un âge au décès précis (sauf pour les décès néonataux), il nous faut imputer à la fois la date de naissance et l'âge au décès. Nous pouvons le faire en utilisant des nombres aléatoires.

Il n'est évidemment pas souhaitable – entre autres par manque de reproductibilité – d'utiliser pour cela un vrai générateur de nombres au hasard. En outre, une « vraie » randomisation risque de produire une fausse impression de précision. À la place, nous proposons de créer des nombres pseudo-aléatoires à partir de variables couramment disponibles dans les données des EDS et qui peuvent être appliquées dans l'algorithme ci-dessus. Il est facile de créer de nouvelles variables en répartissant les enregistrements en déciles sur la base du jour d'interview (*v016* dans une EDS) et du numéro de ménage (*v002*). (Ces variables ont été choisies parce qu'il est peu vraisemblable qu'elles soient corrélées avec la mortalité des jeunes enfants). Ces nouvelles variables prendront des valeurs dans l'intervalle (0, 1 ... 9). En divisant ces valeurs par 10 et en ajoutant 0,05 on obtient deux nouvelles variables distribuées uniformément, *aléa1* et *aléa2*, qui prennent des valeurs dans l'intervalle (0,05, 0,15, ..., 0,95).

Il est ensuite facile d'imputer une date de naissance (*ddn*, en mois), si les naissances du mois de l'interview sont exclues de l'analyse, en ajoutant *aléa1* à *b3* (la date de naissance en nombre de mois depuis le 1^{er} janvier 1900). La méthode d'imputation de l'âge au décès (en unités mensuelles) dépend de l'« unité ». Quand « unité » = 1 (c'est-à-dire que l'âge au décès est mesuré en jours), l'âge au décès (*aad*) peut être estimé comme (« valeur » + *aléa2*)/31 (quand l'âge au décès est en jours, ce n'est pas nécessaire, mais nous le pré-

sentons ici pour être systématique) ; quand « unité » = 2, l'âge au décès est « valeur » + *aléa2*, et quand « unité » = 3, l'âge au décès est (« valeur » + *aléa2*)*12.

Étape 2 : Situer les décès au cours de l'année étudiée

Pour chaque taux de mortalité par mois d'âge, les événements sont les décès à cet âge au cours de la période étudiée. L'étape 1 a imputé l'âge au décès en mois. La date de décès *ddd* est égale à la somme du mois de naissance imputé *ddn* et de l'âge au décès imputé *aad*. Si l'âge au décès imputé appartient à l'intervalle d'âge et si la date de décès imputée tombe dans la période étudiée, l'événement est retenu.

Étape 3 : Calcul de l'exposition au risque

Le calcul de l'exposition au risque est subtil, mais relativement direct. L'intervalle d'âge se réfère aux âges (définis dans les unités appropriées) pour lesquels nous voulons mesurer la mortalité. Nous désignons la limite basse de l'intervalle d'âge comme x_l et la limite haute comme x_h .

La période étudiée est la mesure de l'intervalle de temps pour laquelle nous cherchons à estimer la mortalité ; elle est définie comme la période ($t_2 - t_1$), où t_2 est la date de fin de la période étudiée et t_1 la date de début, mesurées dans les mêmes unités que celle définie par l'intervalle d'âge.

Graphiquement nous cherchons ensuite à mesurer la mortalité à l'âge et dans la période définis par les lignes épaisses sur la figure 17.2.

La trajectoire d'un individu par âge et période s'inscrit sur une diagonale (comme dans un diagramme de Lexis classique). Cinq scénarios possibles sont représentés, étiquetés de (a) à (e). La position d'un individu dans l'espace peut être définie par son âge à t_1 , x_{t1} . Il s'ensuit que toute personne âgée de x à t_1 , si elle ne décède pas avant t_2 , doit être âgée de $x_{t2} = x_{t1} + (t_2 - t_1)$ à t_2 . Nous définissons l'âge au décès des personnes décédées dans un intervalle d'âge donné au cours de la période étudiée comme x_d . La contribution relative de chaque scénario à l'exposition au risque est déterminée par l'algorithme du tableau 17.2.

En appliquant ces règles à chaque individu pour définir l'exposition au risque dans l'intervalle d'âge au cours de la période étudiée et en agrégeant, on obtient l'exposition totale au risque, qui est le dénominateur du taux de mortalité. La somme des décès survenus dans l'intervalle d'âge au cours de la période étudiée fournit le numérateur.

FIGURE 17.2 Diagramme de Lexis illustrant le calcul de l'exposition au risque

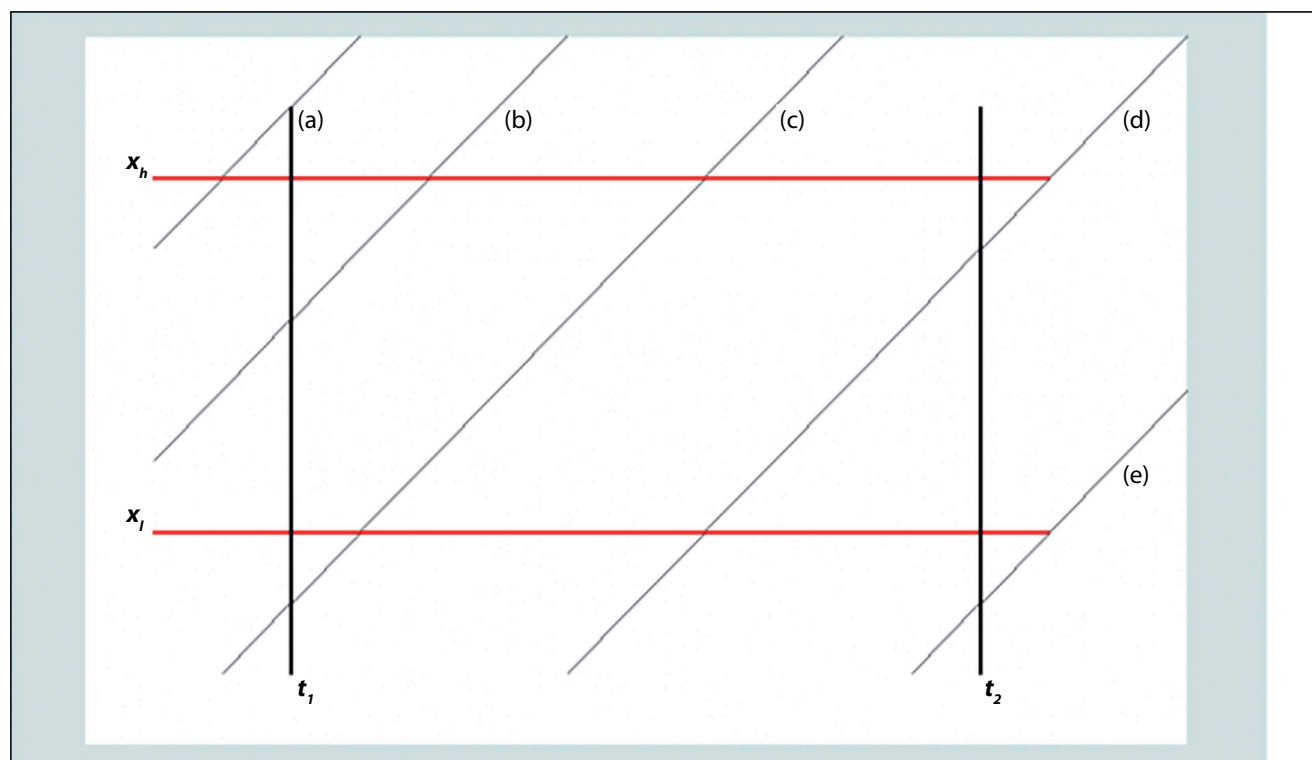


TABLEAU 17.2 Algorithme pour déterminer l'exposition au risque

| Scenario | Description | Règle(s) de définition | Exposition des survivants au cours de la période étudiée | Exposition de décédés (quand le décès survient au cours de la période étudiée) |
|----------|--|--|--|--|
| (a) | Agé de plus de x_h en t_1 | $x_{t1} > x_h$ | 0 | 0 |
| (b) | Agé entre x_l et x_h en t_1 . Atteint x_h au cours de la période étudiée | $x_l < x_{t1} < x_h$ $x_{t1} + (t_2 - t_1) > x_h$ | $x_h - x_{t1}$ | $x_d - x_{t1}$ |
| (c) | Atteint x_l et x_h au cours de la période étudiée | $x_l > x_{t1}$ $x_{t1} + (t_2 - t_1) > x_h$ | $x_h - x_l$ | $x_d - x_l$ |
| (d) | Atteint x_l au cours de la période étudiée mais la période se termine avant que x_h soit atteint | $x_l > x_{t1}$ $x_l < x_{t1} + (t_2 - t_1) < x_h$ | $x_{t1} + (t_2 - t_1) > x_l$ | $x_d - x_l$ |
| (e) | N'atteint pas x_l au cours de la période étudiée | $x_{t1} + (t_2 - t_1) < x_l$ | 0 | 0 |

Étape 4 : Pondération et cumul des événements et de l'exposition au risque

La variable de pondération dans un fichier EDS standard est $v005$. Elle a une moyenne de 1 000 000. Pour éviter de donner l'impression que l'échantillon est énorme (et que les

intervalles de confiance sont très étroits), nous recommandons d'abord de recalculer les pondérations et de définir une nouvelle variable $wgt = v005 / 1\,000\,000$. Les taux de mortalité peuvent être calculés en tenant compte des contributions de chacun des N enfants dans l'enquête au nombre des événe-

ments et au temps total d'exposition. Le taux de mortalité par âge aux âges x à $x + 1$ (en mois) au cours d'une période j est

$$M(x, j) = \frac{\sum_{i=1}^N D(i, x, j) \cdot wgt(i)}{\sum_{i=1}^N E(i, x, j) \cdot wgt(i)}$$

où $D(i, x, j)$ est une variable binaire indiquant le décès de l'enfant i à l'âge x l'année j (1 si le décès survient, sinon 0), $E(i, x, j)$ est le temps d'exposition de l'enfant i à l'âge x l'année j , et où $wgt(i)$ est la pondération d'échantillonnage (moyenne 1,0) de l'enfant i .

Étape 5 : Calcul des probabilités de décéder à partir des taux de mortalité par âge

Les taux calculés à l'étape 4 sont par mois d'âge d'exposition. Il est donc nécessaire d'en tenir compte en adaptant la formule classique de passage d'un taux à une probabilité de décéder par période. Comme nous avons fait diverses hypothèses simplificatrices et que nous travaillons avec des intervalles d'âge étroits, il convient de supposer que les décès sont distribués régulièrement entre chaque intervalle d'âge mensuel. Nous pouvons donc calculer $q(x)$ comme

$$q(x, j) = \frac{\frac{M(x, j)}{12}}{\left(1 + \frac{M(x, j)}{24}\right)}$$

Les probabilités de survie de la naissance jusqu'à n'importe quel âge peuvent être obtenues en combinant les probabilités de survie mensuelles (c'est-à-dire $(1 - q(x, j))$). Ainsi par exemple

$${}_5q_0^j = 1 - \prod_{x=0}^{59} (1 - q(x, j))$$

EXEMPLE

Comme on l'a dit, l'estimation directe de la mortalité des jeunes enfants à partir des historiques de naissances s'appuie sur l'utilisation des enregistrements individuels plutôt que sur des tableaux de données agrégées. C'est pourquoi l'exemple que nous prenons illustre la méthode à partir d'un nombre restreint d'enregistrements tirés d'une EDS, plus précisément la mortalité des enfants âgés de 5 mois en 2002 d'après l'EDS 2004 du Malawi. Seuls les enfants nés entre le 1^{er} juillet 2001 et le 31 juillet 2002 sont exposés

TABEAU 17.3 Données de base d'histoires génésiques pour l'estimation directe de la mortalité des jeunes enfants

| Enregistrement | b3 | b5 | b6 | v005 |
|----------------|------|-----|-----|---------|
| 1 | 1223 | oui | . | 469061 |
| 2 | 1223 | oui | . | 469061 |
| 3 | 1222 | non | 107 | 469061 |
| 4 | 1224 | oui | . | 469061 |
| 5 | 1223 | oui | . | 469061 |
| 6 | 1218 | non | 205 | 469061 |
| 7 | 1230 | oui | . | 2171218 |
| 8 | 1225 | oui | . | 704240 |
| 9 | 1230 | oui | . | 704240 |
| 10 | 1224 | oui | . | 704240 |
| 11 | 1224 | non | 202 | 704240 |
| 12 | 1221 | oui | . | 1106470 |
| 13 | 1225 | oui | . | 1106470 |
| 14 | 1224 | non | 205 | 1106470 |
| 15 | 1221 | oui | . | 1106470 |
| 16 | 1221 | oui | . | 1106470 |
| 17 | 1218 | non | 205 | 1106470 |
| 18 | 1229 | oui | . | 3900164 |
| 19 | 1230 | oui | . | 1247934 |
| 20 | 1224 | oui | . | 1247934 |
| 21 | 1226 | non | 201 | 1247934 |
| 22 | 1221 | oui | . | 537170 |
| 23 | 1218 | oui | . | 537170 |
| 24 | 1227 | oui | . | 537170 |
| 25 | 1226 | oui | . | 537170 |
| 26 | 1224 | oui | . | 1095220 |
| 27 | 1230 | non | 205 | 1594776 |
| 28 | 1225 | oui | . | 1594776 |
| 29 | 1221 | oui | . | 1594776 |
| 30 | 1225 | oui | . | 1594776 |
| 31 | 1229 | non | 208 | 1538303 |
| 32 | 1223 | oui | . | 1538303 |
| 33 | 1220 | oui | . | 1538303 |
| 34 | 1226 | oui | . | 1538303 |
| 35 | 1225 | oui | . | 1538303 |
| 36 | 1220 | oui | . | 1538303 |
| 37 | 1224 | non | 205 | 1538303 |
| 38 | 1228 | oui | . | 1538303 |
| 39 | 1219 | oui | . | 3789587 |
| 40 | 1228 | oui | . | 2011510 |
| 41 | 1223 | non | 302 | 2011510 |
| 42 | 1220 | oui | . | 2011510 |
| 43 | 1220 | oui | . | 2011510 |
| 44 | 1221 | oui | . | 686252 |
| 45 | 1228 | non | 201 | 686252 |
| 46 | 1229 | oui | . | 2451926 |
| 47 | 1219 | oui | . | 2451926 |
| 48 | 1219 | oui | . | 1043244 |
| 49 | 1224 | oui | . | 1043244 |
| 50 | 1230 | non | 205 | 1043244 |

au risque de décéder à l'âge de 5 mois au cours de l'année 2002 (les enfants nés avant le 1^{er} juillet 2002 seraient âgés de 6 mois ou plus au début de l'année 2002, et ceux nés après le 31 juillet 2002 n'auraient pas atteint l'âge de 5 mois au cours de l'année). Nous n'avons fait figurer que les enregistrements pertinents, c'est-à-dire ceux correspondant à des naissances entre les mois 1218 et 1230 (en nombres de mois depuis le 1^{er} janvier 1900, soit de juillet 2001 à juillet 2002). Dans la pratique, nous excluons aussi les naissances ayant donné lieu à un décès avant cinq mois d'âge, mais nous les gardons dans l'exemple pour montrer que nous les excluons des calculs.

Le tableau 17.3 présente les variables essentielles pour 50 enregistrements tirés de l'EDS 2004 du Malawi. Ce sont des enregistrements de naissances, pas de femmes.

Étape 1 : Traitement de l'âge au décès et calcul de la date de naissance et de l'âge au décès estimés

Les nombres aléatoires *aléa1* et *aléa2* sont établis comme il a été expliqué plus haut. Il en résulte les valeurs révisées des dates de naissance et âge au décès, *ddn'* et *aad'*. La date de décès *ddd'* est estimée comme la somme du mois de naissance imputé *ddn'* et du mois de décès imputé *aad'*. Les *ddd'* figurent à la colonne 10 du tableau 17.4.

Étape 2 : Situer les décès au cours de l'année étudiée

Nous retenons les décès survenus entre les mois 1224 et 1235 (dans une numérotation depuis janvier 1900). Les décès correspondant aux enregistrements 3, 6, 31 et 41 au tableau 17.4 sont donc exclus car ils ne se sont pas produits en 2002. Les décès des enregistrements 11 et 45 ne sont pas pris en compte ici car les enfants sont décédés respectivement à l'âge de 2 mois et 1 mois et ne sont donc pas exposés au risque de décéder à 5 mois.

Étape 3 : Calcul de l'exposition au risque

Le tableau 17.5 présente le calcul de l'exposition au risque pour les 50 cas retenus. La règle de détermination de l'exposition figure dans la colonne « Scénario ». L'exposition qui en résulte figure aux deux colonnes suivantes pour ceux qui survivent au terme de la période étudiée et pour ceux qui décèdent au cours de la période.

Pour les enfants qui survivent à l'âge de 6 mois, ceux nés les mois 1219 à 1229 comptent pour un mois entier d'exposition au cours de l'âge-période considéré (c'est-à-dire de 5 mois à 6 mois exacts). Ainsi l'enregistrement 1 (né en

1223,55) contribue pour un mois entier. Un enfant né le mois 1218 contribuera pour (*ddn* – 1218) mois et donc l'enregistrement 23 (né en 1218,85) contribue pour 0,85 mois. Un enfant né le mois 1230 contribuera pour (1231 – *ddn*) mois et donc l'enregistrement 7 contribue pour 1231 – 1230,55 = 0,45 mois. Les enfants nés les mois 1219 à 1229 qui décèdent à 5 mois d'âge contribueront pour (*aad* – 5) mois d'exposition ; ainsi le décès dans l'enregistrement 14 survient à 5,25 mois et contribue pour 0,25 mois d'exposition.

Étape 4 : Pondération et cumul des événements et de l'exposition au risque

L'étape finale avant le calcul du taux de mortalité consiste à tenir compte de la pondération d'échantillonnage de chaque enregistrement à la fois pour les décès et pour le temps d'exposition, puis à faire la somme des décès et des expositions. Les colonnes 6 et 7 du tableau 17.5 présentent l'exposition au risque des survivant et des décédés. Les colonnes 8 et 9 résultent de la multiplication des colonnes 6 et 7 respectivement par la pondération d'échantillonnage $v005/1\ 000\ 000$. Le taux de mortalité par âge $M(5,2002)$ est ensuite calculé en divisant la somme des décès pondérés par la somme des temps d'exposition pondérés :

$$M(x, j) = \frac{\sum_{i=1}^N D(i, x, j) \times wgt(i)}{\sum_{i=1}^N E(i, x, j) \times wgt(i)} = \frac{6,389}{59,317} = 0,1077.$$

Étape 5 : Calcul des probabilités de décéder à partir des taux de mortalité par âge

Les taux calculés à l'étape 4 sont par mois d'exposition. Il est donc nécessaire d'en tenir compte en adaptant la formule classique de passage d'un taux à une probabilité de décéder par période. Comme nous avons fait diverses hypothèses simplificatrices et que nous travaillons avec des intervalles d'âge étroits, il convient de supposer que les décès sont distribués régulièrement entre chaque intervalle d'âge mensuel, même pour le premier mois de vie. Nous pouvons donc calculer $q(x)$ comme

$$\begin{aligned} q(5, 2002) &= \frac{\frac{M(5, 2002)}{(1 + \frac{M(5, 2002)}{24})}}{\frac{0,1077}{(1 + \frac{0,1077}{24})}} \\ &= \frac{0,008975}{1 + 0,004488} = 0,008935. \end{aligned}$$

TABLEAU 17.4 Calcul de la date de naissance, de l'âge au décès et de la date de décès imputés. Malawi, EDS 2004 (50 observations)

| Enregistrement | <i>b3</i> | <i>b5</i> | <i>b6</i> | <i>v005</i> | <i>aléa1</i> | <i>aléa2</i> | <i>ddn'</i> | <i>aad'</i> | <i>ddd'</i> |
|----------------|-----------|-----------|-----------|-------------|--------------|--------------|-------------|-------------|-------------|
| 1 | 1223 | oui | . | 469061 | 0,55 | | 1223,55 | | |
| 2 | 1223 | oui | . | 469061 | 0,85 | | 1223,85 | | |
| 3 | 1222 | non | 107 | 469061 | 0,15 | 0,05 | 1222,15 | 0,28 | 1222,43 |
| 4 | 1224 | oui | . | 469061 | 0,25 | | 1224,25 | | |
| 5 | 1223 | oui | . | 469061 | 0,25 | | 1223,25 | | |
| 6 | 1218 | non | 205 | 469061 | 0,05 | 0,45 | 1218,05 | 5,45 | 1223,5 |
| 7 | 1230 | oui | . | 2171218 | 0,55 | | 1230,55 | | |
| 8 | 1225 | oui | . | 704240 | 0,55 | | 1225,55 | | |
| 9 | 1230 | oui | . | 704240 | 0,25 | | 1230,25 | | |
| 10 | 1224 | oui | . | 704240 | 0,35 | | 1224,35 | | |
| 11 | 1224 | non | 202 | 704240 | 0,55 | 0,75 | 1224,55 | 2,75 | 1227,3 |
| 12 | 1221 | oui | . | 1106470 | 0,45 | | 1221,45 | | |
| 13 | 1225 | oui | . | 1106470 | 0,75 | | 1225,75 | | |
| 14 | 1224 | non | 205 | 1106470 | 0,85 | 0,25 | 1224,85 | 5,25 | 1230,1 |
| 15 | 1221 | oui | . | 1106470 | 0,35 | | 1221,35 | | |
| 16 | 1221 | oui | . | 1106470 | 0,45 | | 1221,45 | | |
| 17 | 1218 | non | 205 | 1106470 | 0,95 | 0,65 | 1218,95 | 5,65 | 1224,6 |
| 18 | 1229 | oui | . | 3900164 | 0,45 | | 1229,45 | | |
| 19 | 1230 | oui | . | 1247934 | 0,65 | | 1230,65 | | |
| 20 | 1224 | oui | . | 1247934 | 0,65 | | 1224,65 | | |
| 21 | 1226 | non | 201 | 1247934 | 0,75 | 0,85 | 1226,75 | 1,85 | 1228,6 |
| 22 | 1221 | oui | . | 537170 | 0,65 | | 1221,65 | | |
| 23 | 1218 | oui | . | 537170 | 0,85 | | 1218,85 | | |
| 24 | 1227 | oui | . | 537170 | 0,95 | | 1227,95 | | |
| 25 | 1226 | oui | . | 537170 | 0,85 | | 1226,85 | | |
| 26 | 1224 | oui | . | 1095220 | 0,95 | | 1224,95 | | |
| 27 | 1230 | non | 205 | 1594776 | 0,15 | 0,65 | 1230,15 | 5,65 | 1235,8 |
| 28 | 1225 | oui | . | 1594776 | 0,15 | | 1225,15 | | |
| 29 | 1221 | oui | . | 1594776 | 0,85 | | 1221,85 | | |
| 30 | 1225 | oui | . | 1594776 | 0,05 | | 1225,05 | | |
| 31 | 1229 | non | 208 | 1538303 | 0,65 | 0,85 | 1229,65 | 8,85 | 1238,5 |
| 32 | 1223 | oui | . | 1538303 | 0,45 | | 1223,45 | | |
| 33 | 1220 | oui | . | 1538303 | 0,15 | | 1220,15 | | |
| 34 | 1226 | oui | . | 1538303 | 0,55 | | 1226,55 | | |
| 35 | 1225 | oui | . | 1538303 | 0,95 | | 1225,95 | | |
| 36 | 1220 | oui | . | 1538303 | 0,45 | | 1220,45 | | |
| 37 | 1224 | non | 205 | 1538303 | 0,25 | 0,85 | 1224,25 | 5,85 | 1230,1 |
| 38 | 1228 | oui | . | 1538303 | 0,35 | | 1228,35 | | |
| 39 | 1219 | oui | . | 3789587 | 0,35 | | 1219,35 | | |
| 40 | 1228 | oui | . | 2011510 | 0,15 | | 1228,15 | | |
| 41 | 1223 | non | 302 | 2011510 | 0,65 | 0,55 | 1223,65 | 30,6 | 1254,25 |
| 42 | 1220 | oui | . | 2011510 | 0,35 | | 1220,35 | | |
| 43 | 1220 | oui | . | 2011510 | 0,25 | | 1220,25 | | |
| 44 | 1221 | oui | . | 686252 | 0,95 | | 1221,95 | | |
| 45 | 1228 | non | 201 | 686252 | 0,85 | 0,35 | 1228,85 | 1,35 | 1230,2 |
| 46 | 1229 | oui | . | 2451926 | 0,25 | | 1229,25 | | |
| 47 | 1219 | oui | . | 2451926 | 0,05 | | 1219,05 | | |
| 48 | 1219 | oui | . | 1043244 | 0,85 | | 1219,85 | | |
| 49 | 1224 | oui | . | 1043244 | 0,95 | | 1224,95 | | |
| 50 | 1230 | non | 205 | 1043244 | 0,35 | 0,35 | 1230,35 | 5,35 | 1235,7 |

TABLEAU 17.5 Calcul de l'exposition au risque pour l'estimation de la mortalité des jeunes enfants. Malawi, EDS 2004 (50 observations)

| Enregistrement | ddn' | aad' | ddd' | v005 | Scénario | Exposition au risque | | Pondéré | |
|----------------|---------|-------|--------|---------|----------|----------------------|-------|------------|-------|
| | | | | | | Survivants | Décès | Exposition | Décès |
| 1 | 1223,55 | | | 469061 | c | 1 | | 0,469 | |
| 2 | 1223,85 | | | 469061 | c | 1 | | 0,469 | |
| 3 | 1222,15 | 0,25 | 1222,4 | 469061 | N/A | N/A | N/A | 0,000 | |
| 4 | 1224,25 | | | 469061 | c | 1 | | 0,469 | |
| 5 | 1223,25 | | | 469061 | c | 1 | | 0,469 | |
| 6 | 1218,05 | 5,45 | 1223,5 | 469061 | N/A | N/A | N/A | 0,000 | |
| 7 | 1230,55 | | | 2171218 | d | 0,45 | | 0,977 | |
| 8 | 1225,55 | | | 704240 | c | 1 | | 0,704 | |
| 9 | 1230,25 | | | 704240 | d | 0,75 | | 0,528 | |
| 10 | 1224,35 | | | 704240 | c | 1 | | 0,704 | |
| 11 | 1224,55 | 2,75 | 1227,3 | 704240 | c | 1 | | 0,704 | |
| 12 | 1221,45 | | | 1106470 | c | 1 | | 1,106 | |
| 13 | 1225,75 | | | 1106470 | c | 1 | | 1,106 | |
| 14 | 1224,85 | 5,25 | 1230,1 | 1106470 | c | | 0,25 | 0,277 | 1,106 |
| 15 | 1221,35 | | | 1106470 | c | 1 | | 1,106 | |
| 16 | 1221,45 | | | 1106470 | c | 1 | | 1,106 | |
| 17 | 1218,95 | 5,65 | 1224,6 | 1106470 | b | | 0,6 | 0,664 | 1,106 |
| 18 | 1229,45 | | | 3900164 | c | 1 | | 3,900 | |
| 19 | 1230,65 | | | 1247934 | d | 0,35 | | 0,437 | |
| 20 | 1224,65 | | | 1247934 | c | 1 | | 1,248 | |
| 21 | 1226,75 | 1,85 | 1228,6 | 1247934 | c | 1 | | 1,248 | |
| 22 | 1221,65 | | | 537170 | c | 1 | | 0,537 | |
| 23 | 1218,85 | | | 537170 | b | 0,85 | | 0,457 | |
| 24 | 1227,95 | | | 537170 | c | 1 | | 0,537 | |
| 25 | 1226,85 | | | 537170 | c | 1 | | 0,537 | |
| 26 | 1224,95 | | | 1095220 | c | 1 | | 1,095 | |
| 27 | 1230,15 | 5,65 | 1235,8 | 1594776 | d | | 0,65 | 1,037 | 1,595 |
| 28 | 1225,15 | | | 1594776 | c | 1 | | 1,595 | |
| 29 | 1221,85 | | | 1594776 | c | 1 | | 1,595 | |
| 30 | 1225,05 | | | 1594776 | c | 1 | | 1,595 | |
| 31 | 1229,65 | 8,85 | 1238,5 | 1538303 | c | 1 | | 1,538 | |
| 32 | 1223,45 | | | 1538303 | c | 1 | | 1,538 | |
| 33 | 1220,15 | | | 1538303 | c | 1 | | 1,538 | |
| 34 | 1226,55 | | | 1538303 | c | 1 | | 1,538 | |
| 35 | 1225,95 | | | 1538303 | c | 1 | | 1,538 | |
| 36 | 1220,45 | | | 1538303 | c | 1 | | 1,538 | |
| 37 | 1224,25 | 5,85 | 1230,1 | 1538303 | c | | 0,85 | 1,308 | 1,538 |
| 38 | 1228,35 | | | 1538303 | c | 1 | | 1,538 | |
| 39 | 1219,35 | | | 3789587 | c | 1 | | 3,790 | |
| 40 | 1228,15 | | | 2011510 | c | 1 | | 2,012 | |
| 41 | 1223,65 | 32,35 | 1256 | 2011510 | c | 1 | | 2,012 | |
| 42 | 1220,35 | | | 2011510 | c | 1 | | 2,012 | |
| 43 | 1220,25 | | | 2011510 | c | 1 | | 2,012 | |
| 44 | 1221,95 | | | 686252 | c | 1 | | 0,686 | |
| 45 | 1228,85 | 1,35 | 1230,2 | 686252 | c | 1 | | 0,686 | |
| 46 | 1229,25 | | | 2451926 | c | 1 | | 2,452 | |
| 47 | 1219,05 | | | 2451926 | c | 1 | | 2,452 | |
| 48 | 1219,85 | | | 1043244 | c | 1 | | 1,043 | |
| 49 | 1224,95 | | | 1043244 | c | 1 | | 1,043 | |
| 50 | 1230,35 | 5,35 | 1235,7 | 1043244 | d | | 0,35 | 0,365 | 1,043 |
| TOTAL | | | | | | | | 59,317 | 6,389 |

Tableau 17.6 Proportions d'enfants décédés selon que la naissance s'est produite avant ou après la date de troncature, Mongolie, Enquêtes sur la santé reproductive 1998, 2003 et 2008

| Groupe d'âge | 1998 (historique complet) | | | 2003 (historique tronqué) | | | 2008 (historique tronqué) | | |
|--------------|----------------------------|-------|-------|----------------------------|-------|-------|----------------------------|-------|-------|
| | Proportion enfants décédés | | Ratio | Proportion enfants décédés | | Ratio | Proportion enfants décédés | | Ratio |
| | Avant | Après | | Avant | Après | | Avant | Après | |
| 20-24 | 0,106 | 0,070 | 1,5 | 0,222 | 0,035 | 6,3 | 0,052 | 0,041 | 1,2 |
| 25-29 | 0,140 | 0,061 | 2,3 | 0,122 | 0,036 | 3,4 | 0,083 | 0,024 | 3,5 |
| 30-34 | 0,128 | 0,082 | 1,6 | 0,117 | 0,022 | 5,4 | 0,081 | 0,015 | 5,3 |
| 35-39 | 0,072 | 0,064 | 1,1 | 0,120 | 0,025 | 4,7 | 0,097 | 0,010 | 10,2 |
| 40-44 | 0,119 | 0,068 | 1,8 | 0,150 | 0,051 | 3,0 | 0,095 | 0,010 | 9,6 |
| 45-49 | 0,213 | 0,000 | * | 0,066 | 0,048 | 1,4 | 0,119 | 0,000 | * |

Une fois tous les $q(x,j)$ calculés, ils peuvent être transformés en leurs compléments, les probabilités de survie qui, combinées entre elles, donnent les probabilités de survivre et de décéder jusqu'aux différents âges.

Pour obtenir des taux et des probabilités couvrant des périodes de plusieurs années, les sommes pondérées obtenues à l'étape 4 le sont sur toutes ces années. L'étape 5 reste la même.

Notez que la procédure décrite ici diffère de celle utilisée par les EDS, qui consiste à calculer les probabilités directement pour des quasi-cohortes (Rutstein and Rojas 2003). Les calculs sont faits dans huit groupes d'âge : néonatal, 1-2 mois, 3-5 mois, 6-11 mois et année par année de 1 à 4 ans. Pour chaque intervalle d'âge, les décès de la période sont déterminés à partir de la date de naissance et de l'âge au décès. Le nombre d'enfants exposés au risque est obtenu par approximation du nombre d'enfants qui entrent dans un intervalle d'âge au cours de la période. Cette approximation est la somme de tous les enfants qui entrent dans l'intervalle d'âge et qui en sortent (ou qui le feraient s'ils survivaient) au cours de la période, plus la moitié de ceux qui entrent dans l'intervalle d'âge au cours de la période mais qui en sortiraient après la période, plus la moitié de ceux qui sont entrés dans l'intervalle d'âge avant le début de la période mais qui en sortiraient au cours de la période.

Quelle que soit la procédure utilisée, il est nécessaire de disposer de données individuelles tirées des histoires génésiques complètes. Les calculs peuvent être réalisés à l'aide de tableaux détaillés, mais ce serait très fastidieux. Nous recom-

mandons fortement le recours à des outils informatiques adaptés.

INTERPRÉTATION

Quand on interprète les résultats, on doit avoir présent à l'esprit le risque de biais de sélection des répondantes, dû au fait que l'information est fournie par les femmes survivantes au sein des ménages enquêtés. En particulier, la mortalité des enfants nés dans un groupe auquel leurs mères n'appartiennent plus au moment de l'enquête n'est pas incluse dans les mesures. Si ces enfants ont une mortalité plus élevée que ceux nés de mères qui appartiennent encore au groupe, la mortalité sera sous-estimée. Le biais le plus important est sans doute celui qui résulte d'une prévalence élevée du VIH dans le groupe, car une telle prévalence, en l'absence d'un recours massif à la thérapie antirétrovirale, entraîne une forte corrélation positive entre la survie de l'enfant et la survie de la mère (voir chapitre 15). Mais dans toute population, il est fort probable qu'il y ait une certaine corrélation positive entre la survie de la mère et celle de l'enfant. Il peut y avoir d'autres sources de biais. Par exemple, quand l'immigration est importante, des femmes déclarent des enfants nés et élevés ailleurs, alors qu'une émigration importante élimine des réponses relatives à des enfants nés et élevés dans le groupe. Il est impossible de savoir *a priori* le sens et l'ampleur de tels biais, mais l'analyste doit avoir leur éventuel effet présent à l'esprit. Les non-réponses peuvent aussi être un problème si des femmes absentes durablement ne peuvent pas être interviewées personnellement, et si cette absence a une influence

sur la mortalité de leurs enfants ou si cette absence est en partie due aux risques différents subis par leurs enfants.

EXTENSION DE LA MÉTHODE : HISTOIRES GÉNÉSIQUES TRONQUÉES

L'histoire génésique tronquée :

Évaluation de la qualité des données

L'histoire génésique tronquée offre moins de possibilités de contrôles de qualité des données que l'histoire génésique complète, du fait même que la série d'événements déclarés est tronquée. Si la troncature porte sur la période, les événements déclarés doivent être représentatifs de la période couverte, alors que si la troncature porte sur le nombre d'événements, les événements déclarés peuvent être représentatifs seulement de l'ensemble des événements sur une courte période antérieure à l'enquête, et ceci va compliquer toute évaluation de la qualité de la suite des événements au fil du temps.

Comme pour l'histoire génésique complète, la première étape doit être un examen des données en valeur manquante. La deuxième étape doit impliquer l'examen des rapports de masculinité à la naissance et de l'attraction de certains âges au décès.

Il n'est pas possible d'envisager directement l'existence de transfert des naissances, puisqu'on ne dispose pas d'information détaillée sur les dates de naissance antérieures au point de troncature. Mais une évaluation indirecte est possible. Une histoire génésique tronquée devrait toujours commencer par le recueil d'une histoire génésique résumée. Les naissances et les décès de jeunes enfants pour les femmes d'un groupe d'âge à l'enquête peuvent donc être calculés à la fois à la date de l'enquête (à partir de l'histoire génésique résumée) et (avec une approximation pour les décès) à la date de troncature, en retranchant les naissances et les décès déclarés dans l'histoire génésique tronquée. Le calcul est précis pour les naissances, mais il est approximatif pour les décès de jeunes enfants car certains de ceux qui ont été déclarés dans l'histoire résumée se seront produits dans la période d'après troncature tout en concernant des enfants nés avant celle-ci ; mais le nombre de ces décès sera généralement faible, du fait que les risques de mortalité diminuent rapidement à mesure que les enfants avancent en âge. L'évaluation de la qualité des données s'appuie donc sur la comparaison entre la proportion d'enfants décédés (par groupe d'âge de la mère au moment de l'enquête) parmi les enfants nés après la date de troncature et celle parmi les enfants nés avant cette date.

La première proportion (enfants nés après la date) sera généralement inférieure à la seconde pour deux raisons. Tout d'abord, les enfants auront été exposés moins longtemps au risque de décéder. Ensuite, si la mortalité diminue au fil des années, ils auront aussi été exposés à des risques moindres aux différents âges. Mais si les enfants décédés sont systématiquement omis de la période post-troncature, ou s'ils sont déclarés dans l'histoire génésique résumée mais pas déclarés comme étant nés au cours de la période, le ratio des deux sera surestimé du fait des erreurs dans les données. Nous pouvons estimer un ratio vraisemblable en l'absence d'erreur à partir d'une histoire génésique complète pour la même population à une date antérieure ou postérieure. Le tableau 17.6 présente des données pour la Mongolie, tirées de trois enquêtes sur la santé reproductive, une en 1998 qui incluait une histoire génésique complète et deux – en 2003 et 2008 – qui n'ont recueilli que des histoires tronquées. Les données de l'histoire génésique complète en 1998 sont utilisées pour calculer les proportions d'enfants décédés avant et après une date de troncature définie comme dans les enquêtes ultérieures ; les résultats sont ensuite comparés aux proportions calculées à partir des histoires tronquées en 2003 et 2008. Comme on peut le voir, les ratios tirés des histoires tronquées sont largement supérieurs à ceux tirés des histoires complètes, ce qui démontre clairement l'existence d'un transfert des enfants décédés hors de la période post-troncature. En l'absence d'une référence propre au pays, comme celle que fournit l'enquête de 1998 pour la Mongolie, on doit tenir des ratios de 3 ou plus comme preuve d'une omission probable d'enfants décédés au cours de la période de référence récente.

L'histoire génésique tronquée : Calcul des indicateurs de mortalité des jeunes enfants par cohorte

Le calcul des probabilités de décès par cohorte à partir d'histoires génésiques tronquées suit le même principe que celui observé avec des histoires génésiques complètes : la probabilité de décéder avant l'âge x est égale au rapport entre le nombre d'enfants décédés et le nombre d'enfants nés vivants dans une cohorte née au moins x années avant l'enquête. Il y a toutefois une différence importante, clairement mise en évidence sur le diagramme de Lexis de la figure 17.1 : la valeur de x est contrainte par la date de troncature. Par exemple, si la date de troncature se situe 5 ans avant l'enquête, aucune cohorte de naissance n'aura été pleinement exposée au risque complet de décéder avant l'âge de 5 ans,

et les cohortes pleinement exposées au risque jusqu'à l'âge de 2 ans seront constituées des naissances survenues 2, 3 et 4 ans avant l'enquête. Il y a donc des limites aux âges pour lesquels des indicateurs de mortalité peuvent être calculés.

L'histoire génésique tronquée : Calcul des indicateurs de mortalité des jeunes enfants par période

Fondamentalement, le calcul des indicateurs classiques à partir d'histoires génésiques tronquées suit les mêmes principes que lorsqu'on utilise des histoires génésiques complètes : calculer des taux par âge pour une période donnée, les convertir en estimations des probabilités de décéder dans des intervalles d'âge successifs, et appliquer les probabilités à une cohorte synthétique de naissance pour créer une

table de mortalité. Le problème dans ce type d'analyse des histoires tronquées est le même que celui rencontré lors du calcul des indicateurs de cohortes, puisque le nombre de cas et le temps d'exposition au risque se réduisent progressivement à mesure que l'âge augmente. Ainsi, si le point de troncature se situe cinq ans avant l'enquête, les mesures aux âges de 3 et 4 ans s'appuieront sur de petits nombres et elles seront entourées d'une large marge d'erreur.

RÉFÉRENCES

- Preston SH, P Heuveline and M Guillot. 2001. *Demography: Measuring and Modelling Population Processes*. Oxford: Blackwell.
- Rutstein S and G Rojas. 2003. *Guide to DHS Statistics*. Calverton, MD: ORC Macro.

Chapter 18 ESTIMATION DE LA MORTALITÉ DES JEUNES ENFANTS À PARTIR DES DONNÉES DES ÉTABLISSEMENTS DE SANTÉ : LA TECHNIQUE DE LA NAISSANCE PRÉCÉDENTE

Allan Hill

CONTEXTE

Dans le chapitre d'introduction à l'analyse de la mortalité des jeunes enfants (chapitre 15), nous avons déjà attiré l'attention sur la possibilité d'utiliser les données collectées dans les établissements de santé pour mesurer la mortalité de la petite enfance. Si on met pour l'instant à part les questions de sélection, la technique la plus utile pour obtenir ces mesures de la mortalité s'appuie sur une question simple posée à une mère qui attend un prochain enfant ou qui le met au monde. Nous désignerons cet enfant comme l'enfant de référence (en anglais *index child*). Il s'agit de savoir si la précédente naissance vivante est encore en vie ou décédée au moment où sa mère est interrogée pendant la grossesse ou lors de l'accouchement de l'enfant de référence. Pour un ensemble de mères (généralement au moins un millier de répondantes), les proportions d'enfants décédés parmi les enfants nés vivants précédents sont d'abord calculées puis converties en une mesure de la mortalité de la petite enfance. Cette mesure est souvent voisine de ${}_2q_0$ mais elle peut se rapprocher davantage de ${}_3q_0$ si les intervalles entre naissances sont longs. Des variantes de la méthode peuvent consister à obtenir les proportions d'enfants précédents décédés avant l'accouchement de l'enfant de référence, par exemple au moment d'une visite prénatale, ou après l'accouchement de l'enfant de référence, pour les convertir en mesure de la survie des jeunes enfants.

L'information sur la survie des enfants recueillie de cette façon auprès des hôpitaux, des cliniques ou autres centres de santé présente plusieurs caractéristiques attrayantes. Premièrement, le recours à des enquêtes et études spéciales est évité, car l'information fait souvent partie d'un système sani-

taire d'enregistrement régulier. Deuxièmement, quand l'information provient d'établissements de santé, des informations complémentaires qu'il est difficile d'obtenir dans des enquêtes rétrospectives peuvent être obtenues relativement aisément. Il s'agit de caractéristiques de la mère ainsi que des naissances, dont le sexe, le rang de naissance et le poids à la naissance. Troisièmement, les données peuvent être désagrégées et fournir des estimations détaillées pour certains établissements de santé (quand la population dans le bassin de recrutement est suffisamment nombreuse), pour des villes et de petites provinces. Cette information à un niveau géographique fin peut être précieuse si les autorités de santé veulent cibler les groupes ayant les taux de mortalité infantiles ou juvéniles les plus élevés. Des tendances au niveau local peuvent aussi servir à apprécier l'efficacité qu'ont eue des actions sanitaires. Enfin, les données des établissements de santé étant généralement recueillies par des personnels éduqués, on peut s'attendre à ce qu'elles soient raisonnablement précises. Par ailleurs, des répondantes placées dans un contexte médical peuvent être mieux préparées que lors d'enquêtes à domicile à déclarer des événements qui peuvent être perçus comme stigmatisants (la distinction entre naissances vivantes, mort-nés, avortements et fausses couches) ou pénibles à rappeler (un décès néonatal ou infantile).

ORIGINE DES MÉTHODES BASÉES SUR LA TECHNIQUE DE L'ENFANT PRÉCÉDENT

La technique a ses origines dans une étude sur la mortalité dans les Iles Salomon dans les années 1980. Au cours de cette étude, il est apparu que parmi les informations régulièrement recueillies dans les maternités figuraient des réponses

aux questions sur le nombre d'enfants déjà nés et survivants ainsi qu'une question sur la survie de l'enfant précédent, si la mère accouchait d'un deuxième enfant ou d'un enfant de rang plus élevé, l'enfant de référence (voir figure 18.1). Brass et Macrae se sont efforcés de montrer comment ces données pouvaient être rattachées aux mesures habituelles de la survie des jeunes enfants, estimées à partir d'histoires génésiques résumées et des proportions d'enfants décédés parmi les naissances précédentes (Brass and Macrae 1984, 1985). Deux méthodes ont été proposées. L'intérêt pour la technique de la naissance précédente a été immédiat, car elle fournit une estimation de la mortalité de la petite enfance proche de la période en cours. La base théorique de la méthode a été étendue et développée pour permettre des applications aux données de survie de l'enfant précédent recueillies avant ou après une naissance (Aguirre 1994 ; Aguirre and Hill 1988 ; Hill and Aguirre 1990). La seconde méthode (Brass and Macrae 1985), fondée sur les nombres totaux d'enfants déjà nés et survivants, est plus élaborée et plus compliquée d'application ; elle n'est pas présentée ici.

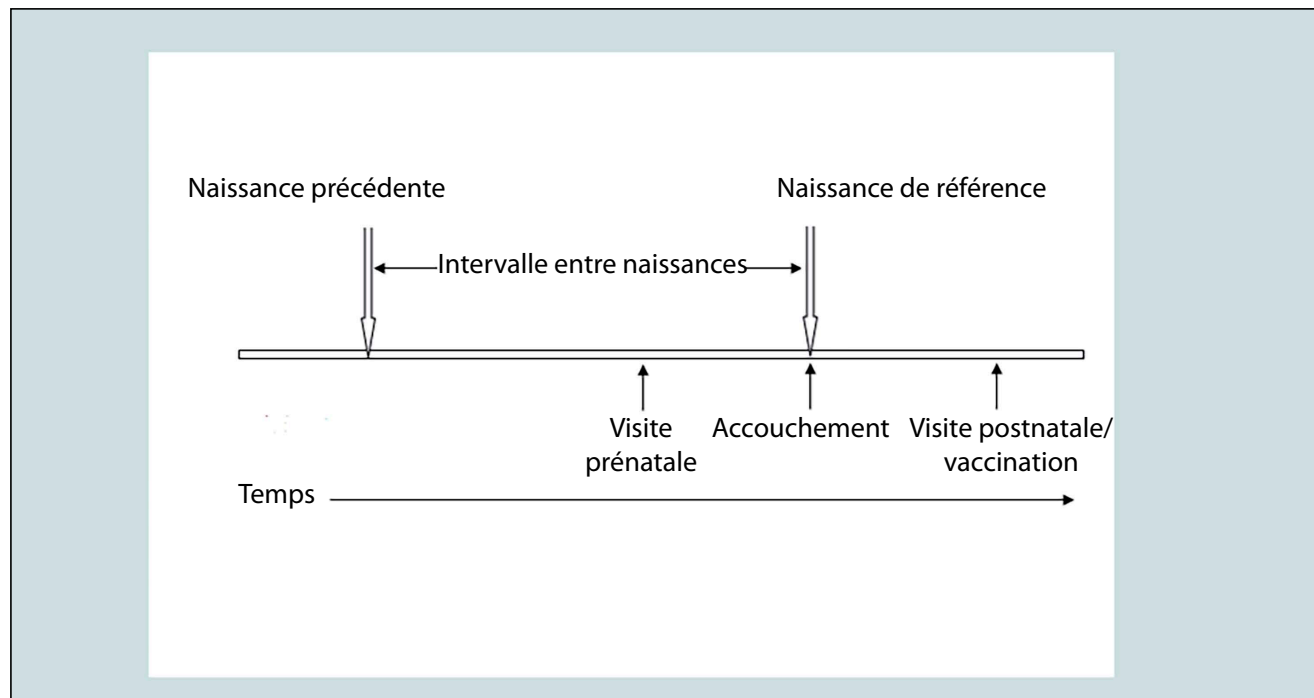
Par la suite, d'autres chercheurs ont appliqué la technique de l'enfant précédent dans diverses circonstances, avec dif-

férents objectifs – par exemple, dans des camps de réfugiés, pour mesurer les taux d'avortement dans des consultations prénatales, pour des estimations sur petits domaines et pour mesurer l'impact d'actions sanitaires (Bicego, Augustin, Musgrave *et al.* 1989 ; Madi 2000 ; Oliveras, Ahiadeke, Adanu *et al.* 2008 ; Rowe, Onikpo, Lama *et al.* 2011). Dans une étude dans un site de surveillance démographique où les naissances et les décès étaient enregistrés de façon fiable et indépendante, on a montré que de bons résultats pouvaient être obtenus, alors que la contraception était largement diffusée et que les intervalles entre naissances s'étaient allongés, aussi bien avec des données recueillies dans des consultations prénatales plutôt qu'en maternité ou au moment de la première vaccination du nouveau-né (Bairagi, Shuaib and Hill 1997).

PRÉCAUTIONS ET MISES EN GARDE

L'analyste est confronté à plusieurs difficultés au moment d'utiliser des données sur la mortalité des jeunes enfants recueillies dans des établissements de santé. La plus importante tient au fait que la population fréquentant ces établissements n'est pas sélectionnée aléatoirement. Les biais qui en résultent sont importants à plusieurs titres.

FIGURE 18.1 Schéma d'un intervalle entre naissances présentant les éventuels points de contact avec les services sanitaires



Le premier biais est dû à la couverture incomplète de la population par les établissements de santé, à la fois publics et privés. Cette sélection fonctionne de différentes façons. Souvent, la population urbaine accède plus facilement aux établissements de santé que la population rurale. Le recours aux services de santé modernes augmente avec le niveau de vie et le niveau d'instruction des personnes. D'où un biais dans les estimations de la mortalité infantile et juvénile, qui tendent à être trop optimistes. Mais dans certains cas, le biais peut jouer à l'inverse. Les centres de référence, qui incluent généralement la maternité centrale ou les hôpitaux universitaires, ont souvent des résultats moins favorables que les centres périphériques, parce que les cas les plus compliqués nécessitant des interventions chirurgicales et d'autres formes de soins avancés leur sont adressés. Les estimations pour de tels établissements tendront donc à surestimer la mortalité infantile et juvénile.

Le premier biais est d'autant moins sensible que les taux de couverture des établissements de santé utilisés comme source de données sont eux-mêmes élevés. En Afrique subsaharienne, d'après les enquêtes EDS les plus récentes dans 38 pays, 51 % des mères ont accouché dans un établissement de santé, 76,5 % des mères en zone urbaine. Les proportions sont encore plus élevées pour les visites prénatales : 93 % des femmes ont été vues par un médecin ou du personnel médical dans les zones urbaines et 76 % dans les zones rurales, une proportion étonnante (Macro International Inc 2012). Grâce à ces taux de couverture élevés, il est possible de traiter certains des biais associés à la couverture incomplète de la population. C'est pourquoi nous présentons plus loin une méthode pour estimer la mortalité de la petite enfance à partir des proportions d'enfants précédents décédés quand les mères se rendent à des consultations prénatales ou même à des centres de vaccination. Ces options sont présentées plus loin dans la section « *Extensions* ».

Quand la couverture de la population par les établissements de santé est beaucoup plus basse, le biais de sélection est nettement plus élevé. On peut utiliser des données d'enquêtes ou du recensement sur l'ensemble de la population pour ajuster les statistiques des établissements de santé, en comparant les caractéristiques des utilisateurs et des non utilisateurs des services de santé. Une complication vient de ce que de nombreux pays s'efforcent d'étendre la couverture de leurs services de santé. Il peut en résulter l'inclusion de nouveaux groupes de population ayant des schémas de mortalité spécifiques dans la base de données sur la survie des jeunes enfants. Ces modifications de couverture peuvent rendre

difficile l'interprétation des évolutions au fil du temps. Toutefois dans la plupart des populations de grande taille, l'installation de nouveaux établissements prend du temps et la couverture n'évolue que lentement.

La seconde source de biais systématique est le fait que presque toutes les femmes qui viennent dans les maternités et les centres de santé le font parce qu'elles sont sur le point d'avoir un enfant. Toutes les informations obtenues de ces femmes sont propres à ces moments de leur vie reproductive. Il en va différemment dans les enquêtes sur des échantillons aléatoires, où les femmes sont interrogées indépendamment de leur position dans leur parcours reproductif, les informations qu'elles donnent étant donc représentatives des déclarations de l'ensemble des femmes ou des femmes de toutes parités. Un ajustement a été proposé pour que les données obtenues dans les établissements de santé sur le nombre total d'enfants déjà nés et décédés se rapprochent davantage des données dans les enquêtes auprès des ménages. Mais l'ajustement semble trop dépendant des diverses hypothèses, comme l'effet du rang de naissance sur la survie des enfants et la répartition dans le temps des décès des enfants (Brass and Macrae 1985).

Un troisième biais de sélection avec cette méthode provient du fait que la dernière naissance de chaque femme n'est jamais représentée, puisqu'elle n'a pas l'occasion d'être déclarée lors de la naissance suivante. Le biais est probablement négligeable quand la fécondité est élevée, mais dans une population où de nombreuses femmes n'ont que deux enfants, il y aura une surreprésentation des premières naissances, qui ont généralement une mortalité supérieure à la moyenne. Dans une population où une importante proportion de femmes n'a qu'un enfant, la technique de la naissance précédente ne donnera pas non plus des estimations sans biais de la mortalité de la petite enfance.

DESCRIPTION DÉTAILLÉE DE LA MÉTHODE

Un diagramme aide à comprendre les termes utilisés dans cette explication. La figure 18.1 schématise un intervalle entre naissances \bar{I} pour une femme ayant trois points de contact possibles avec les services sanitaires – à l'occasion d'une visite prénatale, au moment de l'accouchement et après l'accouchement lors d'un contrôle post-partum ou d'une vaccination du nouveau-né.

Des modèles et des données empiriques ont montré que les proportions d'enfants décédés parmi les enfants précédents, Q , recueillies à proximité de l'accouchement de l'enfant de

référence sont proches de la probabilité de décéder avant un âge de l'ordre de 80 % de l'intervalle moyen entre naissances vivantes, \bar{l} . La compilation des intervalles entre naissances médians pour 35 des enquêtes les plus récentes couvrant la période 1990-2010 dans des pays d'Afrique subsaharienne fait ressortir un intervalle médian de 34,8 mois (Macro International Inc 2012). Les seuls pays avec des intervalles médians de plus de 40 mois sont le Ghana (40), la Namibie (42), l'Afrique du Sud (47) et le Zimbabwe (47). Les proportions d'enfants précédents décédés au moment de l'accouchement de l'enfant de référence sont donc proches des proportions d'enfants décédés avant leur deuxième anniversaire, soit, dans une notation de table de mortalité, ${}_2q_0$.

La proportion d'enfants précédents décédés avant la naissance de référence est très proche de ${}_2q_0$ dans une table de mortalité, parce que la proportion d'enfants décédés est l'intégrale du produit de deux fonctions asymétriques. L'une est la distribution des naissances par durée avant la naissance la plus récente, $b(x)$. La seconde fonction est la probabilité cumulée $q(x)$, prise dans les premiers âges couverts par la table de mortalité. La distribution par mois de la naissance précédente est asymétrique : aucune naissance dans les neuf premiers mois précédant la maternité actuelle mais une concentration des naissances autour de l'intervalle moyen entre naissances, puis une longue queue de distribution à mesure que l'on remonte dans le temps avant la naissance la plus récente. Les probabilités cumulées de décéder aux jeunes âges dans toute table de mortalité s'accroissent rapidement durant les deux premières années de vie, mais par la suite les proportions cumulées, q_x , se stabilisent au-delà de l'âge de deux ans (Hill and Aguirre 1990).

La figure 18.2 illustre la forme de ces deux fonctions avec des données réelles d'intervalles entre naissances (d'où les légères irrégularités dues à des défauts de déclaration des dates) et les probabilités de décéder par mois depuis la naissance tirées de la table-type des Nations Unies (modèle « Général ») avec une espérance de vie à la naissance de 60 ans.

La proportion d'enfants décédés parmi les naissances précédentes, Q , est l'intégrale du produit de ces deux fonctions. Soit

$$Q = \int_0^{\infty} b(x)q(x)dx \quad (1)$$

où $b(x)$ est le nombre de naissances survenues x mois avant la maternité actuelle, et $q(x)$ est la probabilité cumulée des décéder avant l'âge x .

Comme la méthode vise essentiellement à mesurer les variations d'une année à l'autre de la survie des jeunes enfants par établissement, l'intérêt porte davantage sur la valeur de l'indice que sur sa représentation exacte dans la table de mortalité. Le principal effet perturbateur dans la comparaison des résultats au fil du temps ou dans différentes populations résultera de différences dans la longueur des intervalles entre naissances. Mais l'analyse de Rutstein a montré que les intervalles médians n'ont évolué que très lentement entre les premières et les dernières EDS dans chaque pays et en particulier en Afrique subsaharienne, de sorte que l'importance des variations de la longueur des intervalles entre naissances est probablement mineure (voir Rutstein (2011 : Tableau 2.2a)). Par convention, on dénomme simplement « indice de mortalité de la petite enfance » la proportion d'enfants précédents décédés au moment de la naissance de l'enfant de référence, et on le considère comme une approximation de ${}_2q_0$ dans la plupart des populations et de ${}_{2,7}q_0$ dans les populations où les intervalles entre naissances sont de l'ordre de 40 mois. Dans de nombreux pays à basse fécondité, les intervalles entre naissances sont plus longs, mais la plupart de ces pays ont de bons systèmes d'enregistrement des mouvements de population et ils ne sont pas les principaux utilisateurs de la technique de la naissance précédente.

On peut estimer les effets éventuels d'intervalles entre naissances s'écartant de 30 mois sur les mesures de la mortalité des jeunes enfants en recourant aux tables-types de mortalité. Nous utilisons celles du modèle Général des Nations Unies, sexes réunis, avec une espérance de vie à la naissance de 60 ans, pour calculer les probabilités mensuelles de décéder jusqu'à 5 ans. Nous calculons ensuite les différences relatives dans la mesure de la mortalité de la petite enfance lorsque l'intervalle entre naissances, l , s'écarte de 30 mois et que le temps d'exposition au risque n'est donc pas $0,8 \cdot l$ ou 24 mois (${}_2q_0$). Quand l'intervalle entre naissances est de 25 mois et donc le temps d'exposition au risque de $25 \cdot 0,8 = 20$ mois, ce qui peut survenir si les données sont recueillies lors de visites prénatales, les proportions d'enfants précédents décédés sont alors proches de ${}_{1,7}q_0$ ou 5 % plus basses que si l'intervalle entre naissances avait été de 30 mois. Quand les intervalles entre naissances atteignent 40 mois, les proportions d'enfants précédents décédés sont une approximation de ${}_{2,7}q_0$, soit un écart de 7 % avec la valeur de ${}_2q_0$ associée à un intervalle entre naissances de 30 mois. Si l'intervalle entre la naissance de l'enfant précédent et le moment où les don-

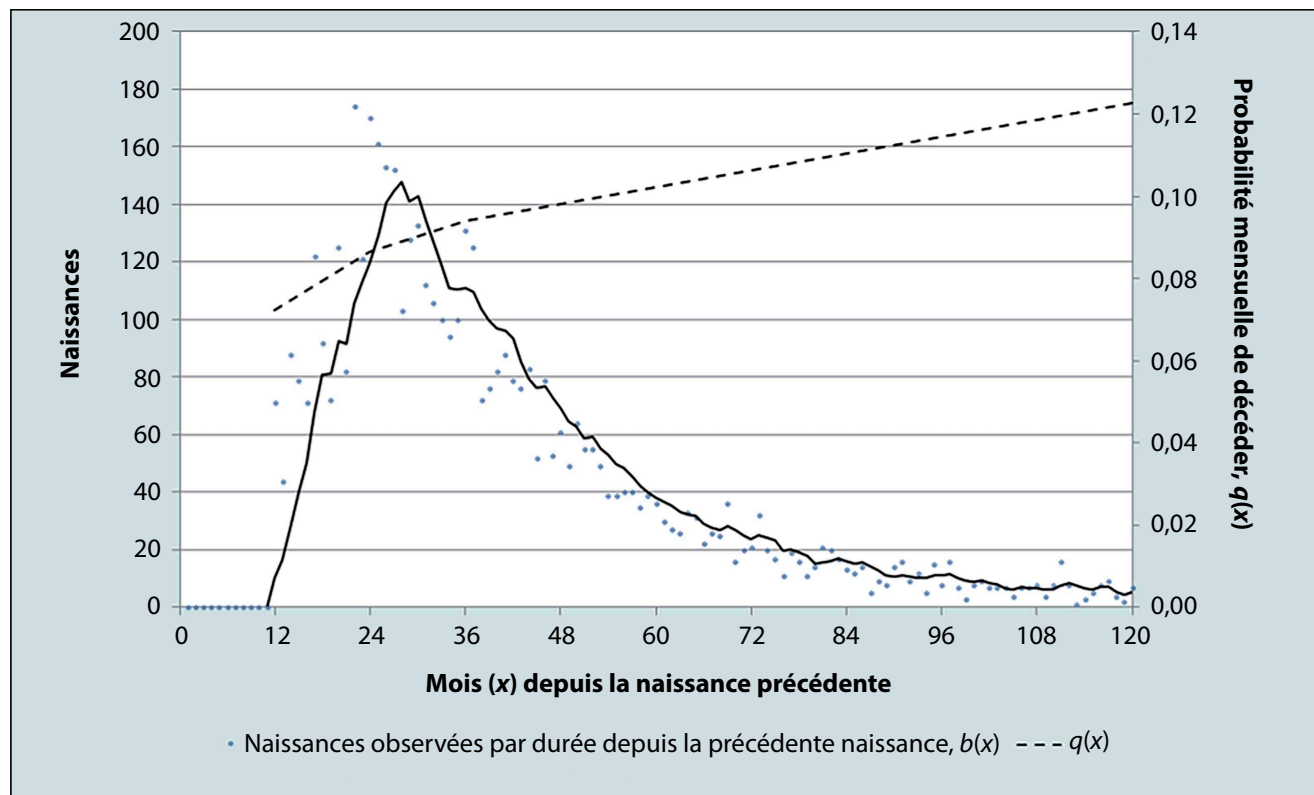
nées sont recueillies atteint 45 mois, ce qui peut arriver si les mères sont interrogées quelque temps après la naissance de l'enfant de référence, alors la mesure estimée de la mortalité des jeunes enfants est à peu près égale à ${}_3q_0$ soit 9 % de plus que ${}_2q_0$ dans la table type de mortalité. Bien qu'importantes, ces différences restent modérées et les pourcentages restent vraisemblablement stables à court terme.

Les tables-types de mortalité peuvent être utilisées pour interpoler entre les diverses mesures de la mortalité des jeunes enfants obtenues par la technique de la naissance précédente. Avec des intervalles entre naissances de 30 mois, nous estimons à peu près ${}_2q_0$. En utilisant des transformations logit des tables-types de mortalité, on peut aisément obtenir les valeurs correspondantes de ${}_1q_0$ (mortalité infantile) et ${}_5q_0$ (taux de mortalité avant 5 ans utilisé par l'Unicef). Un exemple de la méthode d'interpolation figure au tableau 18.2. La même procédure peut être suivie pour obtenir ${}_2q_0$ quand les intervalles entre la naissance précédente et la naissance de référence ne sont pas de 30 mois, en

particulier si l'intervalle est réduit du fait de la collecte des données lors des visites prénatales, ou si la durée depuis la naissance précédente est au contraire allongée du fait de la collecte des données est organisée, disons, lors de la première vaccination de l'enfant de référence. La feuille de calcul disponibles sur <http://demographicestimation.iussp.org/fr> présente la procédure en détail.

Les questions soulevées par ce simple résultat qui veut que $Q \approx (1.0, 8) q_0$ sont nombreuses. La première question concerne l'omission des femmes n'ayant qu'une naissance et donc pas de naissance précédente. Dans une large majorité des populations sans système d'état civil complet, la plupart des femmes ont au moins un deuxième enfant, de sorte que la mortalité des premières naissances est bien prise en compte, le biais étant donc minime. Toutefois, dans les populations à faible fécondité, la question se pose de la surreprésentation des premières naissances dans les données. Deuxièmement, les femmes qui décèdent en couches peuvent ne pas survivre assez longtemps pour déclarer la survie de leurs enfants pré-

FIGURE 18.2 Distributions des intervalles entre naissances par durée depuis la naissance précédente et risques cumulés de décéder avant l'âge (x)



cédents, même si des informations sur les dossiers hospitaliers restent souvent disponibles au-delà du décès. On sait que les risques de perdre l'enfant de référence sont fortement associés à la mortalité maternelle, comme le montrent les données de Bamako (Hill and Aguirre 1990). Heureusement les décès maternels sont suffisamment rares pour avoir peu d'effets sur les données recueillies de cette façon.

Il est important de préciser la période de référence des estimations obtenues par la technique de la naissance précédente. À nouveau, la figure 18.2 nous aide à évaluer la date moyenne des décès des enfants précédents qui sont morts avant la naissance de référence. La combinaison de risques relativement élevés de décéder précocement (voir la fonction $q(x)$) et d'une concentration des naissances autour de l'intervalle moyen entre naissances oriente vers un âge moyen au décès nettement inférieur à la moitié de l'intervalle entre naissances. Aguirre (1990) a montré à l'aide de modèles que la date moyenne de décès des enfants précédents se situait environ aux deux tiers de la longueur de l'intervalle entre naissances avant la date de naissance de l'enfant de référence. À partir de données empiriques couvrant un large éventail d'intervalles entre naissances, le décès de l'enfant précédent se situait à un point entre 54 et 74 % de l'intervalle en mois avant la date de naissance de l'enfant de référence. Dans la plupart des applications, nous recommandons de considérer que la date de référence des décès se situe aux deux tiers de l'intervalle avec la naissance précédente à partir de la naissance de l'enfant de référence. Cette hypothèse a été utilisée dans les feuilles de calcul des estimations.

Données nécessaires

La série de questions nécessaires pour la mise en œuvre de cette technique est simple, en supposant que la femme est interrogée dans l'établissement de santé alors qu'elle est enceinte (visite prénatale), qu'elle vient d'accoucher (dans une maternité ou à l'hôpital) ou qu'elle a amené son nouveau-né pour une vaccination :

- « Avez-vous déjà été enceinte une autre fois avant la grossesse/naissance actuelle ? »
Si « oui », continuer. Si « non », stop.
- « Quelle a été l'issue de cette précédente grossesse ? » (naissance vivante, mort-né, fausse couche ou avortement – spontané ou provoqué).
Si « naissance vivante » continuer. Si autre, stop.
- « Cet enfant précédent est-il encore en vie aujourd'hui ? » (oui/non)

La date de l'interview (généralement la date d'accouchement de l'enfant de référence) est également nécessaire, mais elle fait généralement partie des renseignements administratifs connus.

Il s'agit là des questions de base, mais il s'en ajoute souvent d'autres relatives à la mère et à ses enfants, telles que la date de naissance de l'enfant précédent, son sexe, son poids à la naissance (s'il est connu), si c'était une naissance simple ou multiple, s'il est encore allaité entièrement, etc., selon les besoins et les circonstances. De même, d'autres informations peuvent être utiles pour les soins à la mère et à l'enfant (si la mère est vue après l'accouchement) : l'âge, le niveau d'instruction et la résidence de la mère ainsi que des informations sur l'accouchement de l'enfant de référence tels que son poids à la naissance et le lieu et le type d'accouchement (normal, au forceps, par aspiration, par césarienne, etc.). La connaissance de la date de naissance de l'enfant précédent est utile pour le calcul des intervalles moyens entre naissances dans la population étudiée.

Les données sont souvent obtenues dans les hôpitaux sous la forme de registres ou de grands livres. Un exemple est donné au tableau 18.1. Le contenu de chaque colonne peut varier en fonction des objectifs mais les questions essentielles pour l'estimation de la mortalité de la petite enfance sont reprises dans les colonnes (7) et (8). Notez que dans le système sanitaire, le nombre de grossesses (aussi courtes soient-elles) remplace souvent le nombre de naissances vivantes, mais avec un peu d'accoutumance et de formation le personnel de l'établissement peut aisément distinguer entre les définitions plus médicales (gravidité, parité, parturition) et des termes plus démographiques (grossesses, naissances vivantes et enfants vivants).

EXEMPLE

Dans sa forme basique l'analyse est très simple. Elle consiste à diviser le nombre d'enfants précédents décédés (ligne B du tableau 18.2) au moment de la naissance de référence par le total du nombre d'enfants précédents nés vivants. Les mort-nés sont exclus du calcul.

Dans les cas où l'on souhaite estimer d'autres indices de la table de mortalité, la mortalité infantile et la mortalité avant 5 ans peuvent être obtenues en utilisant des tables-types de mortalité et des transformations logit. Au tableau 18.3, nous illustrons les étapes de la procédure pour produire des estimations de ${}_1q_0$ et ${}_5q_0$ en utilisant des transformations logit de tables-types de mortalité. Ces méthodes sont incluses dans les feuilles de calcul associées.

TABEAU 18.1 Exemple de registre de recueil des données pour la mise en œuvre de la technique de la naissance précédente

| Date d'accouchement (1) | Nom de la mère ou N° d'identité (2) | Âge ou date de naissance de la mère (3) | Gravidité (Total des grossesses) (4) | Enfants nés vivants (5) | Enfants vivants (6) | Accouchement actuel : type (né vivant, mort né, avortement ou fausse couche) (7) | Naissance simple ou multiple ? (8) | Si né vivant : précédent enfant vivant aujourd'hui ? (9) | Sexe du dernier accouchement : m/f (10) |
|----------------------------|--|--|---|----------------------------|------------------------|---|---------------------------------------|---|--|
| 27 Jan 2012 | Mariama Sow | 31 Oct 1980 | 7 | 5 | 4 | Né vivant | Simple | Oui | M |
| 28 Jan 2012 | Comfort Frempong | 27 juin 1991 | 3 | 3 | 2 | Mort né | N.A. | N.A. | N.A. |
| 29 Jan 2012 | Huda Khalaf | 19 Oct 1992 | 3 | 2 | 2 | Né vivant | Jumeaux | 1 – Oui | M |
| 29 Jan 2012 | Huda Khalaf | 19 Oct 1992 | 3 | 2 | 2 | Né vivant | Jumeaux | 2 – Non | F |
| 30 Jan 2012 | Mary Kenyatta | 22 ans | 3 | 2 | 1 | Fausse couche | N.A. | N.A. | N.A. |

Note : Dans cet exemple, des mots ont été utilisés plutôt que des codes, mais dans la plupart des cas, les réponses au choix sont pré-codées afin de standardiser les réponses et de minimiser le travail du personnel médical. Les nombres de grossesses, de naissances vivantes et d'enfants vivants excluent la grossesse ou la naissance la plus récente. Les jumeaux ou les triplés doivent être enregistrés séparément – voir le tableau.

Dans certains cas, des informations sont aussi recueillies et utilisées concernant l'avant-dernier enfant né, c'est-à-dire l'enfant né avant la précédente naissance (s'il y a eu un avant-dernier enfant). En simplifiant, la période d'exposition au risque de décéder pour cet enfant sera légèrement

TABEAU 18.2 Estimation de la mortalité de la petite enfance par la technique de la naissance précédente pour Bamako, Mali en 1985

| Mesures | Naissances précédentes | Avant-dernières naissances |
|--|------------------------|----------------------------|
| Total des enfants précédents nés vivant (A) | 4 778 | 3 737 |
| Total des décédés parmi les enfants précédents nés vivants (B) | 679 | 620 |
| Proportions d'enfants précédents décédés (B/A) | 0,142 | 0,166 |

Source : Données de l'étude sur la maternité de Bamako (Hill and Aguirre 1990).

inférieure à deux fois l'intervalle moyen entre naissances, \bar{I} . En prenant $\bar{I} = 30$ mois, la proportion d'avant-derniers enfants décédés au moment de l'accouchement actuel sera à peu près égale à 2.190 , soit 590 . La proportion déclarée d'avant-derniers enfants décédés est proche de la probabilité de décéder dans leurs cinq premières années de vie et pas à un âge plus jeune, car la distribution mensuelle des naissances et les probabilités cumulées de décéder sont beaucoup plus plates autour de l'âge de cinq ans qu'autour de deux ans. La difficulté est que ces données sur l'avant-dernier enfant ne peuvent provenir que de mères ayant eu au moins trois accouchements ou au moins deux accouchements et une troisième grossesse. La sélection systématique de femmes ayant un nombre d'enfants élevé et probablement une fertilité forte (intervalles entre naissances courts) exacerbe les biais par rapport à la version simple de la méthode (Hill and Aguirre 1990). Bien qu'elle soit incluse dans l'illustration ci-dessous, nous ne recommandons pas l'utilisation de l'information sur la survie de l'avant-dernier enfant né pour estimer la mortalité récente des jeunes enfants.

Sur la figure 18.3, nous faisons apparaître la relation entre les différentes mesures de mortalité estimées au tableau 18.3. À première vue, les données suggèrent une amélioration de la survie des enfants dans la période précédant la collecte des données dans les maternités de Bamako. Mais il faut se

rappeler les effets de sélection implicites dans l'utilisation des données sur les avant-derniers enfants, puisque seules les mères d'au moins trois enfants fournissent les informations utilisées pour estimer les mesures repris à la dernière ligne du tableau 18.3.

FIGURE 18.3 Relation entre les proportions d'enfants précédents et d'avant derniers enfants décédés, Q , et les indices de mortalité des jeunes enfants ${}_1q_0$ and ${}_5q_0$

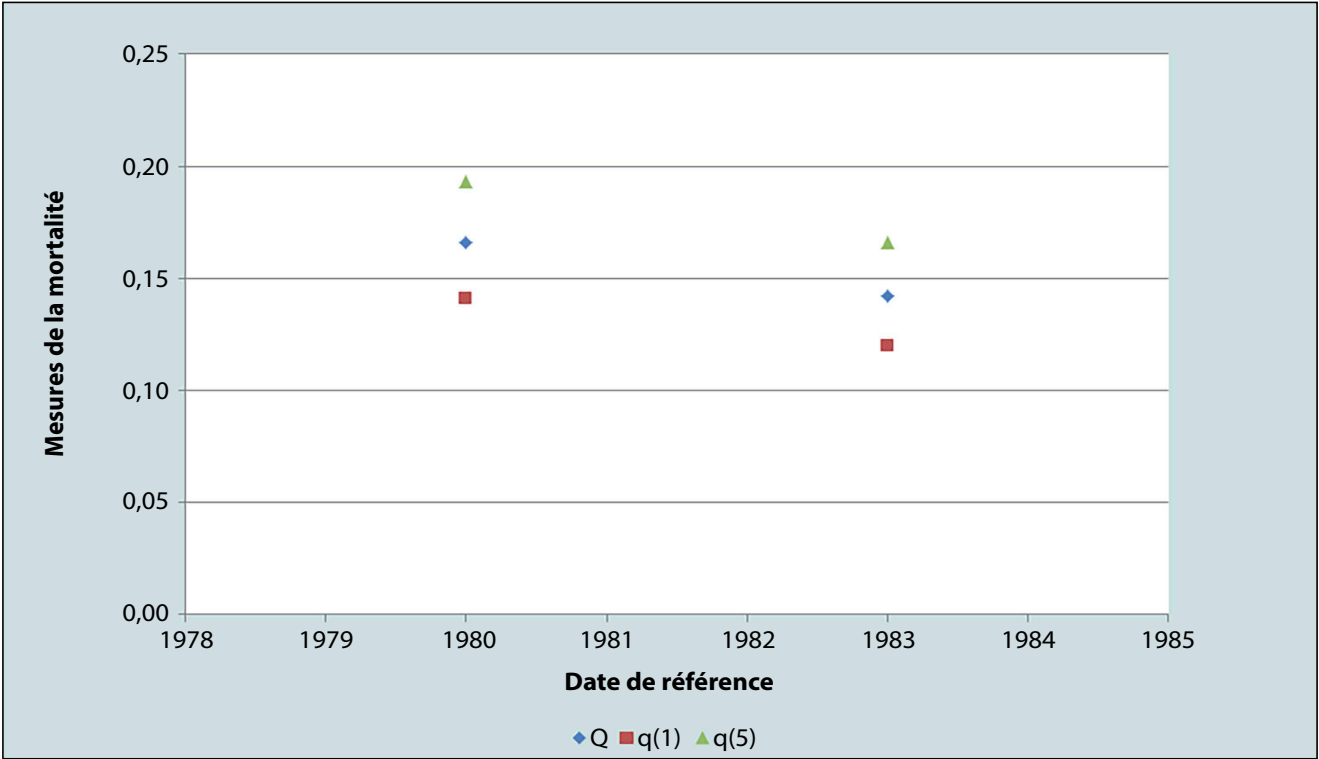


TABLEAU 18.3 Utilisation des logits et des tables types des Nations Unies (modèle Général) pour estimer les taux de mortalité infantile et de mortalité avant 5 ans à partir des proportions d'enfants précédents et d'avant derniers enfants décédés avant l'accouchement de référence

| Mesure | Proportions observées d'enfants décédés | Proportions d'enfants vivants | $l(2)$ dans les tables-types des NU Général: $e(0) = 60$ | logit $l(2)$ | Logit $l(2)$ observé | Alpha | ${}_1q_0$: mortalité infantile estimée | ${}_5q_0$: mortalité avant 5 ans estimée |
|--|---|-------------------------------|--|--------------|----------------------|-------|---|---|
| Proportion d'enfants précédents décédés | 0,142 | 0,858 | 0,914 | - 1,179 | - 0,899 | 0,280 | 0,120 | 0,166 |
| Proportions d'avant derniers enfants décédés | 0,166 | 0,834 | 0,914 | - 1,179 | - 0,807 | 0,372 | 0,141 | 0,193 |

Source : Étude dans les maternités de Bamako (Hill and Aguirre 1990).

Extensions de la méthode

Une critique courante de la méthode de l'enfant précédent est que les taux sont inévitablement calculés pour la population fréquentant les centres de santé et les maternités, mais qu'ils ne nous disent rien sur les taux de mortalité des jeunes enfants pour les mères ne fréquentant pas ces centres. Comme on l'a vu, dans la plupart des pays, des mères de plus en plus nombreuses accouchent dans des centres de santé d'un type ou l'autre, de sorte que peu à peu les estimations par la technique de l'enfant précédent deviennent plus représentatives. Entre temps, la recherche d'informations essentielles via d'autres sources vaut souvent la peine.

L'occasion la plus évidente de contacter une proportion plus importante de mères survient au moment de la première visite prénatale. Comme on l'a vu, la proportion de femmes enceintes qui fréquentent de tels centres est maintenant assez élevée, même là où l'enregistrement à l'état civil est très lacunaire (et l'accouchement en établissement de santé peu fréquent). Il suffit de très peu d'ajustements pour adapter aux consultations prénatales la collecte des informations essentielles pour la technique de la naissance précédente. Le principal risque est que l'information soit recueillie plusieurs fois auprès de la même mère. On doit donc être attentif à ce que l'information soit recueillie à la première visite mais pas aux suivantes.

Le principal problème technique à résoudre est que, comparé au moment de l'accouchement, l'intervalle depuis la naissance précédente est raccourci quand l'information est obtenue lors d'une visite prénatale. La proportion d'enfants décédés parmi les enfants précédents sous-estimerait donc légèrement la vraie valeur de ${}_2q_0$ dans la population. Toutefois, dans de nombreux cas, les femmes enceintes ne se présentent à la première visite prénatale que lorsque la grossesse est bien avancée. Dans ce cas, la réduction de la longueur de l'intervalle entre naissances sera de quelques mois et l'effet sur la mesure de la mortalité de la petite enfance sera très limité (Bairagi, Shuaib and Hill 1997 ; Hill and Aguirre 1990).

Une autre façon d'obtenir l'information nécessaire à l'application de la technique de l'enfant précédent consiste à inclure les questions lors de la première vaccination du bébé de référence, puisque le taux de couverture des opérations de vaccination est très élevé, aussi bien pour les femmes qui ont accouché en milieu hospitalier et celles qui l'ont fait à domicile. Pour l'essentiel, le format des questions reste le même. La formulation et la formation doivent être soignées afin de bien distinguer le nouveau-né qui reçoit la vaccination

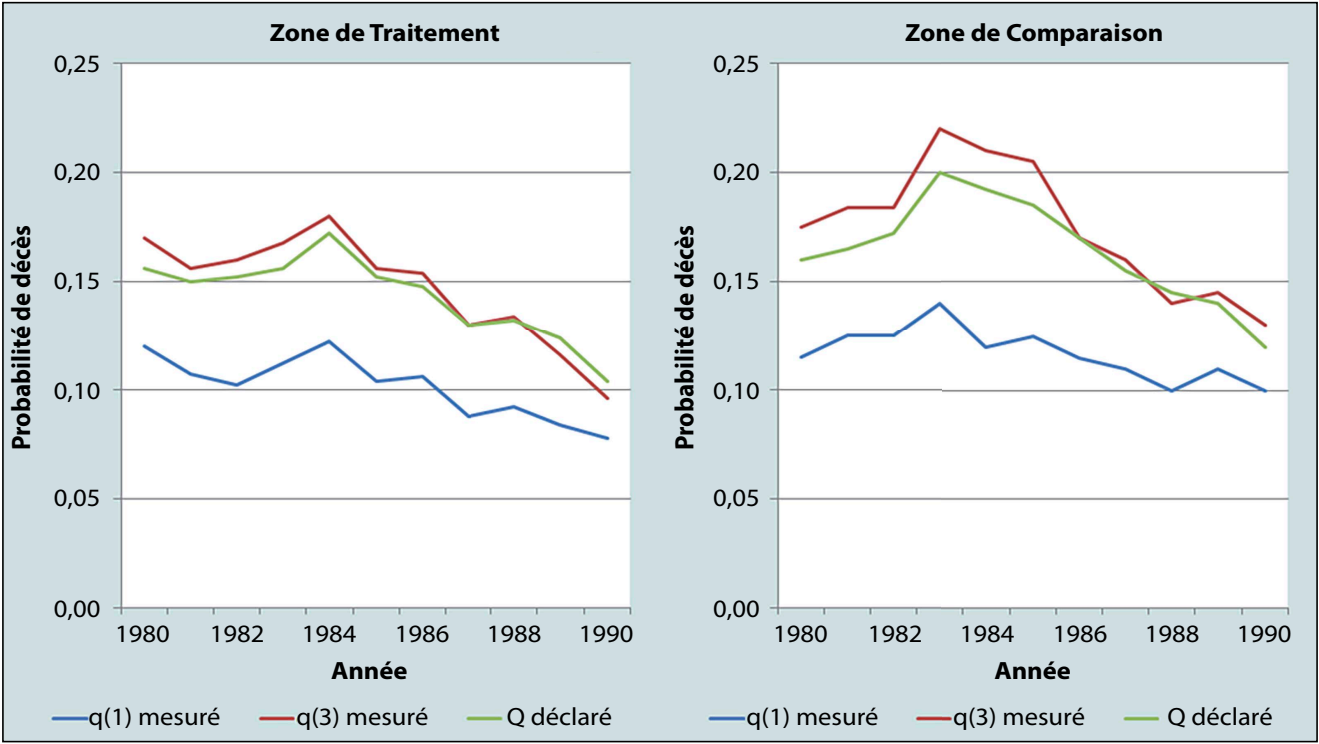
de la naissance précédente dont la survie doit être établie, pour que la technique de la naissance précédente puisse être appliquée. Comme le programme de vaccination débute – ou devrait débiter – dans les premiers mois de la vie du nouveau-né, l'extension de l'exposition de l'enfant précédent au risque de décéder est modeste par rapport à l'exposition jusqu'à l'accouchement. C'est pourquoi les proportions d'enfants précédents décédés au moment de la vaccination peuvent être considérées comme une bonne approximation de ${}_2q_0$. Il convient d'éviter que des mères soient interrogées plusieurs fois à l'occasion de plusieurs séries de vaccination, mais c'est un problème relativement facile à gérer (Hill and Kelly 1996). L'effet de la mortalité du nouveau-né est plus important, car les mères qui perdent leur enfant en bas âge ne participeront pas aux programmes de vaccination. Si la mortalité des enfants successifs n'est pas indépendante, il en résultera une sous-estimation globale de l'indice de mortalité de la petite enfance.

Évolution au fil du temps et comparaisons entre zones

La technique de la naissance précédente trouve une importante application dans l'étude des tendances de la mortalité de la petite enfance dans des sous-populations suivies au fil du temps, et dans la comparaison des tendances de la mortalité des jeunes enfants au fil du temps dans différentes sous-populations. Dans plusieurs pays, du Sénégal et du Mali au Sultanat d'Oman, les estimations courantes de la mortalité de la petite enfance sont tirées d'informations collectées dans des maternités. L'exemple ci-dessous est tiré du site de surveillance sanitaire de Matlab Thana au Bangladesh, où nous bénéficions en outre des données d'un système d'état civil performant, en même temps que des informations sur la survie de l'enfant précédent recueillies au moment de la naissance de référence. Dans la zone d'étude de Matlab Thana il est, de plus, possible de comparer la mortalité des jeunes enfants dans les zones « Traitement » et « Comparaison », ce qui offre une occasion supplémentaire de tester la validité des informations fournies par la méthode.

Sur la figure 18.4 les courbes des zones « Traitement » et « Comparaison » représentent l'évolution de la mortalité de la petite enfance, tirée des questions sur la survie de l'enfant précédent, recueillies au moment de la naissance de référence. Les courbes inférieures représentent l'évolution de la mortalité infantile tirée du système d'état civil, alors que les courbes supérieures correspondent au ${}_3q_0$, également calculé à partir des données d'état civil. La mesure de la mortalité

FIGURE 18.4 Proportions d'enfants précédents décédés au moment de l'accouchement de référence, Q , comparées à la mortalité infantile, $1q0$ et à la mortalité $3q0$, mesurées directement à partir des données de surveillance du Matlab, Bangladesh



de la petite enfance reproduit de très près l'évolution de $3q0$ puisque dans cette population, où les intervalles entre naissance sont de l'ordre de 40 mois, la technique de la naissance précédente mesure la mortalité jusque vers 2,7 ans soit $2,7q0$. La qualité de l'ajustement entre les mesures directes de la mortalité des jeunes enfants et les proportions d'enfants précédents décédés est encourageante.

Comparaison des tendances de la mortalité dans des sous-populations

La dernière illustration provient de l'étude sur la mortalité de la petite enfance dans les maternités de Bamako, Mali, où l'information sur le poids à la naissance du dernier né (l'enfant né à la clinique) était aussi enregistrée. Comme le montre le tableau 18.4, il existe une corrélation très forte entre le poids à la naissance de l'enfant de référence, le dernier né, et la survie de l'enfant précédent. L'analyse fait clairement apparaître la concentration des risques élevés chez certaines mères et montre l'intérêt de cibler ces mères pour

TABLEAU 18.4 Proportions d'enfants précédents et d'avant-derniers enfants décédés au moment de la naissance de référence, par poids à la naissance du dernier né

| Poids à la naissance de l'enfant de référence (grammes) | Naissances précédentes | |
|---|------------------------|-----------------------------------|
| | N | Proportions décédés $\approx 2q0$ |
| 1500-1999 | 76 | 0,197 |
| 2000-2499 | 409 | 0,161 |
| 2500-2999 | 1 389 | 0,153 |
| 3000-3499 | 1 827 | 0,136 |
| 3500-3999 | 607 | 0,104 |
| 4000 ou plus | 98 | 0,092 |

Source : Étude dans les maternités de Bamako (Hill and Aguirre 1990).

réduire la mortalité de la petite enfance. Dans la même étude, le nombre de naissances survenues dans chaque maternité au cours d'une année était suffisamment élevé pour permettre le calcul de l'indice de mortalité de la petite enfance dans

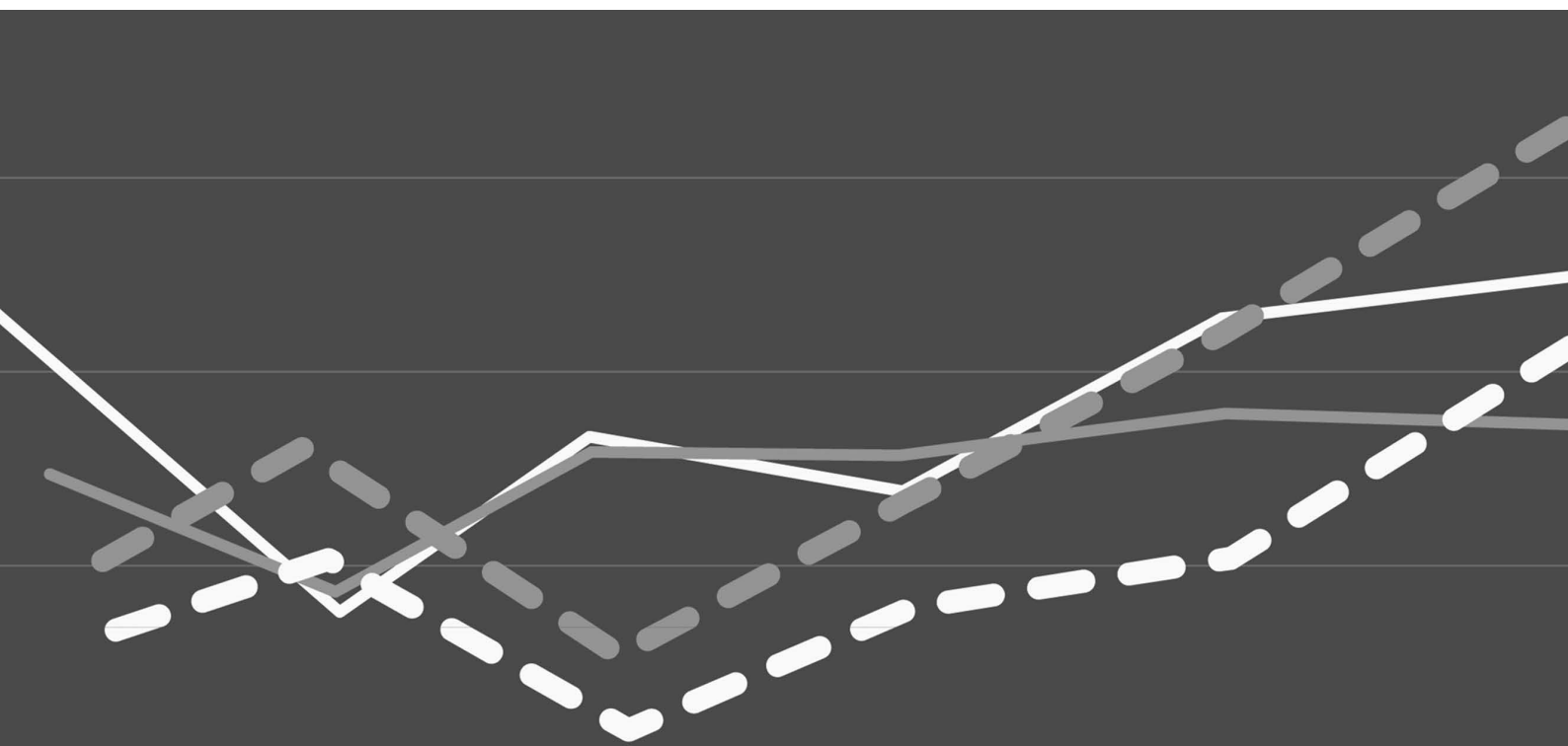
chaque établissement. Il a été ainsi possible de classer les établissements en fonction de la proportion d'enfants précédents décédés au moment de l'accouchement de référence

et d'identifier ainsi les établissements les moins performants, ainsi que les populations dans leur zone de recrutement nécessitant des ressources supplémentaires.

RÉFÉRENCES

- Aguirre A. 1990. "The Preceding Birth Technique for the Estimation of Childhood Mortality: Theory, Extensions and Applications." Unpublished PhD thesis, London: University of London.
- Aguirre A. 1994. "Extension of the preceding birth technique", *Genus* **50**(3-4):151-169.
- Aguirre A and AG Hill. 1988. «Estimacion de la mortalidad de la ninez mediante la tecnica del hijo previo con datos provenientes de centros de salud o de encuestas de hogares: aspectos metodologicos. [Estimating child mortality using the previous child technique, with data from health centers and household surveys: methodological aspects]», *Notas Poblacion* **16**(46-47):9-39.
- Bairagi R, M Shuaib and AG Hill. 1997. "Estimating childhood mortality trends from routine data: a simulation using the preceding birth technique in Bangladesh", *Demography* **34**(3):411-420. doi: <http://dx.doi.org/10.2307/3038293>
- Bicego G, A Augustin, S Musgrave, J Allman and P Kelly. 1989. "Evaluation of a simplified method for estimation of early childhood mortality in small populations", *International Journal of Epidemiology* **18**(4 Suppl 2):S20-32. doi: http://dx.doi.org/10.1093/ije/18.Supplement_2.S20
- Brass W and S Macrae. 1984. "Childhood mortality estimated from reports on previous births given by mothers at the time of a maternity: I. Preceding-births technique", *Asian and Pacific Census Forum* **11**(2):5-8. <http://hdl.handle.net/10125/3561>.
- Brass W and S Macrae. 1985. "Childhood mortality estimated from reports on previous births given by mothers at the time of a maternity: II. Adapted multiplying factor technique", *Asian and Pacific Census Forum* **11**(4):5-9. <http://hdl.handle.net/10125/3562>.
- Hill AG and A Aguirre. 1990. "Childhood mortality estimates using the preceding birth technique - some applications and extensions", *Population Studies* **44**(2):317-340. <http://dx.doi.org/10.1080/0032472031000144616>
- Hill AG and PG Kelly. 1996. *Sur la mise en place de la technique de l'accouchement precedent, Senegal, 19 au 26 janvier 1996 [On the implementation of the Preceding Birth Technique, Senegal 19-26 January 1996]*. Arlington,VA: Partnership for Child Health Care.
- Macro International Inc. MEASURE DHS STAT compiler.<http://www.measuredhs.com>.
- Madi HH. 2000. "Infant and child mortality rates among Palestinian refugee populations", *Lancet* **356**(9226):312. doi: [http://dx.doi.org/10.1016/S0140-6736\(00\)02511-3](http://dx.doi.org/10.1016/S0140-6736(00)02511-3)
- Oliveras E, C Ahiadeke, RM Adanu and AG Hill. 2008. "Clinic-based surveillance of adverse pregnancy outcomes to identify induced abortions in Accra, Ghana", *Studies in Family Planning* **39**(2):133-140. doi: <http://dx.doi.org/10.1111/j.1728-4465.2008.00160.x>
- Rowe AK, F Onikpo, M Lama, DM Osterholt and MS Deming. 2011. "Impact of a malaria-control project in Benin that included the integrated management of childhood illness strategy", *American Journal of Public Health* **101**(12):2333-2341. doi: <http://dx.doi.org/10.2105/AJPH.2010.300068>
- Rutstein S. 2011. *Trends in Birth Spacing*. DHS Comparative Reports No. 28. Calverton, MD: ICF Macro. <http://www.measuredhs.com/pubs/pdf/CR28/CR28.pdf>

La mortalité adulte



Chapitre 19 INTRODUCTION À L'ANALYSE DE LA MORTALITÉ ADULTE

Ian M Timæus, Rob Dorrington et Kenneth Hill

La connaissance précise des niveaux et tendances de la mortalité adulte dans les pays en développement est limitée par une absence généralisée de systèmes d'état civil dont la couverture est complète. Alors que la connaissance de la mortalité infanto-juvénile était, autrefois, tout aussi limitée, les méthodes appliquées aux enquêtes, c'est-à-dire les méthodes indirectes et les histoires génésiques, ont bien mieux réussi à permettre de mesurer la mortalité des enfants que celle des adultes, et l'on sait proportionnellement moins de choses sur ces derniers que sur les premiers.

Pour ce qui concerne l'analyse démographique, la mortalité adulte est en général définie comme la mortalité des personnes âgées de 15 ans et plus. Cependant, dans certains contextes, le terme de « mortalité adulte » se réfère plus particulièrement à la mortalité entre les âges de 15 et 60 ans. On y oppose alors la « mortalité aux âges élevés », c'est-à-dire la mortalité des 60 ans et plus. La probabilité qu'une personne décède entre son 15^e et son 60^e anniversaire (notée $_{45}q_{15}$ dans la table de mortalité), est devenue l'indicateur usuel de la mortalité adulte définie dans ce sens restrictif.

Dans les pays qui n'ont pas de système d'état civil complet, les sources de données et les méthodes utilisées pour étudier la mortalité adulte sont différentes de celles qui sont utilisées pour étudier la mortalité des enfants. Certaines méthodes utilisées pour les adultes peuvent être étendues pour étudier la mortalité des enfants de 5 ans et plus, mais aucune d'elles n'est une source d'information pour la mortalité des moins de 5 ans.

Plusieurs difficultés spécifiques rendent l'étude de la mortalité adulte intrinsèquement plus délicate que celle des enfants. Premièrement, d'une manière générale, l'ordre de grandeur des taux de mortalité adulte est considérablement plus faible que celui des enfants pour la plupart des âges. Les décès d'adultes sont des événements relativement

rare. Obtenir des mesures précises de la mortalité adulte demande donc soit des échantillons plus importants, soit une période d'observation plus longue. Deuxièmement, il est souvent difficile d'identifier un informateur approprié qui soit capable de fournir une information fiable sur les adultes décédés, alors que les informations concernant les décès d'enfants peuvent en général être recueillies auprès de la mère. De plus, les caractéristiques des parents comptent parmi les principaux déterminants du risque de décès dans l'enfance. Comme on ne dispose pas de l'équivalent pour les adultes, c'est-à-dire d'un informateur unique qui convienne pour fournir les différents types de données sur les décès adultes, les problèmes de sous-déclaration et de déclaration multiples sont fréquents. De plus, pour l'étude de la mortalité différentielle, il est souvent peu réaliste d'utiliser les caractéristiques économiques et sociales des répondants comme une approximation de celles des personnes décédées.

Les erreurs de déclaration d'âge constituent un autre problème sérieux qui affecte toutes les sources de données sur la mortalité adulte dans les pays à revenus faibles et intermédiaires. Plusieurs facteurs viennent gêner l'obtention d'une information précise sur les âges des adultes et les âges au décès. Les personnes âgées sont moins susceptibles d'avoir des actes de naissance ou des carnets de santé que les très jeunes, et, dans les pays en développement, ils sont moins susceptibles d'avoir été scolarisés. De plus, même si les personnes décédées connaissent leur propre âge, l'informateur qui déclare leur décès peut ne pas le connaître. L'âge déclaré des adultes les plus âgés est souvent exagéré, et les âges au décès le sont encore plus. Ainsi donc, les estimations « brutes » de la mortalité adulte pour les pays à revenus faibles ou modérés nécessitent souvent un ajustement par une table-type de mortalité avant de pouvoir être utilisés pour estimer l'espérance de vie, ou pour les projections

démographiques. Les estimations faites pour la population très âgée doivent souvent être rejetées et remplacées par des données extrapolées à partir des tables-type de mortalité.

LES DONNÉES POUR L'ESTIMATION DE LA MORTALITÉ ADULTE

Un petit nombre de pays à revenus faibles et intermédiaires, encore plus petit si l'on tient compte de leur population, ont un enregistrement complet des décès adultes et disposent de recensements de bonne qualité. Un grand nombre de pays disposent de systèmes d'état civil, nationaux ou basés sur un sondage, qui soient suffisamment complets pour qu'on y applique les méthodes décrites dans ce manuel d'évaluation de la complétude relative de l'enregistrement de l'état civil par rapport aux effectifs du recensement. De plus, un nombre croissant de pays a introduit dans les recensements (ou dans les grandes enquêtes de ménages par sondage) des questions concernant les décès par sexe et âge au cours d'une période précédente (le plus souvent les 12 derniers mois). La complétude de l'enregistrement de ces décès peut être estimée par les mêmes méthodes que celles qui permettent d'estimer la qualité des données sur les décès adultes enregistrés à l'état civil, en supposant que la majorité des décès se produisent au sein des ménages, et que ces ménages ne sont pas dissous par ces décès.

Un certain nombre de pays, situés surtout en Afrique au sud du Sahara, ont conduit des enquêtes par sondage, le plus souvent sous l'égide du programme des enquêtes démographiques et de santé (EDS), qui incluent des histoires de fratries avec l'information sur la survie de chaque frère et sœur de l'enquêté(e) et éventuellement la date de leurs décès. Certains pays ont essayé de mesurer la mortalité adulte en incluant dans les recensements et enquêtes la question sur la survie des ascendants, père ou mère, des personnes enquêtées. Ces données, comme celles concernant la survie des frères et sœurs, peuvent être tabulées selon l'âge de la personne qui a répondu à la question, et analysées par des méthodes indirectes qui utilisent les modèles démographiques pour les convertir en indicateurs conventionnels de la mortalité adulte que l'on trouve dans les tables de mortalité.

DESCRIPTION DES MÉTHODES

Les méthodes qui utilisent à la fois les décès et la population à risque par sexe et âge, pour estimer la mortalité adulte sont appelées ici « méthodes de répartition des décès ». Ces méthodes peuvent être classées en deux groupes dis-

tincts, selon l'usage qui est fait des données : les méthodes de la balance de « l'accroissement démographique », et la méthode de « l'extinction des cohortes synthétiques » (Hill, You et Choi 2009). Les deux types de méthode nécessitent des données sur les décès provenant soit de l'état civil soit d'une question dans le recensement, et des données sur la population à risque, classées par âge.

La première méthode de la balance des croissances, développée par William Brass, est appelée « méthode de Brass de la balance de l'accroissement démographique » (Brass 1975) (growth balance method). Cette méthode ne nécessite que la population par âge à un point dans le temps, mais elle ne s'applique que si la population adulte est stable, au moins approximativement, c'est-à-dire si la population dispose d'une structure par âge régulière et constante dans le temps.

La seconde méthode de la balance des croissances, développée par Kenneth Hill, est une généralisation de la première méthode pour les populations non-stables (Hill 1987). On l'appelle « méthode généralisée de la balance de l'accroissement démographique » (generalized growth balance method). Cette méthode nécessite des données par âge à deux points dans le temps.

La première méthode de l'extinction des cohortes synthétiques, développée par Samuel Preston et Ansley Coale, nécessite des données sur la population par âge en un point dans le temps et repose sur l'hypothèse selon laquelle la population adulte est stable, au moins approximativement (Preston, Coale *et al.* 1980). La seconde méthode de l'extinction des cohortes synthétiques, développée par Neil Bennett et Shiro Horiuchi, est une généralisation de la première approche pour les populations non-stables (Bennett and Horiuchi 1981 ; 1984). Elle nécessite des données sur la population à risque en deux points dans le temps. En supposant que les hypothèses de complétude constante par âge de la population et des décès sont valides, que les migrations sont négligeables, et que la déclaration de l'âge ne montre pas de distorsion majeure dans les groupes d'âge quinquennaux, alors on préférera les méthodes de répartition des décès pour estimer la mortalité adulte. En effet, d'une part elles produisent des estimations des taux de mortalité par âge, et d'autre part elles sont capables de donner une datation raisonnable (Hill 2001). En pratique, apprécier lesquelles de ces conditions sont vérifiées demande beaucoup d'expérience, ce qui implique que ces méthodes figurent parmi les plus subjectives des méthodes indirectes.

On peut calculer les taux de mortalité par âge directement à partir des décomptes de décès et de personnes-années vécues classées par âge et année dans les histoires de fratries (frères et sœurs) recueillies dans les enquêtes. Mais les tailles d'échantillon des enquêtes, telles que les enquêtes EDS, sont plutôt faibles pour une bonne estimation de la mortalité adulte par année de calendrier. On conseille donc d'agréger les données dans des périodes couvrant plusieurs années. De plus, comme le nombre de frères et sœurs décédés omis dans l'enquête tend à augmenter avec la durée depuis le décès, on devra se limiter aux estimations de mortalité pour les périodes récentes.

Les méthodes de collecte des données telles que les histoires de fratries ne sont pas appropriées au recueil de données sur les causes de décès, car, en général, le répondant ne vit pas dans le même ménage que la personne décédée. Le répondant a donc peu de chances de connaître la cause précise du décès, certifiée médicalement, surtout si la personne décédée n'a reçu que peu ou pas de soins médicaux. Les méthodes alternatives, comme les autopsies verbales, qui sont basées sur le recueil des signes et symptômes précédant le décès, ont peu de chance de marcher correctement, car le répondant n'aura pas d'information de première main concernant ces caractéristiques. Certaines de ces limites s'appliquent aussi aux décès recueillis au sein des ménages lors des recensements : les membres du ménage peuvent ne pas avoir d'information sur la véritable cause du décès. De plus, la formation des enquêteurs et le temps disponible pour l'interview sont en général insuffisants pour permettre une interrogation détaillée. Par contre, l'approche du recensement peut fournir un cadre pour une enquête de suivi par autopsie verbale, sur un échantillon de ménages qui ont déclaré des décès. Ceci nécessite d'avoir recours à des enquêteurs formés à cet effet, et ces enquêtes sont complexes et coûteuses.

Deux autres exceptions à cette règle générale sont constituées par le fait qu'il est possible de distinguer les accidents et morts violentes des causes naturelles, et d'identifier les « décès liés à la grossesse », c'est-à-dire les décès qui se produisent lorsqu'une femme est enceinte, durant l'accouchement ou pendant les six semaines qui suivent la fin de la grossesse. La méthode d'estimation de la mortalité liée à la grossesse à partir des questions posées lors d'un recensement ou lors d'une collecte des histoires de fratries est décrite dans la section sur la mortalité maternelle.

L'alternative à la collecte de données précises sur les décès et la population à risque par âge pour estimer la mortalité

adulte est d'utiliser les méthodes d'estimation indirecte. Ces méthodes ne requièrent aucune information sur l'âge ni sur la date du décès. À la place, il suffit de tabuler selon l'âge du répondant la proportion de personnes encore en vie parmi certaines catégories spécifiques de personnes apparentées aux personnes qui répondent au questionnaire. Les mesures conventionnelles des probabilités de survie de la table de mortalité sont ensuite calculées à partir de ces proportions à l'aide d'un modèle de régression qui a été élaboré à partir de données pour lesquelles on disposait de la relation entre ces deux types d'information.

La technique la plus utilisée pour analyser la mortalité des hommes et des femmes adultes à l'aide de la survie des apparentés est celle qui étudie la survie des ascendants, pères et mères. Cette méthode, appelée la « méthode des orphelins », a été développée par William Brass et Kenneth Hill (Brass et Hill 1973). Les applications récentes qui visent à calculer les survivants de la table de mortalité utilisent plutôt les coefficients de régression proposés par Ian Timæus (Timæus 1992), et non pas les pondérations proposées initialement, car ces coefficients fournissent des estimations plus précises pour les hommes. Des variantes de cette méthode sont aussi présentées dans ce volume, dont le but est de s'appliquer à des populations ayant une forte prévalence de l'infection à VIH. Ces variantes s'appliquent aussi au cas où l'on demande aux répondants si leurs parents étaient toujours en vie lorsqu'ils ont atteints à l'âge adulte, un âge identifié, par exemple, à partir de l'âge au mariage.

Lorsque plusieurs ensembles de données ont été recueillis sur la survie des ascendants, pères et mères, dans des enquêtes successives conduites dans la même population, on peut construire une cohorte synthétique d'orphelins pour estimer la mortalité adulte au cours de la période intermédiaire (Zlotnik et Hill 1981). Ces estimations peuvent se faire à partir des données sur les orphelins de tous âges en utilisant les coefficients de régression relatifs à la méthode de base. Ce manuel propose cependant une autre méthode d'analyse des cohortes synthétiques d'orphelins développée par Ian Timæus (Timæus 1991). Cette variante est moins sensible au sous-enregistrement des orphelins par les répondants dont les parents biologiques sont décédés lorsque le répondant était encore un jeune enfant.

Enfin, ce manuel décrit les méthodes développées par Ian Timæus et collègues, qui permettent d'estimer indirectement la mortalité adulte à partir des données sur les frères et sœurs, pour lesquelles on demande au répondant combien

de frères et sœurs sont décédés avant l'âge adulte, et combien sont décédés après (Timæus *et al.* 2001).

D'autres méthodes ont été proposées pour l'estimation indirecte de la mortalité adulte à partir des questions sur la survie d'autres apparentés, en particulier la survie du premier conjoint du répondant, premier époux ou première épouse. L'expérimentation de ces questions a montré que les répondants omettent souvent de déclarer le fait qu'ils sont devenus veuf ou veuve. En conséquence, cette méthode produit souvent des sous-estimations sévères de la mortalité adulte. La méthode du veuvage et les autres méthodes basées sur la survie d'autres apparentés qui se sont révélées défectueuses ne sont pas décrites dans ce manuel.

Une dernière approche qui a été utilisée pour estimer la mortalité adulte repose sur les variations en effectifs d'une population entre deux recensements. Dans une population fermée aux migrations, bénéficiant de données précises, toute personne

enquêtée au premier recensement mais non recensée au second doit être décédée entre temps. En pratique, dans la plupart des pays, le nombre de migrations internationales annuelles nettes représente une fraction non-négligeable du nombre de décès adultes. Peu de pays ont une mesure suffisamment précise des flux migratoires internationaux qui permettrait d'ajuster les variations intercensitaires de la population avant d'estimer la mortalité. De plus, même une faible variation de la complétude du dénombrement de la population peut produire des biais sérieux dans l'estimation de la mortalité adulte calculée de cette manière. C'est pourquoi cette approche n'est en général pas recommandée comme méthode pour estimer la mortalité adulte, et ce manuel ne la traite pas en détail. On conseille aux lecteurs qui souhaitent néanmoins en savoir plus sur cette approche pour estimer la mortalité de consulter d'abord la variante de cette méthode proposée par Samuel Preston et Neil Bennett (Preston et Bennett 1983).

AUTRES LECTURES ET RÉFÉRENCES

Il n'existe pas d'article récent qui décrive et analyse complètement toutes les méthodes disponibles pour estimer la mortalité adulte dans les pays aux données limitées et déficientes. On trouvera cependant des discussions approfondies de l'estimation de la mortalité des adultes, ainsi que de la mortalité des enfants, dans l'article de Hill *et al.* (2005), et le chapitre de Reniers *et al.* (2011) donne une discussion brève, mais à jour, des méthodes d'estimation, et présente les résultats en mettant les méthodes en pratique.

Bennett NG and S Horiuchi. 1981. "Estimating the completeness of death registration in a closed population", *Population Index* 47(2):207-221.

Bennett NG and S Horiuchi. 1984. "Mortality estimation from registered deaths in less developed countries", *Demography* 21(2):217-233. doi: <http://dx.doi.org/10.2307/2061041>

Brass W. 1975. *Methods for Estimating Fertility and Mortality from Limited and Defective Data*. Chapel Hill: International Program of Laboratories for Population Statistics.

Brass W and KH Hill. 1973. "Estimating adult mortality from orphanhood," in *International Population Conference, Liège, 1973*. Vol. 3 Liège: International Union for the Scientific Study of Population, pp. 111-123.

Hill K. 1987. "Estimating census and death registration completeness", *Asian and Pacific Population Forum* 1(3):8-13, 23-24. <http://hdl.handle.net/10125/3602>

Hill K. 2001. "Methods for measuring adult mortality in developing countries: A comparative review", Paper presented at XXIV IUSSP General Conference, Salvador, Brazil.

Hill K, Y Choi and IM Timæus. 2005. "Unconventional approaches to mortality estimation", *Demographic Research* 13:281-300. doi: <http://dx.doi.org/10.4054/DemRes.2005.13.12>

Hill K, D You and Y Choi. 2009. "Death distribution methods for estimating adult mortality: Sensitivity analysis with simulated data error", *Demographic Research* 21:235-254. doi: <http://dx.doi.org/10.4054/DemRes.2009.21.9>

Preston SH and NG Bennett. 1983. "A census-based method for estimating adult mortality", *Population Studies* 37(1):91-104. doi: <http://dx.doi.org/10.1080/00324728.1983.10405926>

Preston SH, AJ Coale, J Trussell and M Weinstein. 1980. "Estimating the completeness of reporting of adult deaths in populations that are approximately stable", *Population Index* 46(2):179-202.

Reniers G, B Masquelier and P Gerland. 2011. "Adult mortality trends in Africa," in Rogers, RG and EM Crimmins (eds). *International Handbook of Adult Mortality*. Springer, pp. 151-170. doi: <http://dx.doi.org/10.1007/978-90-481-9996-9>

Timæus IM. 1991. "Estimation of mortality from orphanhood in adulthood", *Demography* 28(2):213-227. doi: <http://dx.doi.org/10.2307/2061276>

Timæus IM. 1992. "Estimation of adult mortality from paternal orphanhood: a reassessment and a new approach", *Population Bulletin of The United Nations* 33:47-63.

Timæus IM, B Zaba and M Ali. 2001. "Estimation of adult mortality from data on adult siblings," in Zaba, B and J Blacker (eds). *Brass Tacks: Essays in Medical Demography*. London: Athlone, pp. 43-66.

Zlotnik H and K Hill. 1981. "The use of hypothetical cohorts in estimating demographic parameters under conditions of changing fertility and mortality", *Demography* 18(1):103-122. doi: <http://dx.doi.org/10.2307/2061052>

Méthodes fondées
sur un seul recensement

Chapitre 20 LA MÉTHODE DE LA BALANCE DE L'ACCROISSEMENT DÉMOGRAPHIQUE DE BRASS

Rob Dorrington

DESCRIPTION DE LA MÉTHODE

La méthode de la balance de l'accroissement démographique de Brass (*growth balance method*) est la première de la série des méthodes dites de la répartition des décès pour estimer la complétude de la déclaration des décès relativement à une estimation de la population (Brass 1975). La méthode part de l'observation que dans une population stable (c'est-à-dire une population ayant une structure par âge constante dans le temps, au moins pour les âges adultes, croissant à un taux annuel constant r , fermée aux migrations et avec une bonne déclaration de l'âge), le taux de croissance r est égal au taux de natalité b moins le taux de mortalité d . La même relation est vérifiée dans le groupe d'âge de la population âgée d'au moins x années, c'est-à-dire : $r = b(x+) - d(x+)$ où $b(x+)$ représente le taux auquel la population atteint le groupe d'âge $x+$ (les personnes âgées d'au moins x années), et $d(x+)$ est le taux de mortalité dans le même groupe d'âge $x+$. Si, dans cette population, les décès sont sous-représentés dans la même proportion à chaque âge, alors $b(x+) = r + d^r(x+) / c$, où $d^r(x+)$ est le taux de mortalité basé sur les décès déclarés dans le groupe d'âge $x+$, et c est la proportion de décès qui sont déclarés. On peut estimer c à partir de la pente d'une ligne droite ajustée aux points de données $[b(x+), d^r(x+)]$. Cette estimation est en général restreinte aux âges adultes, car la complétude de l'enregistrement des décès d'enfants diffère souvent de celle des décès d'adultes. Les taux de mortalité sont ensuite calculés en divisant le nombre de décès déclarés par c , et en divisant ce nombre par une estimation de la population exposée au risque basée sur la population utilisée pour calculer les taux partiels de natalité et de mortalité avant correction.

Cette méthode est un cas particulier de la méthode plus générale connue sous le nom de « méthode généralisée de

la balance de l'accroissement démographique » (*generalized growth balance method*), qui requiert des estimations de la population à deux instants donnés, mais s'affranchit de l'hypothèse de stabilité de la population. On renvoie le lecteur à la section correspondante pour plus de détails sur la méthode. La version présentée ici figure dans ce manuel comme une méthode que l'on peut considérer lorsqu'on ne dispose que d'un point dans le temps pour les effectifs de population.

DONNÉES NÉCESSAIRES ET HYPOTHÈSES

Tabulation des données nécessaires

- Nombre de décès de femmes (ou hommes) au cours d'une période donnée, répartis par groupe d'âge de 5 ans, et pour le dernier groupe d'âge $A+$, A étant aussi élevé que possible.
- Nombre de femmes (ou hommes) à un point dans le temps aussi proche que possible de la période considérée, aussi répartis par groupe d'âge de 5 ans, et pour le dernier groupe d'âge $A+$, A étant aussi élevé que possible.

Hypothèses importantes

- La population est stable, quoique cette hypothèse puisse être relâchée dans certaines limites (voir ci-dessous).
- La complétude de l'enregistrement des décès est la même pour tous les âges au-dessus d'un âge limite inférieure (en général 15 ans).
- La population est fermée aux migrations, quoique cette hypothèse puisse également être relâchée si les migrations nettes sont faibles par rapport aux taux de mortalité, ou si on dispose d'estimations raisonnables de migrants par âge qui peuvent être prises en compte dans l'équation de la balance de l'accroissement démographique (ce qui est un cas très rare en pratique).

TRAVAUX PRÉPARATOIRES ET RECHERCHES PRÉLIMINAIRES

Avant de mettre en œuvre cette méthode, on doit au préalable analyser la qualité des données, et au moins les points suivants :

- structure par âge de la population ;
- répartition par sexe de la population ;
- structure par âge des décès ; et
- répartition par sexe des décès.

PRÉCAUTIONS ET MISES EN GARDE

Lorsqu'ils appliquent cette méthode, les analystes doivent prendre en considération les points suivants :

- Le processus d'interprétation et d'estimation doit prendre en compte la source des données de mortalité (décès déclarés à l'état civil, décès déclarés par les ménages dans les recensements, ou décès déclarés dans les structures sanitaires), comme cela est expliqué plus loin. Cependant, les biais associés à la source des données tendent à avoir moins d'effet sur l'estimation de la complétude avec la méthode de la balance de l'accroissement démographique (chapitre 24) qu'avec la méthode de l'extinction des cohortes synthétiques (chapitre 25).
- Le problème des migrations devient plus sérieux quand on applique la méthode à des zones géographiques régionales, et non plus nationales.
- Il est nécessaire de choisir l'intervalle d'âges sur lequel on ajuste une droite reliant les taux partiels de natalité et de mortalité, et qui permet de calculer la complétude. Les problèmes spécifiques sont : quel est le meilleur âge pour choisir l'intervalle ouvert lorsqu'on suspecte une tendance à l'exagération de l'âge ; comment traiter les points qui se situent au-dessus de la ligne aux âges élevés lorsque la complétude tend à baisser avec l'âge, par exemple dans le cas de migrations associées à la retraite entre les milieux urbain et rural où l'enregistrement est moins complet ; et faut-il exclure les âges en dessous de 30 ou 35 ans, car l'impact des migrations n'y est pas pris en compte spécifiquement.
- Lorsque la complétude des décès déclarés semble être inférieure à 60 %, on conseille de prendre garde aux résultats, car l'incertitude sur les estimations devient importante.

APPLICATION DE LA MÉTHODE

Étape 1 : Cumuler la population et les décès à partir du bas

Afin de calculer les taux partiels de natalité et de mortalité, il faut d'abord cumuler les effectifs de population et de décès au cours d'une période de t années, pour tous les âges au-delà de l'âge x . Pour la population, on utilise la formule suivante :

$$N(x+) = \sum_{y=x}^{A-5} {}_5N_y + {}_{\infty}N_A$$

où A est l'âge au début du dernier groupe d'âge ouvert.

On utilise une équation analogue pour calculer le nombre de décès au-delà de l'âge x , notée : $D(x+)$.

Étape 2 : Calcul des personnes-années vécues : $PYL(x+)$

Pour calculer les taux partiels de natalité et de mortalité, on doit calculer les personnes-années d'exposition au risque de décès. Pour ce faire, on utilise la formule suivante :

$$PYL(x+) = t \cdot N(x+)$$

où t est la longueur de la période au cours de laquelle les décès ont été enregistrés.

Étape 3 : Calcul du nombre de personnes qui atteignent l'âge x dans la population, $N(x)$

Le nombre de personnes qui atteignent l'âge x (qui correspond à une arrivée, ou « naissance », dans le groupe d'âge $x+$) dans la population se calcule comme la moyenne géométrique des effectifs dans les deux groupes quinquennaux adjacents, divisée par 5, et multipliée par t , la longueur de l'intervalle au cours duquel les décès sont déclarés, exprimé en années, selon la formule :

$$N(x) = \frac{t}{5} \left({}_5N_{x-5} \times {}_5N_x \right)^{\frac{1}{2}}$$

Étape 4 : Calcul des taux partiels de natalité et de mortalité, $b(x+)$ et $d(x+)$

Les taux partiels de natalité et mortalité sont calculés selon les formules suivantes :

$$b(x+) = \frac{N(x)}{PYL(x+)}$$

et

$$d(x+) = \frac{D(x+)}{PYL(x+)}$$

respectivement.

Étape 5 : Représentation graphique, ajustement d'une droite et estimation de la complétude (c)

Pour estimer la complétude de l'enregistrement des décès relativement à la population, on commence par tracer sur un graphique la relation entre $b(x+)$ et $d(x+)$. Les coefficients (pente et ordonnée à l'origine) de la droite de régression orthogonale qui les lient sont calculés selon les formules suivantes :

$$b = \frac{\sigma_y}{\sigma_x} = \frac{\frac{1}{n-1} \sum_{i=1}^n (y_i - \mu_y)^2}{\frac{1}{n-1} \sum_{i=1}^n (x_i - \mu_x)^2}$$

et

$$a = \mu_y - b\mu_x$$

où b est la pente et a l'ordonnée à l'origine de la droite liant les ordonnées y_i , c'est à dire les $b(x+)$ et les abscisses x_i , c'est-à-dire les $d(x+)$; et représentent respectivement les moyennes des deux séries.

Il convient alors d'analyser l'alignement des points et les résidus, de manière à choisir le meilleur intervalle d'âge à garder pour calculer la complétude de l'enregistrement des décès. Ce choix est discuté en détail ci-dessous. En pratique, tout point ayant un résidu supérieur à 1 % en valeur absolue devrait être exclu. Une ligne finale est alors tracée à partir des points restants, et on calcule une nouvelle valeur des coefficients a et b pour cette nouvelle ligne.

La complétude de l'enregistrement des décès se calcule alors à partir des valeurs de a et de b selon la formule suivante :

$$c = \frac{1}{b} \exp(a(t_c - t_m))$$

où t_c désigne la date du recensement, et t_m désigne le milieu de la période au cours de laquelle les décès ont été enregistrés. La justification de cette équation est la suivante : l'inverse de la pente donne une estimation de la complétude de l'enregistrement, en faisant l'hypothèse que le recensement est placé au milieu de la période d'enregistrement des

décès. Le second terme est utilisé pour corriger la différence entre la population à la date t et la population en milieu de période t_m . En faisant l'hypothèse d'une population stable, croissant à un taux estimé par a , ce ratio se calcule comme suit : $\exp(a(t_c - t_m))$.

Étape 6 : Estimation des taux de mortalité après correction pour tenir compte de la sous-déclaration des décès

Pour calculer les taux de mortalité au cours de la période d'enregistrement, il faut d'abord estimer la population moyenne, qui, là encore, doit tenir compte de la croissance de la population. On l'obtient en multipliant les effectifs du recensement répartis par groupes d'âge quinquennaux par le coefficient suivant : $\exp(-a(t_c - t_m))$.

Ensuite, on doit corriger le nombre de décès observés, en les divisant par l'estimation de la complétude c .

Les personnes-années d'exposition sont calculées en multipliant la population moyenne (au temps t_m) par la longueur de la période d'observation t .

Les taux de mortalité, après correction pour tenir compte du fait que l'enregistrement des décès est incomplet, sont alors calculés selon la formule :

$${}_5m_x = \frac{{}_5D_x / c}{t \times {}_5N_x \exp(-a(t_c - t_m))}$$

Le numérateur et le dénominateur sont tous deux ajustés par le coefficient de croissance de la population :

$$\exp(-a(t_c - t_m)).$$

On pourrait donc ignorer cette correction et obtenir en fin de compte les mêmes chiffres pour les taux de mortalité corrigés. Mais l'estimation de la complétude serait alors équivalente à celle qui s'appliquerait si la population au temps t_m était égale à celle observée au temps t_c .

Étape 7 : Lissage à l'aide d'un système relationnel logit de tables-types de mortalité

Puisque les taux de mortalité par âge peuvent être erratiques, ils doivent être lissés. Ceci peut être réalisé à l'aide d'une fonction relationnelle logit (ajustement de Brass) à partir d'une série de taux de mortalité d'une table type que l'on considère comme ayant le même schéma par âge que la population analysée.

Le classeur Excel associé à ce manuel contient une feuille de calcul qui permet de produire un ensemble de taux de

mortalité lissés par un système relationnel logit appliqué à une table-type. L'utilisateur peut choisir une table-type parmi le modèle Général des tables-type des Nations Unies, ou un des quatre modèles régionaux des tables-type de Princeton. On peut aussi choisir comme référence une table de mortalité spéciale, si l'on a une bonne raison de penser que ce schéma par âge correspond mieux au schéma de la mortalité adulte de la population étudiée.

Pour effectuer le lissage, les quotients quinquennaux de mortalité, notés ${}_5q_x$, sont calculés à partir des taux de mortalité corrigés pour la complétude comme suit :

$${}_5q_x = \frac{{}_5m_x}{1 + 2,5{}_5m_x}.$$

Puis, on part d'une racine de 1 à cinq ans ($l_5 = 1$), et on calcule de proche en proche les survivants de la table selon la formule : $l_{x+5} = l_x(1 - {}_5q_x)$.

Les coefficients α et β s'obtiennent en ajustant le modèle relationnel suivant :

$$\gamma_x = \alpha + \beta\gamma_x^s$$

où la fonction logit est définie comme :

$$\gamma_x = 0,5 \ln \left(\frac{1 - l_x}{l_x} \right)$$

et où l'exposant s désigne les valeurs de la table-type (le standard). Enfin, on peut établir la table de mortalité lissée en appliquant les coefficients α et β comme suit :

$$\gamma_x^{fitted} = \alpha + \beta\gamma_x^s$$

et

$$l_x^{fitted} = \frac{1}{\exp(1 + 2\gamma_x^{fitted})}$$

Les taux de mortalité lissés se calculent à partir de cette table de mortalité selon les formules :

$${}_5m_x^{fitted} = \frac{l_x^{fitted} - l_{x+5}^{fitted}}{T_x - T_{x+5}}$$

et

$${}_om_x^{fitted} = \frac{l_x^{fitted}}{T_x}$$

où

$$T_x = \sum_{x=x,5}^{\omega} \frac{5}{2} (l_x^{fitted} + l_{x+5}^{fitted})$$

soit

$$T_x = T_{x+5} + \frac{5}{2} (l_x^{fitted} + l_{x+5}^{fitted})$$

et ω représente l'âge ultime de la table de mortalité, au-delà duquel il ne reste plus aucun survivant.

EXEMPLE DÉTAILLÉ

Cet exemple utilise les données du Salvador : les données de la population féminine viennent du recensement de 1961, et les décès féminins de l'état civil pour l'année 1961. Cet exemple apparaît dans la feuille de calcul intitulée : *BGB_El Salvador*. La date de référence du recensement est la nuit du 5 au 6 mai 1961, c'est cette date qui est entrée comme 1961/05/06 dans la feuille d'introduction.

Étape 1 : Cumul de la population, des décès et des migrants à partir du bas

On cumule les effectifs de population et de décès à partir du bas, c'est-à-dire à partir du dernier groupe d'âge (tableau 20.1).

TABLEAU 20.1 Calcul des effectifs cumulés de population et de décès, El Salvador, Recensement de 1961

| Âge | ${}_5N_x$ | ${}_5D_x$ | $N(x+)$ | $D(x+)$ |
|-------|-----------|-----------|-----------|---------|
| 0-4 | 214 089 | 6 909 | 1 274 253 | 13 652 |
| 5-9 | 190 234 | 610 | 1 060 164 | 6 743 |
| 10-14 | 149 538 | 214 | 869 930 | 6 133 |
| 15-19 | 125 040 | 266 | 720 392 | 5 919 |
| 20-24 | 113 490 | 291 | 595 352 | 5 653 |
| 25-29 | 91 663 | 271 | 481 862 | 5 362 |
| 30-34 | 77 711 | 315 | 390 199 | 5 091 |
| 35-39 | 72 936 | 349 | 312 488 | 4 776 |
| 40-44 | 56 942 | 338 | 239 552 | 4 427 |
| 45-49 | 46 205 | 357 | 182 610 | 4 089 |
| 50-54 | 38 616 | 385 | 136 405 | 3 732 |
| 55-59 | 26 154 | 387 | 97 789 | 3 347 |
| 60-64 | 29 273 | 647 | 71 635 | 2 960 |
| 65-69 | 14 964 | 449 | 42 362 | 2 313 |
| 70-74 | 11 205 | 504 | 27 398 | 1 864 |
| 75+ | 16 193 | 1 360 | 16 193 | 1 360 |

Étape 2 : Calcul des personnes-années vécues, $PYL(x+)$

Comme des décès sont enregistrés selon l'année d'âge, les personnes-années vécues dans chaque groupe d'âge (colonne 2 du tableau 20.2) sont simplement la somme des effectifs à chaque âge dans le recensement (soit la même chose que la colonne 4 du tableau 20.1).

Étape 3 : Calcul des effectifs de personnes atteignant l'âge x dans la population, $N(x)$

Les effectifs de personnes qui atteignent l'âge x figurent dans la troisième colonne du tableau 20.2. Par exemple, le nombre de personnes atteignant l'âge de 70 ans est calculé comme suit :

$$N(70) = \frac{1}{5}(14964 \times 11205)^{\frac{1}{2}} = 2589,8.$$

Étape 4 : Calcul des taux partiels de natalité et de mortalité, $b(x+)$ and $d(x+)$

Les taux partiels de natalité et de mortalité figurent dans la quatrième et la cinquième colonne du tableau 20.2. Par exemple, pour l'âge de 20 ans, on a :

$$b(20+) = \frac{23825}{595352} = 0,0400$$

et

$$d(20+) = \frac{5653}{595352} = 0,0095.$$

Étape 5 : Représentation graphique, ajustement de la droite de régression, et estimation de la complétude, c

Pour tracer le graphique et ajuster la droite de régression à partir de tous les points représentant les données, on choisit l'âge minimal à 5 ans et l'âge maximal à $A-1$, où A représente l'âge au début du dernier groupe d'âge, c'est-à-dire l'intervalle ouvert (soit 75 ans dans cet exemple). La figure 20.1 présente les points d'ordonnée $b(x+)$ et d'abscisse $d(x+)$. Les coefficients de la droite de régression liant ses points sont calculés selon les formules suivantes :

$$b = \frac{\sigma_y}{\sigma_x} = \frac{0,020547}{0,019103} = 1,0756$$

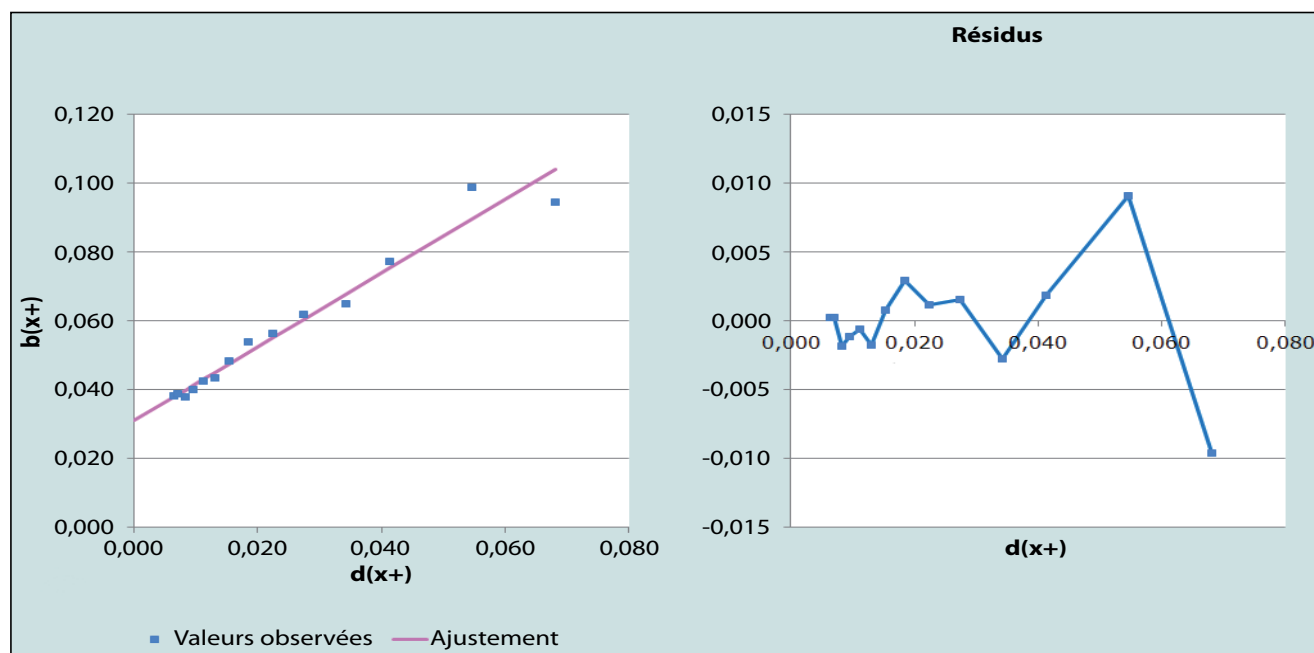
$$a = 0,05686 - 1,0756 \times 0,02407 = 0,031.$$

L'analyse des points figurant sur la figure 20.1 suggère que la plupart des points sont approximativement alignés, à l'exception des deux derniers, ceux qui sont situés le plus à droite. On n'observe pas de preuve évidente d'un effet des migrations. Bien que les résidus des deux der-

TABLEAU 20.2 Calcul des personnes-années vécues, du nombre de personnes atteignant l'âge x , des taux partiels de natalité et de mortalité, et des résidus, El Salvador, Recensement de 1961

| Âge | $PYL(x+)$ | $N(x)$ | $b(x+)$ | $d(x+) = X$ | $b(x+) = Y$ | $a + bx$ | Résidus $y - (a + bx)$ |
|-------|-----------|--------|---------|-------------|-------------|----------|---------------------------|
| 0-4 | 1 274 253 | | | 0,00000 | | 0,03097 | |
| 5-9 | 1 060 164 | 40 362 | 0,03807 | 0,00636 | 0,03807 | 0,03782 | 0,00026 |
| 10-14 | 869 930 | 33 733 | 0,03878 | 0,00705 | 0,03878 | 0,03856 | 0,00022 |
| 15-19 | 720 392 | 27 348 | 0,03796 | 0,00822 | 0,03796 | 0,03981 | - 0,00185 |
| 20-24 | 595 352 | 23 825 | 0,04002 | 0,00950 | 0,04002 | 0,04119 | - 0,00117 |
| 25-29 | 481 862 | 20 399 | 0,04233 | 0,01113 | 0,04233 | 0,04294 | - 0,00061 |
| 30-34 | 390 199 | 16 880 | 0,04326 | 0,01305 | 0,04326 | 0,04501 | - 0,00175 |
| 35-39 | 312 488 | 15 057 | 0,04818 | 0,01528 | 0,04818 | 0,04741 | 0,00077 |
| 40-44 | 239 552 | 12 889 | 0,05380 | 0,01848 | 0,05380 | 0,05085 | 0,00295 |
| 45-49 | 182 610 | 10 259 | 0,05618 | 0,02239 | 0,05618 | 0,05506 | 0,00112 |
| 50-54 | 136 405 | 8 448 | 0,06193 | 0,02736 | 0,06193 | 0,06040 | 0,00153 |
| 55-59 | 97 789 | 6 356 | 0,06500 | 0,03423 | 0,06500 | 0,06779 | - 0,00279 |
| 60-64 | 71 635 | 5 534 | 0,07725 | 0,04132 | 0,07725 | 0,07542 | 0,00183 |
| 65-69 | 42 362 | 4 186 | 0,09881 | 0,05460 | 0,09881 | 0,08970 | 0,00911 |
| 70-74 | 27 398 | 2 590 | 0,09452 | 0,06803 | 0,09452 | 0,10415 | - 0,00963 |
| 75+ | 16 193 | | | | | | |

FIGURE 20.1 Diagnostic graphique, avec un intervalle d'âge ouvert de 75 ans et plus, El Salvador, Recensement de 1961



niers points tombent juste dans la limite tolérée de moins de 1 %, et que nous pouvons donc estimer la complétude de l'enregistrement des décès à 93 %, il est utile d'examiner le cas où ces deux points seraient éliminés. Le calcul de la complétude se fait à partir des coefficients a et b , comme suit :

$$c = \frac{1}{1,07559} \exp(-0,031(1961,34 - 1961,5))$$

$$= \frac{1}{1,07559} e^{0,031 \times 0,16} = 92,51\%.$$

Omettre les deux derniers points, ce qui revient à choisir 64 ans comme la limite supérieure de l'âge, produit un meilleur ajustement, qui apparaît sur la figure 20.2, et qui donne une complétude de 89 %, ce qui est assez proche de l'estimation précédente, et suggère qu'il n'est pas nécessaire d'omettre ces deux derniers points. Par contre, omettre le dernier point seulement induit un changement notable dans

l'estimation, soit une complétude de 82 % et un ajustement linéaire moins précis pour certains points situés à gauche, ce qui suggère que cette stratégie n'est pas la meilleure. Pour les populations qui souffrent d'une forte attirance pour les chiffres ronds, une règle générale est de ne pas terminer à un âge qui se termine par zéro.

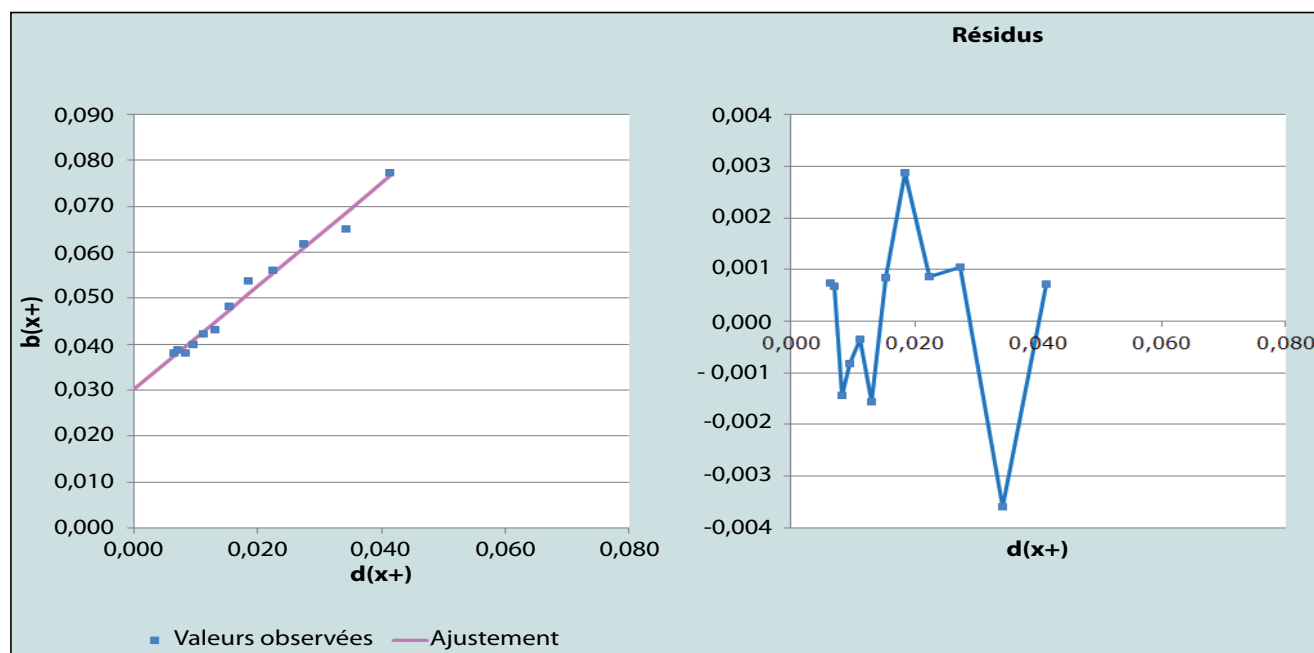
Étape 6 : Calcul des taux de mortalité après correction pour tenir compte de la sous-déclaration des décès

La population au milieu de la période au cours de laquelle les décès sont déclarés est estimée en ajustant la population recensée en tenant compte de la croissance démographique entre les deux dates, soit un taux de 3,1 %. Cette estimation apparaît dans la seconde colonne du tableau 20.3. Par exemple, pour le groupe d'âge 15-19 ans, l'effectif est calculé de la manière suivante :

$${}_5N_{15}(t_m) = 125040 \times \exp(-0,031(1961,34 - 1961,50))$$

$$= 125040 e^{0,031 \times 0,16} = 125662,3.$$

FIGURE 20.2 Diagnostic graphique, avec un intervalle d'âge ouvert de 65 ans et plus, El Salvador, Recensement de 1961



Ensuite, les décès sont corrigés pour tenir compte de leur sous-déclaration, en divisant le nombre de décès déclarés dans chaque groupe d'âge par l'estimation de la complétude. Ces calculs apparaissent dans la colonne 3 du tableau 20.3. Ainsi, pour le groupe d'âge 15-19 ans, le nombre observé (266 décès figurant en colonne 3 du tableau 20.1) est corrigé en 287,5 décès, comme suit :

$$\frac{266}{0,9251} = 287,5.$$

Les personnes-années vécues apparaissent en colonne 4 du tableau 20.3. Ce sont les effectifs de la population au milieu de la période où les décès sont enregistrés (colonne 2 du tableau 20.3), multipliés par la durée de la période (en années), qui dans ce cas est simplement 1 an.

Les taux de mortalité, après correction pour tenir compte de la sous-déclaration des décès, figurent en colonne 5 du tableau 20.3. Ils sont calculés en divisant le nombre de décès corrigés par les personnes-années vécues dans l'intervalle. Par exemple, pour le groupe d'âge 15-19 ans, le taux de mortalité corrigé se calcule comme suit :

TABLEAU 20.3 Calcul des taux de mortalité corrigés, El Salvador, Recensement de 1961

| Âge | Population ajustée ${}_5N_x(t_m)$ | Décès ajustés ${}_5D_x$ | Personnes-années ajustées $PYL(x,5)$ | Taux ajustés ${}_5m_x$ |
|-------|--------------------------------------|----------------------------|---|---------------------------|
| 0-4 | | | | |
| 5-9 | 191 181 | 659 | 191 181 | 0,0034 |
| 10-14 | 150 282 | 231 | 150 282 | 0,0015 |
| 15-19 | 125 662 | 288 | 125 662 | 0,0023 |
| 20-24 | 114 055 | 315 | 114 055 | 0,0028 |
| 25-29 | 92 119 | 293 | 92 119 | 0,0032 |
| 30-34 | 78 098 | 340 | 78 098 | 0,0044 |
| 35-39 | 73 299 | 377 | 73 299 | 0,0051 |
| 40-44 | 57 225 | 365 | 57 225 | 0,0064 |
| 45-49 | 46 435 | 386 | 46 435 | 0,0083 |
| 50-54 | 38 808 | 416 | 38 808 | 0,0107 |
| 55-59 | 26 284 | 418 | 26 284 | 0,0159 |
| 60-64 | 29 419 | 699 | 29 419 | 0,0238 |
| 65-69 | 15 038 | 485 | 15 038 | 0,0323 |
| 70-74 | 11 261 | 545 | 11 261 | 0,0484 |
| 75+ | 16 274 | 1 470 | 16 274 | 0,0903 |

$$\frac{287,60}{125694} = 0,0023.$$

Étape 7 : Lissage utilisant un système relationnel logit de tables-types de mortalité

Les estimations des quotients quinquennaux de mortalité féminine, notés ${}_5q_x$, calculés ci-dessus à partir des taux de mortalité corrigés, apparaissent dans la deuxième colonne du tableau 20.4. Par exemple, la probabilité pour une femme de 15 ans de décéder avant d'atteindre le 20^e anniversaire, se calcule comme suit :

$${}_5q_{15} = \frac{5 \times 0,00229}{1 + 2,5 \times 0,00229} = 0,0114.$$

Les survivants de la table de mortalité se calculent de proche en proche dans les groupes d'âge quinquennaux. Ils appa-

raissent dans la colonne 3 du tableau 20.4. Par exemple, la proportion de personnes survivantes à l'âge de 20 ans se calcule comme suit :

$$l_{20} = 0,9754(1 - 0,0114) = 0,9643.$$

La transformation logit des proportions de survivants apparaît dans la colonne 4 du tableau 20.4. Par exemple, le logit de l_{20} se calcule comme suit :

$$\gamma_{20} = 0,5 \ln \left(\frac{1 - 0,9643}{0,9643} \right) = -1,6477.$$

La transformation logit de la table de mortalité féminine tirée du modèle Ouest des tables-types de Princeton avec une espérance de vie $e^o = 60$ ans (colonne 5 du tableau 20.4) apparaît dans la colonne 6 du même tableau. Comme on peut le voir sur la figure 20.3, le modèle Ouest semble bien ajuster les données, sauf peut-être pour les âges jeunes.

TABLEAU 20.4 Calcul des taux de mortalité lissés par le biais d'un système relationnel logit de tables-types de mortalité, El Salvador, Recensement de 1961

| Âge | ${}_5q_x$ | l_x/l_5 | $Y(x)$ obs. | Table-type Princeton West $l^s(x)$ | Logit $Y_s(x)$ | Ajusté $Y(x)$ | $l(x)$ | $T(x)$ | Valeurs lissées ${}_5m_x$ |
|-----|-----------|-----------|-------------|--|----------------|---------------|--------|--------|---------------------------------|
| 0 | | | | | | | | | |
| 5 | 0,0171 | 1 | | 1,0000 | | | 1 | 61,957 | 0,0025 |
| 10 | 0,0077 | 0,9829 | - 2,0258 | 0,9890 | - 2,2506 | - 2,1978 | 0,9878 | 56,987 | 0,0018 |
| 15 | 0,0114 | 0,9754 | - 1,8394 | 0,9805 | - 1,9585 | - 1,9153 | 0,9788 | 52,071 | 0,0027 |
| 20 | 0,0137 | 0,9643 | - 1,6477 | 0,9681 | - 1,7060 | - 1,6710 | 0,9658 | 47,209 | 0,0035 |
| 25 | 0,0158 | 0,9511 | - 1,4836 | 0,9519 | - 1,4928 | - 1,4649 | 0,9493 | 42,421 | 0,0039 |
| 30 | 0,0216 | 0,9361 | - 1,3419 | 0,9337 | - 1,3226 | - 1,3003 | 0,9309 | 37,721 | 0,0045 |
| 35 | 0,0254 | 0,9159 | - 1,1938 | 0,9132 | - 1,1766 | - 1,1590 | 0,9104 | 33,118 | 0,0051 |
| 40 | 0,0314 | 0,8926 | - 1,0588 | 0,8899 | - 1,0447 | - 1,0314 | 0,8872 | 28,624 | 0,0061 |
| 45 | 0,0407 | 0,8646 | - 0,9269 | 0,8628 | - 0,9194 | - 0,9103 | 0,8606 | 24,254 | 0,0076 |
| 50 | 0,0522 | 0,8294 | - 0,7906 | 0,8299 | - 0,7925 | - 0,7875 | 0,8285 | 20,031 | 0,0105 |
| 55 | 0,0765 | 0,7861 | - 0,6507 | 0,7863 | - 0,6514 | - 0,6511 | 0,7862 | 15,994 | 0,0146 |
| 60 | 0,1122 | 0,7259 | - 0,4870 | 0,7289 | - 0,4946 | - 0,4995 | 0,7308 | 12,202 | 0,0222 |
| 65 | 0,1493 | 0,6445 | - 0,2974 | 0,6490 | - 0,3074 | - 0,3184 | 0,6540 | 8,740 | 0,0339 |
| 70 | 0,2158 | 0,5482 | - 0,0968 | 0,5427 | - 0,0856 | - 0,1039 | 0,5517 | 5,725 | 0,0545 |
| 75 | #N/A | 0,4299 | 0,1411 | 0,4062 | 0,1898 | 0,1625 | 0,4194 | 3,297 | 0,0871 |
| 80 | #N/A | #N/A | #N/A | 0,2545 | 0,5373 | 0,4986 | 0,2695 | 1,575 | 0,1370 |
| 85 | #N/A | #N/A | #N/A | 0,1201 | 0,9956 | 0,9419 | 0,1320 | 0,571 | 0,2084 |

Les coefficients α et β sont respectivement l'ordonnée à l'origine et la pente de la droite de régression liant le logit des probabilités de survie qui apparaissent dans les colonnes 4 et 6 du tableau 20.4, pour l'intervalle d'âge choisi par l'utilisateur, comme entre 45 et 75 ans dans cet exemple, soit $\alpha = 0,0211$ et $\beta = 0,9672$.

Ces coefficients sont ensuite appliqués à la transformation logit de la table de mortalité conditionnelle pour fournir les logits lissés qui figurent dans la colonne 7 du tableau 20.4. Ainsi, par exemple, le logit lissé à l'âge de 20 ans se calcule comme suit :

$$\gamma_{20}^{fitted} = -0,0211 + 0,9672 \times (-1,7060) = -1,671.$$

Ces valeurs sont ensuite utilisées pour produire la table de mortalité lissée, qui apparaît en colonne 8 du tableau 20.4. Par exemple, la valeur à l'âge de 20 ans se calcule comme suit :

$$l_{20}^{fitted} = \frac{1}{\exp(1 + 2 \times (-1,671))} = 0,9658.$$

Les personnes-années cumulées au-delà de l'âge x , notées T_x , apparaissent en colonne 9 du tableau 20.4. Elles sont calculées à partir de la table de mortalité lissée, et ces nombres sont utilisés pour produire les taux de mortalité lissés qui apparaissent en colonne 10 du tableau 20.4. Par exemple, pour l'âge de 80 ans :

$$T_{80} = 0,577 + \frac{5}{2}(0,2695 + 0,132) = 1,575$$

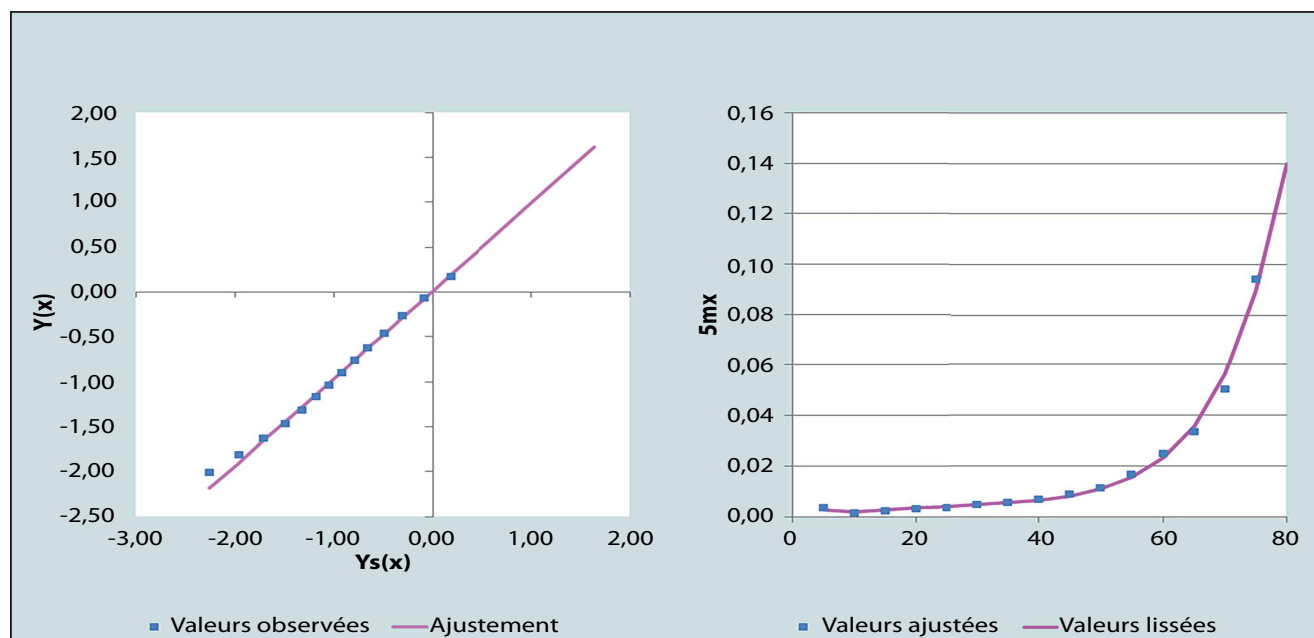
$${}_5m_{80}^{fitted} = \frac{0,2695 - 0,132}{1,575 - 0,571} = 0,137.$$

DIAGNOSTICS, ANALYSE ET INTERPRÉTATION

Contrôles et validation

L'exemple ci-dessus est tiré du *Manuel X* des Nations Unies (Division de la Population des Nations Unies 1984), qui donnait une estimation de la complétude de 83 % après avoir appliqué cette méthode et celle de Preston et Coale. La différence entre ces estimations antérieures et l'estimation présentée ici (soit 93 %) semble due en grande partie à la

FIGURE 20.3 Logit et taux de mortalité féminine ajustés à l'aide du modèle Ouest des tables-type de mortalité de Princeton, El Salvador, Recensement de 1961



méthode utilisée pour ajuster la droite et le choix des points retenus pour l'ajustement. La méthode mise en œuvre dans la feuille de calcul (*BGB_El Salvador*) disponible sur le site web associé à cet ouvrage utilise une régression orthogonale, alors que le *Manuel X* utilise une procédure plus simple, qui consiste à utiliser des « moyennes groupées » jusqu'à l'âge de 60 ans, et des « moyennes tronquées », ce qui revient en fait à supprimer l'impact du dernier point. Cette différence suggère que l'on pourrait proposer d'ignorer les deux derniers points dans cet exemple puisque la droite de régression est fortement influencée par les points extrêmes sur les axes. Il faut cependant rappeler, comme cela a été noté plus haut, que lorsqu'on utilise une régression orthogonale l'effet d'ignorer ces points reste mineur.

Interprétation

Un problème qui se pose souvent lorsqu'on doit décider quelle limite d'âge supérieure retenir pour ajuster la ligne droite est le fait que la complétude peut varier considérablement du fait de l'exclusion d'un seul point. Par exemple, si on choisit 70 ans et plus comme intervalle ouvert, le graphique permettant de porter un diagnostic ressemblerait à celui qui apparaît dans la

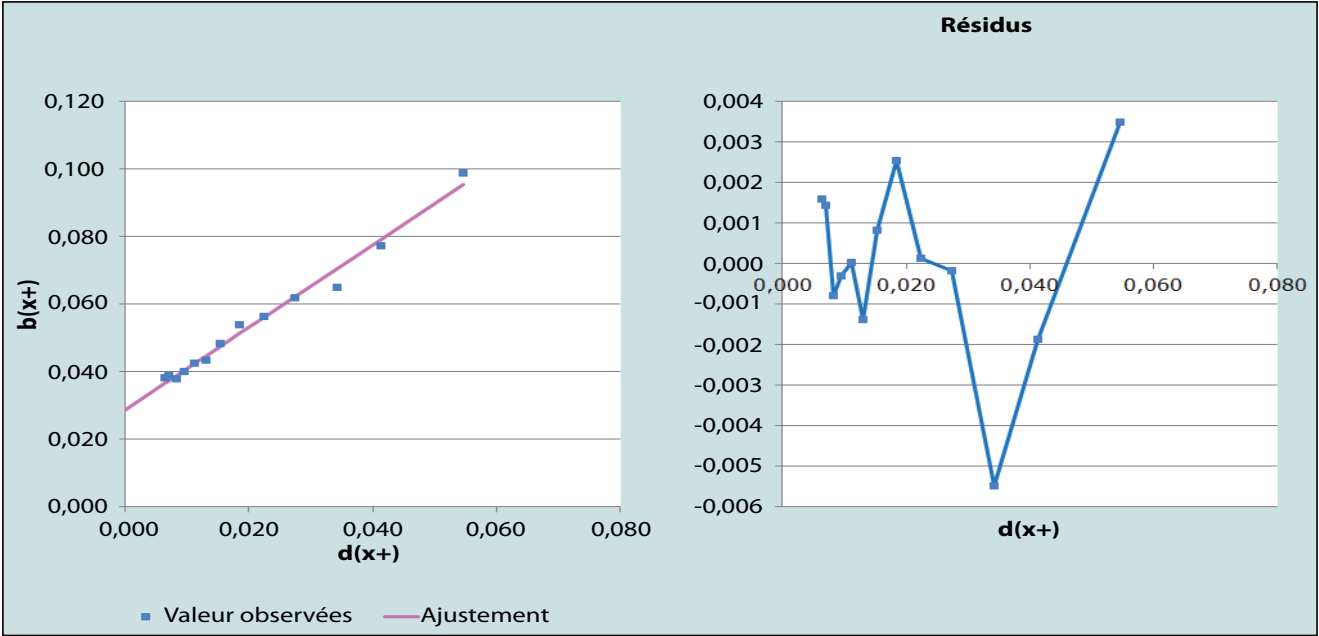
figure 20.4, et l'estimation de la complétude serait de 82 %. En elle-même, ce graphique n'indique pas que l'ajustement est meilleur ou pire que celui obtenu avec le groupe d'âge ouvert de 65 ans et plus. Dans de tels cas, on devrait refaire les calculs pour plusieurs intervalles ouverts, et choisir celui qui est le plus proche de la moyenne ou de la médiane. Dans le cas mentionné ci-dessus, l'estimation de la complétude calculée avec l'intervalle de 60 ans et plus est de 91 %, ce qui suggère que l'enregistrement des décès est approximativement complet à 90 %. Cependant, comme cela a été mentionné plus haut, on recommande d'une manière générale de ne pas terminer par un groupe d'âge se terminant par zéro dans les populations qui font preuve d'une forte préférence pour les âges se terminant par un chiffre rond.

**PROBLÈMES D'INTERPRÉTATION
DÛS À LA MÉTHODE**

Source des données sur les décès déclarés

En général, on doit faire face à deux sortes de problèmes relatifs aux données sur les décès : les cas où le sous-enregistrement (ou le sur-enregistrement) est constant avec l'âge, ce qui

FIGURE 20.4 Diagnostic graphique, avec un groupe d'âge ouvert de 70 ans et plus, El Salvador, Recensement de 1961



est précisément le cas pour lequel la méthode est conçue, et les autres cas, où on observe une complétude variant selon l'âge, qui peut introduire des distorsions dans les estimations. Bien que l'approche générale demeure essentiellement la même quelle que soit la source des données sur les décès, les différentes sources ont chacune leurs propres biais, qui ont un impact sur l'interprétation des résultats. Ces différences sont illustrées ici par des exemples particuliers, mais en règle générale il convient de faire attention aux biais exposés ci-dessous.

1) Données de l'état civil

Lorsque la répartition de la population entre les milieux urbain et rural (ou des catégories semblables) diffère significativement selon l'âge, et que la complétude de la déclaration des décès en milieu urbain est nettement plus forte qu'en milieu rural, alors l'hypothèse que la complétude est indépendante de l'âge est vraisemblablement violée, avec une baisse de la complétude au-delà de 50 ans. Ceci est souvent le cas lorsque une proportion importante de personnes déménagent de l'urbain vers le rural au moment de la retraite. Si on ignore ce fait, on peut obtenir une sous-estimation du niveau moyen de la complétude.

2) Décès déclarés dans les recensements et enquêtes auprès des ménages

Ce type de données est potentiellement sujet à trois problèmes :

- Si une proportion importante de ménages se trouve dissuade à un décès d'une personne clé (par exemple celui ou celle qui pourvoit aux dépenses), alors les décès de ces personnes seront sous-déclarés, et l'hypothèse de l'indépendance de la complétude avec l'âge ne sera plus vérifiée. Si une proportion importante de décès dans certains groupes d'âge élevés relève de personnes qui ne vivent pas dans des ménages ordinaires, par exemple ceux qui vivent dans des maisons de retraite pour personnes âgées, la violation de l'hypothèse est encore plus forte. Mais ceci n'est pas encore un problème dans la majorité des pays en développement.
- Dans les situations où les jeunes adultes quittent la maison familiale où ils ont grandi pour aller travailler en milieu urbain, il est possible qu'ils soient considérés comme membres de plusieurs ménages (voire d'aucun ménage), et que leur décès soit déclaré plusieurs fois (voire omis), ce qui à nouveau entraîne une violation de l'hypothèse de complétude constante par âge. Dans ce cas, on peut en limiter l'effet en ignorant les données en dessous d'un certain âge pour estimer la complétude.

- Il peut souvent y avoir confusion sur la période exacte de référence au cours de laquelle les décès doivent être déclarés, en plus des erreurs sur l'âge au décès. Ce fait peut conduire soit à un sous-enregistrement, soit à un sur-enregistrement des décès. Cependant, lorsque ces biais sont indépendants de l'âge des décédés, cette distorsion sera prise en compte dans l'estimation de la complétude, et ne devrait pas poser de problème pour estimer les taux de mortalité corrigés.

3) Décès enregistrés dans les structures sanitaires

On sait peu de choses sur cette source de données. On peut cependant supposer que la complétude va dépendre de la répartition géographique des structures sanitaires auprès desquelles ces données seront recueillies. Dans de nombreux pays en développement, ces services sont souvent concentrés dans les zones urbaines. Ainsi donc, si la proportion de personnes vivant en milieu urbain plutôt qu'en milieu rural varie selon l'âge, la complétude ne sera de nouveau pas indépendante de l'âge. Il est aussi possible que certaines causes de décès prédominent dans les structures sanitaires, qui, si elles sont importantes quantitativement et sont spécifiques à certains âges, vont probablement violer encore plus l'hypothèse d'indépendance de la complétude constante selon l'âge.

Dans tous ces cas, les points figurant sur le graphique vont se situer de plus en plus au-dessus de la droite de régression aux âges élevés, ce qui risque de conduire à une sous-estimation de la complétude. Mais on pourra améliorer l'estimation en excluant les points qui représentent les âges élevés pour le choix de la droite d'ajustement, quoique l'estimation restera encore biaisée malgré tout.

DESCRIPTION DÉTAILLÉE DE LA MÉTHODE

Exposé mathématique

Quoique la méthode de Brass de la balance de l'accroissement démographique (*growth balance method*) soit simplement un cas particulier de la méthode généralisée de la balance de l'accroissement démographique (*generalized growth balance method*), avec un taux de croissance de la population âgée de x années et plus ($x+$) qui reste constant à tous les âges, il peut être utile de bien décrire ce cas spécifique pour assurer une bonne compréhension.

L'origine de la méthode de Brass de la balance de l'accroissement démographique (Brass 1975) se trouve dans les travaux de Carrier (1958) qui, le premier, proposa une

méthode d'estimation de la mortalité en utilisant la distribution par âge des décès. Cette méthode part de la relation que l'on trouve dans l'équation d'équilibre d'une population fermée aux migrations. Dans une telle population, le nombre de personnes dans la population au temps t_2 est égal au nombre au temps t_1 , plus les naissances qui se sont produites entre t_1 et t_2 , et moins les décès qui se sont produits entre t_1 et t_2 , soit :

$${}_t N_0(t_2) = {}_t N_0(t_1) + B - D$$

où B sont les naissances et D les décès qui se sont produites entre t_1 et t_2 . Cette équation peut se généraliser à tous les groupes d'âge $x+$. Ceci suppose que l'on ait une mesure du nombre de personnes qui atteignent l'âge x au cours de l'intervalle de temps situé entre t_1 et t_2 , noté N_x , ainsi que du nombre de décès d'âge supérieur ou égal à x qui se sont produits entre t_1 et t_2 , noté : ${}_t D_x$.

La formule devient alors :

$${}_t N_x(t_2) = {}_t N_x(t_1) + N_x - {}_t D_x \quad (1)$$

L'équation (1) peut aussi s'écrire :

$${}_t N_x(t_2) - {}_t N_x(t_1) = N_x - {}_t D_x$$

et en divisant par le nombre de personnes-années vécues au risqué entre t_1 et t_2 , on peut exprimer l'équation de la balance de l'accroissement démographique comme suit :

$$r(x+) = b(x+) - d(x+) \quad (2)$$

où les taux sont calculés comme :

$$r(x+) = \frac{{}_t N_x(t_2) - {}_t N_x(t_1)}{\int_{t_1}^{t_2} {}_t N_x(t) dt}$$

$$b(x+) \approx \frac{N_x}{\int_{t_1}^{t_2} {}_t N_x(t) dt}$$

et

$$d(x+) \approx \frac{{}_t D_x}{\int_{t_1}^{t_2} {}_t N_x(t) dt}$$

On appelle souvent les taux $b(x+)$ et $d(x+)$ les « taux partiels » de natalité et de mortalité respectivement, ou parfois les « taux segmentiels » (*segmental birth and death rates*).

Ces relations ne sont vérifiées que dans le cas où l'enregistrement de la population (c'est-à-dire celle qui atteint son anniversaire) et des décès est complet entre t_1 et t_2 .

Supposons maintenant que toutes les données soient complètes, sauf les décès qui sont enregistrés avec imprécision. Supposons de plus que l'on puisse faire l'hypothèse qu'une proportion constante, notée c , de décès est déclarée, indépendamment de l'âge au décès, pour tous les âges sauf au-delà d'un certain âge et confiné aux personnes adultes. Alors, ${}_t D_x = {}_t D_x^r / c$ ou ${}_t D_x^r$ représente le nombre de décès effectivement déclarés d'âge supérieur ou égal à x .

L'équation 2 devient alors :

$$r(x+) = b(x+) - d^r(x+) / c$$

où

$$d^r(x+) = \frac{{}_t D_x^r}{\int_{t_1}^{t_2} {}_t N_x(t) dt}$$

Si l'on suppose en outre que la population est stable, et croissante à un taux annuel constant r , alors l'équation peut être réécrite comme suit : $b(x+) = r + k d^r(x+)$. En conséquence, on doit pouvoir ajuster une ligne droite entre les points de coordonnées : $(b(x+), d^r(x+))$. L'ordonnée à l'origine de la droite donnera la valeur du taux de croissance r , et l'inverse de la pente k donnera la valeur de la complétude de l'enregistrement des décès c .

Les taux de mortalité par groupe d'âge quinquennal, notés, ${}_5 m_x$ se calculent alors selon la formule :

$${}_5 m_x = \frac{{}_5 D_x^r / c}{\int_{t_1}^{t_2} {}_t N_x(t) dt}$$

MISE EN ŒUVRE DE LA MÉTHODE

En pratique, on dispose des données suivantes : le nombre de décès déclarés au cours de la période allant de t_1 à t_2 , présentés par groupes d'âge quinquennaux, notés ${}_5 D_x^r$; jusqu'à un dernier intervalle d'âge ouvert au-delà de l'âge A , noté : ${}_t D_A^r$; et les nombres de personnes dans la population au milieu de la période, présentés selon les mêmes groupes d'âge, notés : ${}_5 N_x$ jusqu'à ${}_t N_A$. On part de ces données pour effectuer tous les calculs relatifs à la méthode : calcul des effectifs cumulés : ${}_t D_x^r$ et ${}_t N_x$ des personnes atteignant leur anniversaire : $N_x = ({}_5 N_{x-5} + {}_5 N_x) / 10$

ou $N_x = ({}_5N_{x-5} \cdot {}_5N_x)^{\frac{1}{2}} / 5$ et des personnes-années vécues cumulées :

$$\int_{t_1}^{t_2} {}_{\infty}N_x(t)dt = {}_{\infty}N_x(t_2 - t_1).$$

Dans le cas où on ne dispose pas de la population en milieu de période, mais seulement à un autre point dans le temps, disons t , on peut tout de même appliquer la méthode avec cette population. La seule différence est que l'estimation de la complétude sera relative à cette population au temps t , comme si elle représentait le milieu de période. En d'autres termes, en supposant la population stable, on a :

$${}_{\infty}N_x(t) = {}_{\infty}N_x e^{r(t-(t_2-t_1)/2)}$$

et la complétude relative de cette population devient :

$$ce^{r(t-(t_2-t_1)/2)}$$

Les taux de mortalité obtenus en divisant les nombre de décès corrigés pour ce niveau de complétude par ${}_{\infty}N_x(t)$ fourniront les mêmes taux que les taux que l'on aurait obtenus si on avait utilisé la population par âge au milieu de la période.

Ajustement de la ligne droite

Deux questions se posent pour déterminer la ligne droite qui représente au mieux la relation entre les taux partiels de natalité et de mortalité, à savoir : le choix de la méthode d'ajustement, et le choix des points utilisés pour calculer la pente et l'ordonnée à l'origine.

Ajuster une droite de régression en utilisant la méthode des moindres carrés sans pondération n'est pas recommandé, car la méthode donnerait trop de poids aux valeurs obtenues aux âges élevés qui sont moins fiables. On recommande plutôt d'ajuster une droite en utilisant une méthode plus robuste, comme la droite « moyenne », c'est-à-dire la droite qui joint les points moyens obtenus en répartissant, en fonction de l'âge, les points en deux groupes d'importance égale, et en prenant la moyenne des coordonnées horizontales et verticales de ces deux séries de points. On peut aussi utiliser la méthode de la « moyenne tronquée », c'est-à-dire une méthode analogue à celle de la droite « moyenne », mais où on donne une pondération inférieure aux points qui sont moins fiables, en général les extrémités de l'intervalle. Ces méthodes sont expliquées

en détail dans le *Manuel X* (Division de la Population des Nations Unies 1984). Une autre alternative est décrite en détail en anglais dans *Methods for Estimating Adult Mortality* (UN Population Division 2002 : 105-110). Cette alternative est semblable à celle de la droite « moyenne », mais ici on divise les points en trois groupes de mêmes effectifs,¹ et on détermine la droite qui joint les médianes des points des groupes extrêmes, c'est-à-dire du premier et du dernier tiers.

Mari Bhat signale que chaque méthode a ses points faibles (Bhat 2002). Puisque les taux partiels de natalité et de mortalité sont traités tous deux comme des variables dépendantes, il propose d'utiliser la régression orthogonale comme la meilleure méthode pour tenir compte d'éventuelles erreurs sur les âges. Cette stratégie permet de tenir compte à la fois des distances horizontales et verticales à la ligne de régression, puisqu'elle minimise la somme des carrés des résidus orthogonaux (ORSS), définis comme suit :

$$\sum_i \frac{x_i^2 y_i^2}{x_i^2 + y_i^2}.$$

L'utilisation de cette méthode conduit à calculer la complétude de la déclaration des décès comme le rapport des écarts-types des taux partiels de mortalité et de natalité. La valeur de l'ordonnée à l'origine est la différence entre la moyenne des taux partiels de natalité et la moyenne des taux partiels mortalité divisés par c . C'est cette approche qui est utilisée dans l'application de la méthode de Brass de la balance de l'accroissement démographique présentée dans le classeur Excel disponible sur le site web associé à ce manuel.

Limites de la méthode

Les principales limites de la méthode ont été décrites ci-dessus, et on en tient compte dans la feuille de calcul. Ce sont les conditions suivantes : la population doit être stable et fermée aux migrations. La méthode ne devrait pas être utilisée lorsque ces conditions ne sont pas vraiment remplies. Un bon exemple d'un usage de cette méthode qui serait inapproprié vient des données d'Afrique du Sud, pour les décès déclarés entre le recensement de 2001 et l'enquête dite de communauté de 2007, qui est un recensement sur un

1 Lorsqu'il n'est pas possible de diviser le nombre total de points, n , en trois groupes de même taille, on peut alors prendre comme groupes ceux formés par les k premiers et k derniers points, k étant calculé comme : $\text{ent}(n/3) + 1$.

large échantillon de population. (Ces données sont présentées dans le classeur Excel : *GGB_hommes Sud Africains*). Le calcul de la population en milieu de période obtenue en prenant la moyenne des deux recensements donne une estimation de la complétude de 85 %. Si on augmente l'âge minimal utilisé pour ajuster la droite à 35 ans, la complétude augmente à 88 %, mais reste inférieure à l'estimation obtenue en utilisant la méthode généralisée de la balance de l'accroissement démographique (chapitre 24) qui donne 91 %.

Cette méthode est moins vulnérable aux erreurs sur l'âge que celle proposée par Preston et Coale (chapitre 21). En particulier, par exemple, la tendance fréquente à exagérer les âges déclarés au décès (relativement à ceux déclarés au recensement) se manifestera par des points sur le graphique plus à droite (c'est-à-dire situés en-dessous de la ligne) pour tous les âges surestimés. Ces points pourront être éliminés lorsqu'on choisira les points à garder pour ajuster la droite. Par contre, cette méthode est moins vulnérable aux effets de l'écart à la situation stable qui résultent d'une baisse rapide de la mortalité (Martin 1980). Dans ce cas, on tendra à sous-estimer la complétude, car la plus faible mortalité sera interprétée par le modèle comme une augmentation du sous-enregistrement, mesurée par une pente plus forte. Cependant, des simulations ont montré que les biais résultants d'une amélioration lente et régulière de la mortalité, comme certains pays en développement en font l'expérience en l'absence d'épidémie, de famine et de guerre, seront peu importants (Rashad 1978).

En ce qui concerne les changements des taux de fécondité, en supposant que ceux-ci se sont produits au cours des 15 années précédentes, ils auront peu d'impact sur les performances de la méthode car ils ne vont affecter que les groupes d'âges les plus jeunes.

Les migrations vont affecter plutôt la population de jeunes adultes, surtout entre 20 et 35 ans, mais auront moins d'effet sur les décès qui eux se produisent surtout aux âges élevés. Les immigrations dont on ne peut pas tenir compte vont tendre à faire diminuer la pente, et par conséquent à surestimer la complétude de l'enregistrement des décès, et à sous-estimer les taux de mortalité. Par contre, les émigrations dont on ne peut pas tenir compte vont tendre à avoir l'effet inverse. Certains démographes recommandent d'ajuster la ligne droite aussi loin que possible à gauche, jusqu'à l'âge de 5 ans, pour réduire les effets des migrations inconnues. Ceci suppose que les différences de complétude de

l'enregistrement des décès entre les enfants et les personnes âgées n'induisent pas de distorsion majeure, car la mortalité est très faible entre 5 et 14 ans. Mais il est peu probable que cette adaptation de la méthode puisse réellement permettre de réduire les biais.

Une stratégie alternative consisterait à limiter les points au-delà de l'âge de 35 ans, de manière à réduire l'essentiel de l'impact potentiel des migrations. Cependant, la qualité des données aux âges élevés est souvent suspecte, ce qui diminue la fiabilité des estimations de la complétude. Il reste possible que ces adaptations produisent de meilleurs estimations que celles obtenues en ignorant les migrations, mais on ne dispose que de peu de recherches sur ce point, et sur la précision des estimations de la complétude obtenues en procédant à de telles adaptations.

D'un point de vue technique, si on dispose de données fiables sur les migrations nettes par âge, on devrait pouvoir adapter la méthode en remplaçant les taux partiels de natalité $b(x+)$ par $b(x+) - i(x+)$, où $i(x+)$ est le taux net de migration dans le groupe d'âge correspondant, pour ajuster une nouvelle ligne droite. Mais en pratique, dans les situations où il faudrait faire cette adaptation, on ne dispose que rarement de données suffisamment fiables de migrations nettes par âge pour justifier cette adaptation.

Les variations de la complétude de l'enregistrement des décès selon l'âge peuvent introduire une courbure dans l'alignement des points du graphique. En conséquence, une des forces de cette méthode est que si les points qui représentent les groupes d'âge successifs sont approximativement situés sur une ligne droite, alors on peut raisonnablement supposer que la complétude est approximativement constante selon l'âge. Mais si certains points sont sur la ligne, et que d'autres en sont éloignés, une méthode pour décider quels points rejeter consiste à calculer le taux partiel de croissance pour chacun des intervalles ouverts successifs, et de garder seulement les points pour lesquels les valeurs de r_{a+} sont assez cohérentes.

Une des principales limites de la méthode est que le graphique liant les taux partiels de natalité et les taux partiels de mortalité, à part les exceptions mentionnées ci-dessus, reste un outil de diagnostic peu performant dans les cas extrêmes. En particulier, la simulation d'une augmentation rapide de la mortalité due au VIH/sida, avec une prévalence de 11 % en population générale, à partir d'une population stable, produit des points qui en tous cas sont bien alignés, mais qui conduisent à sous-estimer le niveau de complétude,

même si on restreint l'analyse aux âges au-delà de 45 ans. Les leçons à tirer de tout ceci sont que, si les points ne sont pas alignés, on a probablement des données déficientes, mais si les points sont bien alignés on ne peut pas être certain que les estimations de la complétude seront correctes.

Extensions

Lorsqu'on dispose indépendamment de données précises et fiables sur le taux de croissance r , et si la population est

stable, alors on peut reformuler l'équation 2 pour estimer $c(x+)$, la complétude de chaque groupe d'âge ouvert, comme suit :

$$c(x+) = {}_{\infty} D_x^r / (N_x - r {}_{\infty} N_x).$$

Mais, en pratique il est rare de trouver une population suffisamment stable avec des âges suffisamment précis pour rendre utile cet exercice.

AUTRES LECTURES ET RÉFÉRENCES

Comme cette méthode est un cas particulier de la méthode généralisée de la balance de l'accroissement démographique, on renvoie le lecteur à la section correspondante pour y trouver des informations complémentaires.

- Bhat M. 2002. "General Growth Balance method: A reformulation for populations open to migration", *Population Studies* 56(1):23–34. doi: <http://dx.doi.org/10.1080/00324720213798>
- Brass W. 1975. *Methods for Estimating Fertility and Mortality from Limited and Defective Data*. Chapel Hill NC: Carolina Population Centre.
- Carrier NH. 1958. "A note on the estimation of mortality and other population characteristics, given death by age", *Population Studies* 12:149–163. doi: <http://dx.doi.org/10.2307/2172187>
- Division de la Population des Nations Unies. 1984. *Manuel X. Techniques indirectes d'estimation démographique*. New York :

- Nations Unies, Département des affaires économiques et sociales internationales, ST/ESA/SER.A/81. http://unstats.un.org/unsd/demographic/standmeth/handbooks/Manuel_X-fr.pdf
- Martin LG. 1980. "A modification for use in destabilized populations of Brass's Technique for estimating completeness of death registration", *Population Studies* 34:381–395. doi: <http://dx.doi.org/10.2307/2175194>
- Rashad HM. 1978. "The Estimation of Adult Mortality from Defective Registration Data." Unpublished PhD thesis, London: University of London.
- UN Population Division. 2002. *Methods for Estimating Adult Mortality*. New York: United Nations, Department of Economic and Social Affairs, ESA/P/WP.175. http://www.un.org/esa/population/techcoop/DemEst/methods_adultmort/methods_adultmort.html

Chapitre 21 LA MÉTHODE DE PRESTON ET COALE

Rob Dorrington

DESCRIPTION DE LA MÉTHODE

La méthode dite de Preston et Coale (Preston, Coale, Trussell *et al.* 1980) est la seconde des méthodes connues maintenant sous le nom des méthodes de la répartition des décès. Ces méthodes visent à estimer la complétude de la déclaration des décès par rapport à une estimation de la population faite à un instant donné. La méthode de Preston et Coale part de l'observation que le nombre de personnes d'un âge donné vivantes à un instant donné doit être égal au nombre cumulé de décès à venir dans cette cohorte à partir de cet instant. Si la population est stable (c'est-à-dire si la structure par âge est constante, au moins pour les âges adultes, croissante à un taux annuel r , et fermée aux migrations), et si l'enregistrement des décès est complet et précis, alors le nombre de décès d'âge x au cours des t années à venir sera égal au nombre de décès actuel multiplié par e^{rt} . Il est alors possible d'estimer la population actuelle d'âge y à partir des décès au-delà de l'âge y et du taux de croissance r . Si les décès sont sous-enregistrés, et que l'on peut faire l'hypothèse d'un sous-enregistrement constant c à tous les âges, alors le nombre de décès à venir dans la cohorte sera lui aussi sous-estimé dans la même proportion. Il sera donc possible d'estimer la complétude de l'enregistrement des décès en divisant la somme des décès attendus dans la cohorte par le nombre de décès enregistrés au cours d'une période donnée. Les taux de mortalité peuvent dès lors être calculés en divisant le nombre de décès déclarés dans chaque groupe d'âge par le coefficient c , puis en divisant ces nombres corrigés par une estimation de la population exposée au risque.

Cette méthode est un cas particulier du cas plus général de la méthode de l'extinction des cohortes synthétiques (chapitre 25), qui, elle, demande des estimations de la population à deux instants donnés, mais n'exige pas que la population soit stable. Le lecteur est invité à consulter la section correspondante pour avoir plus d'informations sur cette méthode plus générale. La méthode présentée ici est celle

que l'on peut utiliser lorsqu'on ne dispose d'estimations de la population qu'à un seul instant donné.

DONNÉES NÉCESSAIRES ET HYPOTHÈSES

Tabulations des données nécessaires

- Nombre de décès de femmes (ou d'hommes) au cours d'une période donnée, répartis par groupe d'âge de 5 ans, et pour le groupe d'âge ouvert $A+$ (A étant aussi élevé que possible).
- Effectifs de la population féminine (ou masculine) à un instant donné, aussi proche que possible de la période au cours de laquelle les décès sont enregistrés, également répartis par groupe d'âge de 5 ans, et pour le groupe d'âge ouvert $A+$.

Hypothèses importantes

- La population est stable, quoique cette hypothèse puisse être quelque peu relâchée (voir ci-dessous).
- La complétude de l'enregistrement des décès est la même pour tous les âges au-delà d'un âge minimal (en général 15 ans).
- La population est fermée aux migrations, quoique cette hypothèse puisse être relâchée si les migrations nettes sont négligeables par rapport aux taux de mortalité, ou si on dispose d'une estimation suffisamment fiable du nombre de migrants par âge pour permettre de corriger l'équation de la balance de l'accroissement démographique (ce qui est rarement le cas).

TRAVAUX PRÉPARATOIRES ET RECHERCHES PRÉLIMINAIRES

Avant de mettre en œuvre cette méthode, on doit analyser la qualité des données, au moins en ce qui concerne les points suivants :

- structure par âge de la population ;
- répartition par sexe de la population ;
- structure par âge des décès ; et
- répartition par sexe des décès.

PRÉCAUTIONS ET MISES EN GARDE

Lorsqu'ils appliquent cette méthode, les analystes doivent prendre en considération les points suivants :

- Le processus d'interprétation et d'estimation doit prendre en compte la source des données de mortalité (décès déclarés à l'état civil, décès déclarés par les ménages dans les recensements, ou décès déclarés dans les structures sanitaires), comme cela est expliqué plus loin.
- Le problème des migrations devient plus sérieux quand on applique la méthode à des zones géographiques régionales, et non plus nationales.
- Le choix de l'intervalle d'âges sur lequel on calcule le taux de croissance est important (c'est-à-dire celui qui minimise la différence absolue entre la complétude moyenne et la complétude calculée sur chaque groupe d'âge). Le problème ici est de savoir si la meilleure estimation du taux de croissance à retenir est celle qui est fournie par l'ordonnée à l'origine calculée sur les mêmes données en utilisant la méthode de la balance de l'accroissement démographique de Brass (voir le chapitre 20). Ceci serait le cas si on pensait que la complétude diminue chez les personnes âgées, comme dans le cas des retraités. Un autre problème est de savoir s'il convient d'exclure les âges en dessous de 30 ou 35 ans du fait de l'impact des migrations que l'on ne peut pas corriger par manque de données.
- Il convient de choisir l'intervalle d'âges à utiliser pour calculer la complétude. En pratique, on devra souvent exclure les jeunes adultes lorsque les migrations sont importantes, et les personnes âgées si les résultats montrent que les décès sont moins bien déclarés à ces âges que chez les jeunes adultes.
- Il faut s'assurer que le module *Solveur* du logiciel *Excel* a fonctionné correctement, c'est-à-dire qu'il fournit des résultats réalistes. Il peut arriver que ce logiciel produise une solution qui est manifestement trop faible. Dans un tel cas, il sera préférable d'ajuster d'abord le coefficient *delta* à la main dans une direction appropriée, puis d'appliquer le module *Solveur* à cette nouvelle valeur.
- Il faut également s'assurer que l'estimation de l'espérance de vie dans le groupe d'âge ouvert est réaliste. Il est fré-

quent que les données concernant les personnes âgées soient déficientes et soient sujettes à des erreurs particulières. Dans ce cas, les estimations de l'espérance de vie calculées d'après ces données erronées sont irréalistes (trop élevées le plus souvent).

- Si la complétude de l'enregistrement des décès tombe en dessous de 60 %, on invite à la prudence dans l'application de la méthode, car la marge d'erreur de cette estimation peut être considérable.

APPLICATION DE LA MÉTHODE

La méthode s'applique en suivant les étapes détaillées ci-dessous :

Étape 1 : Calculer le taux de croissance initial

Au départ, le taux de croissance peut se calculer soit à partir de la croissance de la population au-dessus d'un âge donné (choisi pour ajuster au mieux l'hypothèse de stabilité de la population) au cours d'une période, soit à partir de l'estimation fournie par la méthode de la balance de l'accroissement démographique de Brass. Dans le premier cas, si l'on dispose de la population à deux instants donnés (t_1 et t_2), le taux de croissance se calcule comme suit :

$$r = \frac{\ln(\infty N_x(t_2) / \infty N_x(t_1))}{t_2 - t_1}$$

où $\infty N_x(t)$ est la population âgée de x années et plus au temps t .

Étape 2 : Estimer l'espérance de vie à l'âge A et dans les groupes d'âge quinquennaux en remontant jusqu'à 65 ans

Ce calcul peut se faire de plusieurs manières :

1. Utiliser des estimations provenant de sources indépendantes si elles sont fiables. On pourra utiliser des estimations faites au cours de recherches antérieures, ou des projections de population telles que les « Perspectives de la Population Mondiale » réalisées par les Nations Unies (UN Population Division 2011).
2. Utiliser l'estimation fournie par la méthode de la balance de l'accroissement démographique de Brass, appliquée aux mêmes données. Le classeur Excel qui met en œuvre cette méthode produit de telles estimations parmi ses différents résultats.
3. Utiliser une table-type de mortalité. On peut par exemple partir du rapport entre les décès de 10 à 39 ans (âge au dernier anniversaire) aux décès de 40 à 59 ans (âge au

dernier anniversaire), noté ${}_{30}D_{10/20}D_{40}$, pour trouver le niveau correspondant de la table de mortalité dans le modèle Ouest de Princeton, et en déduire ainsi les espérances de vie. Ce mode de calcul est inclus dans le classeur Excel de cette méthode. Mais comme la table-type du modèle Ouest ne reflète pas la mortalité due au VIH/sida, cette approche n'est pas appropriée pour les pays qui ont un nombre élevé de décès par sida.

4. Travailler par approximations successives pour trouver l'espérance de vie qui convient. On part d'une valeur raisonnable, par exemple à partir d'une table-type du modèle Ouest (quoique cette stratégie puisse ne pas marcher dans les pays fortement affectés par le sida), ou de sources indépendantes. On fait une première estimation de la complétude selon la méthode exposée ci-dessous, puis on copie et colle ces valeurs dans la feuille de calcul *Méthode* du même classeur Excel, et on effectue une nouvelle estimation de la complétude. On répétera la procédure plusieurs fois si nécessaire, jusqu'à ce que l'on obtienne une estimation stable de l'espérance de vie. Mais si on a de bonnes raisons de soupçonner que la mortalité reste sous-estimée aux âges élevés, même après la correction pour tenir compte de la sous-déclaration des décès (par exemple lorsque les âges élevés sont exagérés, ou que la complétude est beaucoup plus faible aux âges élevés), cette approche risque de surestimer les espérances de vie, et donc de surestimer le niveau général de complétude de l'enregistrement des décès.

Étape 3 : Estimer le nombre de personnes qui atteignent l'âge x , et le groupe d'âge de x à $x+4$, à partir du nombre de décès déclarés

Le nombre de personnes qui atteignent l'âge x (leur anniversaire) au cours de la période concernée (celle où les décès sont déclarés) s'estime à partir des décès déclarés comme suit :

$$\widehat{N}_x = \widehat{N}_{x+5} \exp(5r) + {}_5D_x \exp(2,5r)$$

et

$$\widehat{N}_A = {}_\infty D_A \left(\exp(r \times e_A) - (r \times e_A)^2 / 6 \right)$$

où A représente l'âge au début de l'intervalle ouvert, r est le taux annuel de croissance de la population, et e_A est l'espérance de vie à l'âge A .

Le nombre de personnes qui atteignent le groupe d'âge quinquennal entre x et $x+4$ (âge au dernier anniversaire) au

cours de la période concernée se calcule à partir des effectifs quinquennaux comme suit :

$${}_5\widehat{N}_x = 2,5(\widehat{N}_x + \widehat{N}_{x+5}).$$

Étape 4 : Estimation du nombre de personnes dans le groupe d'âge x à $x+4$ au dernier anniversaire au cours de la période concernée, où les décès sont enregistrés, à partir d'un recensement de population

Les effectifs de la population du groupe d'âge x à $x+4$ (âge au dernier anniversaire) au cours de la période concernée (celle où les décès sont enregistrés) sont calculés à partir d'un recensement de population, en multipliant les effectifs de chaque groupe d'âge par la durée de la période concernée (exprimée en années).

Étape 5 : Calcul des rapports entre les estimations de la population âgée de x à $x+4$ au dernier anniversaire et celles de la population âgée de x à $A-1$ au dernier anniversaire, mesurées d'après les décès enregistrés et d'après le recensement de population

On calcule les rapports des deux ensembles d'effectifs par âge : ceux provenant des décès déclarés et ceux provenant du recensement de population. On calcule d'abord directement les rapports des groupes d'âge quinquennaux. On calcule ensuite les rapports des effectifs dans les groupes d'âge cumulés entre l'âge x et l'âge A du début de l'intervalle ouvert et les effectifs cumulés des personnes qui atteignent le groupe d'âge entre x et $A-1$ au cours de la période concernée, qui se calcule comme la somme des effectifs dans les groupes d'âges quinquennaux situés entre x et $A-5$. En formule, cela donne :

$${}_{A-x}\widehat{N}_x = \sum_{a=x,5}^{A-5} {}_5\widehat{N}_a$$

Étape 6 : Estimation de la complétude de la déclaration de décès

Pour déterminer le niveau de complétude de la déclaration des décès, on doit d'abord décider si le choix initial du taux de croissance est correct. L'interprétation des graphiques des rapports est discutée en détail ci-dessous. En résumé, un taux de croissance correct se repère graphiquement comme celui qui produit un ensemble de rapports par âge de même niveau, c'est-à-dire horizontal. La feuille de calcul *Méthode* est faite pour réaliser ce calcul à l'aide du module *Solveur*

(Données / Solveur / Résoudre) : elle permet de trouver le taux de croissance qui minimise l'écart absolu par rapport à la moyenne des rapports sur un intervalle d'âge spécifié par l'utilisateur.

Si la valeur initiale choisie pour le taux de croissance produit une série de rapports de même niveau pour les groupes d'âges des jeunes adultes, mais qu'elle donne une courbure vers le bas aux âges élevés, cette situation peut indiquer une baisse de la complétude aux âges élevés (ce qui peut être le cas si, par exemple, les personnes qui prennent leur retraite migrent des zones urbaines vers les zones rurales où la complétude de l'enregistrement est plus faible). Dans de tels cas, il est important de ne pas prendre un taux de croissance qui produirait un ensemble de rapports de niveau très horizontal, mais de garder plutôt le taux de croissance choisi initialement.

Si l'utilisateur cherche à ajuster par itération à la fois le taux de croissance et l'espérance de vie, il faudra copier et coller les valeurs trouvées dans la feuille de calcul **Espérances de vie** dans la feuille de calcul **Méthode**, et calculer un nouveau taux de croissance. Il faudra alors répéter deux ou trois fois l'opération, jusqu'à ce que l'estimation de l'espérance de vie devienne stable.

Enfin, il faudra décider quel intervalle d'âge utiliser pour l'estimation de la complétude à partir de l'analyse graphique des rapports par âge. Si l'on observe une courbure significative vers le haut aux âges élevés, on est probablement en présence d'une surestimation de l'âge, surtout pour les décès. Il conviendra alors d'essayer d'identifier cet âge en dessous duquel l'exagération de l'âge n'est plus importante, et de le prendre pour début de l'intervalle ouvert. Si la complétude baisse en dessous de l'âge de 35 ans, il pourra s'agir d'émigrations nettes qui n'ont pas été prises en compte. Dans ce cas, il conviendra d'exclure ce groupe d'âge pour recalculer le taux de croissance et la complétude.

La complétude se calcule à partir des rapports par groupe d'âge. Afin de produire des estimations robustes, on propose de la calculer comme la somme pondérée suivante : 50 % de la médiane, plus 25 % de premier et du dernier quartile de la distribution de ces rapports (les quartiles sont définis comme les percentiles situés à 25 % et 75 % de la distribution).

Cependant, comme le calcul de la complétude se fait en prenant pour hypothèse que le recensement de population est réalisé au milieu de la période au cours de laquelle les

décès sont enregistrés, il est souhaitable d'effectuer une correction pour tenir compte de la croissance de la population entre l'instant précis du recensement (t_c) et l'instant correspondant au milieu de la période d'enregistrement des décès (t_m). Pour ce faire, on multiplie l'estimation de la complétude par le rapport des deux populations au temps t_c et t_m . En faisant l'hypothèse que la croissance est exponentielle au cours de l'intervalle, à un taux annuel constant a , cela donne : $\exp(a(t_c - t_m))$ où t_c est l'instant du recensement, et t_m est le milieu de la période au cours de laquelle les décès sont enregistrés.

Étape 7 : Estimation des taux de mortalité après correction du sous-enregistrement des décès

Afin de calculer les taux de mortalité corrigés, on doit d'abord estimer la population par groupes d'âge quinquennaux au milieu de la période au cours de laquelle les décès sont enregistrés, en multipliant les effectifs du recensement par : $\exp(-a(t_c - t_m))$.

Ensuite, il faut corriger les nombres de décès déclarés en les divisant par l'estimation de la complétude, c .

Les personne-années vécues au risque sont calculées en multipliant la population moyenne au temps t_m par la durée de la période au cours de laquelle les décès sont déclarés, t .

Les taux de mortalité corrigés du sous-enregistrement des décès sont alors calculés comme suit :

$${}_5m_x = \frac{{}_5D_x / c}{t \times {}_5N_x \exp(-a(t_c - t_m))}$$

Comme à la fois le numérateur (à travers l'estimation de c), et le dénominateur sont multipliés par le même coefficient : $\exp(-a(t_c - t_m))$, omettre ce coefficient haut et bas (dans les étapes 6 et 7) fournirait les mêmes taux de mortalité. Mais l'estimation de la complétude serait biaisée, et équivalente à ce qu'elle serait si la taille de la population au temps t_m était la même que celle mesurée au temps t_c .

Étape 8 : Lissage en utilisant un système logit de tables-type de mortalité

Lorsque les taux de mortalité par âge sont erratiques, ils doivent être lissés. Ceci peut se faire en ajustant un modèle logit (modèle relationnel de Brass) à un ensemble de taux de mortalité tirés d'une table-type, choisie en considérant qu'elle a approximativement le même schéma par âge que celle de la table de mortalité de la population étudiée.

Le classeur Excel associé à la méthode contient une feuille de calcul qui permet de produire un ensemble de taux de mortalité lissés par un système relationnel logit, qui ajuste les taux corrigés de mortalité calculés selon la méthode décrite ci-dessus. L'utilisateur a le choix entre un standard de la famille du modèle Général des tables-type des Nations Unies, ou d'une des quatre familles des modèles régionaux des tables-type de Princeton. On peut aussi utiliser arbitrairement une autre table choisie comme standard lorsqu'on a de bonnes raisons de penser qu'elle ressemble plus au schéma par âge de la mortalité de la population étudiée.

Afin de réaliser le lissage, les quotients de mortalité des personnes entre les âges x et $x + 5$, notés ${}_5q_x$, sont calculés à partir des taux de mortalité corrigés, selon la formule :

$${}_5q_x = \frac{5{}_5m_x}{1 + 2,5{}_5m_x}.$$

Ces quotients permettent de calculer les survivants à chaque âge dans la table de mortalité, qui commence avec une racine $l_5 = 1$, en utilisant de proche en proche la formule suivante :

$$l_{x+5} = l_x(1 - {}_5q_x).$$

Les coefficients α et β du modèle relationnel logit sont déterminés par la régression linéaire suivante :

$$\gamma_x = \alpha + \beta\gamma_x^s$$

où

$$\gamma_x = 0,5 \ln \left(\frac{1 - l_x}{l_x} \right)$$

et l'exposant s désigne les valeurs de la table type.

On peut alors en déduire la table de mortalité, corrigée et lissée, à partir de la table-type choisie et des coefficients α and β comme suit :

$$\gamma_x^{fitted} = \alpha + \beta\gamma_x^s$$

et

$$l_x^{fitted} = \frac{1}{\exp(1 + 2\gamma_x^{fitted})}.$$

La série des survivants permet de calculer les taux de mortalité selon les formules suivantes :

$${}_5m_x^{fitted} = \frac{l_x^{fitted} - l_{x+5}^{fitted}}{T_x - T_{x+5}}$$

et

$${}_5m_x^{fitted} = \frac{l_x^{fitted}}{T_x}$$

où

$$T_x = \sum_{x=x,5}^{\omega} \frac{5}{2} (l_x^{fitted} + l_{x+5}^{fitted})$$

c'est-à-dire :

$$T_x = T_{x+5} + \frac{5}{2} (l_x^{fitted} + l_{x+5}^{fitted})$$

et ω est l'âge au-delà duquel il ne reste plus de survivant dans la table de mortalité.

On calcule alors l'espérance de vie comme suit :

$$e_x = \frac{T_x}{l_x}.$$

Dans le cas où l'on souhaite estimer l'espérance de vie aux âges élevés par itération, ces valeurs peuvent être utilisées pour ré-estimer la complétude.

EXEMPLE

Cet exemple utilise les données de la population féminine du Salvador au recensement de 1961, et les décès de l'état civil de l'année 1961. Le classeur Excel « PrestonCoale_El Salvador » (voir le site web) est basé sur ces données. La date de référence du recensement de 1961 est la nuit du 5 au 6 mai, et c'est ainsi que la date du recensement est entrée (06/05/1961) dans la feuille de calcul intitulée « *Introduction* ».

Étape 1 : Choix du taux de croissance initial

Le taux de croissance estimé à partir de la population âgée de 10 ans et plus aux recensements de 1950, 1961 et 1971 est de 2,8 %, et c'est cette valeur qui figure dans le *Manuel X*. Le taux de croissance calculé en appliquant la méthode de Brass de la balance de l'accroissement démographique à ces données est de 3,1 %, qui est très proche de l'estimation que l'on peut faire à partir de la population moyenne en 1955 et 1965 qui figure dans International Data Base of the US Census Bureau, et qui se calcule comme suit :

$$r = \frac{\ln(3017852 / 2221139)}{1965 - 1955} = 0,03065.$$

Étape 2 : Estimer l'espérance de vie à l'âge A et pour les intervalles d'âges de 5 ans en 5 ans, jusqu'à 65 ans

Ces estimations proviennent des données, après leur avoir appliqué la méthode de Brass de la balance de l'accroissement démographique, comme cela est indiqué dans la colonne 2 du tableau 21.1.

Les rapports du nombre de décès déclarés dans le groupe d'âge 10-39 ans, âge au dernier anniversaire (1 706 décès), au nombre de décès dans le groupe d'âge 40-59 ans, (1 467 décès) est de : $\frac{1706}{1467} = 1,16$. Les espérances de vie par âge qui correspondent à ce niveau de mortalité dans la table-type du modèle Ouest (que l'on trouve dans la feuille du calcul *Espérances de vie* du classeur Excel), se calculent par interpolation entre les deux niveaux encadrant, comme indiqué dans la colonne 3 du tableau 21.1. Par exemple, pour l'âge de 65 ans :

$$e_{65} = 9,50 + \frac{1,171 - 1,16}{1,171 - 1,115}(9,86 - 9,50) = 9,552 .$$

L'itération pour trouver l'espérance de vie à partir des estimations de la table-type du model Ouest donne une estimation du taux de croissance (comme cela est expliqué plus en détail ci-dessous) de 3,065 %, et l'estimation finale de l'espérance de vie apparaît dans la colonne 4 du tableau 21.1.

Comme le VIH/sida n'existait pas au Salvador en 1961, on peut utiliser les estimations dérivées de la table-type du modèle Ouest de Princeton, qui apparaissent dans la feuille de calcul intitulée *Espérances de vie* du classeur Excel pour calculer la complétude de la déclaration des décès. Cependant, pour mieux illustrer la méthode, le classeur Excel utilise la méthode itérative, bien que

la comparaison des estimations qui figurent dans le tableau 21.1 (et celle des valeurs observées des taux de mortalité par âge dans l'intervalle ouvert 75 ans et plus avec les taux lissés) suggère plutôt que l'âge soit exagéré, ou que la complétude baisse au-delà de 75 ans. Dans un cas comme dans l'autre, ceci conduirait à une surestimation de la complétude.

Étape 3 : Estimer le nombre de personnes qui atteignent l'âge x, et ceux qui atteignent le groupe d'âge x à x + 4, d'après les décès déclarés

Le nombre de personnes qui atteignent l'âge x au cours de la période de référence (celle où sont déclarés les décès) se calcule à partir du nombre de décès dans chaque groupe d'âge, en utilisant le groupe d'âge ouvert de 75 ans et plus, le taux de croissance de 3,065 %, et l'estimation de l'espérance de vie qui figure dans la quatrième colonne du tableau 21.1. Les résultats figurent en colonne 4 du tableau 21.2. Par exemple, l'estimation du nombre de personnes qui atteignent 70 ans au cours de la période de référence s'obtient comme suit :

$$\begin{aligned}\widehat{N}_{75} &= 1360(\exp(0,03065 \times 7,76) - (0,03065 \times 7,76)^2 / 6) \\ &= 1712,11 \\ \widehat{N}_{70} &= 1712 \exp(5 \times 0,03065) + 504 \exp(2,5 \times 0,03065) \\ &= 2539,76.\end{aligned}$$

Le nombre de personnes qui atteignent le groupe d'âge x à x + 4 au cours de la période de référence est indiqué en colonne 5 du tableau 21.2. Par exemple, le nombre de personnes qui atteignent le groupe d'âge 20-24 ans se calcule comme suit :

$${}_5\widehat{N}_{20} = 2,5(21542 + 18212) = 99382,90.$$

TABLEAU 21.1 *Espérances de vie par âge selon la méthode utilisée, Salvador, recensement de 1961*

| Âge x | Méthode de Brass de la balance de la croissance démographique | Modèle Ouest de Princeton | Itérations successives |
|-------|---|---------------------------|------------------------|
| 65 | 13,4 | 9,55 | 13,1 |
| 70 | 10,4 | 7,38 | 10,2 |
| 75 | 7,9 | 5,57 | 7,8 |
| 80 | 5,9 | 4,06 | 5,8 |
| 85 | 4,4 | 2,88 | 4,3 |

Étape 4 : Estimer le nombre de personnes qui atteignent le groupe d'âge x à x + 4 au cours de la période de référence, d'après le recensement de population

Dans ce cas particulier, comme les décès sont enregistrés au cours d'une seule année, le nombre de personnes qui atteignent le groupe d'âge x à x + 4 au cours de la période de référence est simplement égal au nombre de personnes dans le groupe d'âge (c'est-à-dire que les effectifs figurant en colonne 6 sont les mêmes que ceux qui figurent en colonne 2 du tableau 21.2), car les multiplier par 1 ne les change pas.

Étape 5 : Calculer les rapports des estimations faites à partir des décès déclarés aux estimations faites à partir du recensement de population

Les rapports des effectifs de personnes atteignant le groupe d'âge x à $x + 4$ au cours de la période de référence, calculés en rapportant ceux obtenus d'après les décès déclarés sur ceux obtenus d'après le recensement, apparaissent en colonnes 7 et 8 du tableau 21.2. Par exemple, le calcul pour les groupes d'âge 65-69 ans et 65-74 ans donne :

$$\frac{{}_5\widehat{N}_{65}}{{}_5N_{65}} = \frac{14962}{14964} = 0,9999$$

$$\frac{{}_{10}\widehat{N}_{65}}{{}_{10}N_{65}} = \frac{14962 + 10630}{14964 + 11205} = 0,9779 .$$

Étape 6 : Estimer la complétude de la déclaration des décès

Dans cet exemple, l'estimation du taux de croissance fournie par l'application de la méthode de Brass de la balance de l'accroissement démographique produit une série cohérente et satisfaisante des rapports. Cependant, nous proposons ici,

à titre illustratif, de procéder à une itération pour trouver une nouvelle estimation du taux de croissance et de l'espérance de vie. Cette procédure permet de mettre dans un graphique les points correspondants aux rapports, comme le montre la figure 21.1.

Comme on n'observe aucune tendance nette, ni vers le haut ni vers le bas, dans cette figure 21.1, on peut déterminer le taux de croissance à partir des groupes d'âge de 5 à 74 ans, en minimisant les écarts à la moyenne à l'aide du module *Solveur*. La complétude se calcule à partir des rapports calculés pour les âges de 15 à 64 ans, afin d'éviter les fluctuations dues aux âges élevés (quoique cette méthode de calcul basée sur médiane et quartiles soit assez robuste aux fluctuations dans les groupes d'âge particuliers). Ce calcul donne une estimation de la complétude de 88 %, et se fait comme suit :

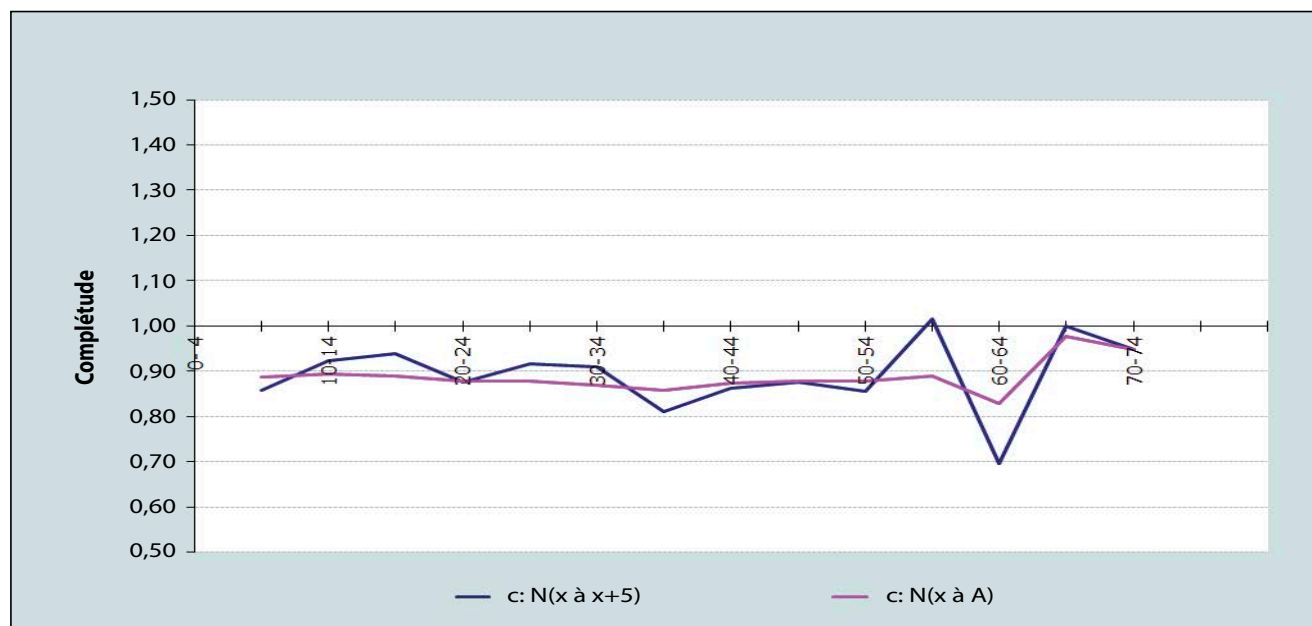
$$c = (0,5 \times 0,8764 + 0,25(0,8575 + 0,9144)) \exp(0,03065(1961,34 - 1961,50))$$

$$= 0,8812 \times 0,9951 = 87,7\% .$$

TABLEAU 21.2 Calcul du nombre de personnes atteignant le groupe d'âge x à $x + 4$ d'après les décès déclarés et le recensement, et rapports des estimations, Salvador, recensement de 1961

| Âge | ${}_5N_x(t_c)$ | ${}_5D_x$ | Est N_x | Est ${}_5N_x$ | ${}_5N_x$ obs | $c : {}_5N_x$ | $c : A-xN_x$ |
|-------|----------------|-----------|-----------|---------------|---------------|---------------|--------------|
| 0-4 | 214 089 | 6 909 | | | 214 089 | | |
| 5-9 | 190 234 | 610 | 35 431 | 163 158 | 190 234 | 0,8577 | 0,8879 |
| 10-14 | 149 538 | 214 | 29 832 | 138 071 | 149 538 | 0,9233 | 0,8946 |
| 15-19 | 125 040 | 266 | 25 396 | 117 344 | 125 040 | 0,9384 | 0,8885 |
| 20-24 | 113 490 | 291 | 21 542 | 99 383 | 113 490 | 0,8757 | 0,8778 |
| 25-29 | 91 663 | 271 | 18 212 | 83 962 | 91 663 | 0,9160 | 0,8783 |
| 30-34 | 77 711 | 315 | 15 373 | 70 677 | 77 711 | 0,9095 | 0,8690 |
| 35-39 | 72 936 | 349 | 12 897 | 59 098 | 72 936 | 0,8103 | 0,8584 |
| 40-44 | 56 942 | 338 | 10 742 | 49 112 | 56 942 | 0,8625 | 0,8741 |
| 45-49 | 46 205 | 357 | 8 903 | 40 525 | 46 205 | 0,8771 | 0,8781 |
| 50-54 | 38 616 | 385 | 7 307 | 33 049 | 38 616 | 0,8558 | 0,8785 |
| 55-59 | 26 154 | 387 | 5 913 | 26 567 | 26 154 | 1,0158 | 0,8892 |
| 60-64 | 29 273 | 647 | 4 714 | 20 398 | 29 273 | 0,6968 | 0,8295 |
| 65-69 | 14 964 | 449 | 3 445 | 14 962 | 14 964 | 0,9999 | 0,9779 |
| 70-74 | 11 205 | 504 | 2 540 | 10 630 | 11 205 | 0,9487 | 0,9487 |
| 75+ | 16 193 | 1 360 | | | | | |

FIGURE 21.1 Estimations de la complétude de l'enregistrement des décès, selon l'âge, Salvador, recensement de 1961



Étape 7 : Estimer les taux de mortalité après correction pour tenir compte du sous-enregistrement des décès

La population moyenne, au milieu de la période de référence, se calcule à partir du recensement de population et du taux de croissance estimé entre les points extrêmes, soit 3,1 %. Ces calculs figurent en colonne 2 du tableau 21.3. Par exemple, pour le groupe d'âge 15-19 ans, l'estimation se fait comme suit :

$${}_5N_{15}(t_m) = 125040 \times \exp(-0,03065(1961,34 - 1961,50)) \\ = 125040e^{0,03065 \times 0,16} = 125655,$$

Puis les décès sont corrigés pour tenir compte du sous-enregistrement, en divisant le nombre de décès déclarés dans chaque groupe d'âge par l'estimation de la complétude. Ces nombres apparaissent en colonne 3 du tableau 21.3. Par exemple, pour le groupe d'âge 15-19 ans, le nombre se calcule à partir du nombre de décès déclarés (qui figure en colonne 3 du tableau 21.1, soit 266), comme suit :

$$\frac{266}{0,8768} = 303,4.$$

Les personnes-années vécues soumises au risque (colonne 4 du tableau 21.3) se calculent comme le produit de la population moyenne, c'est-à-dire celle observée en milieu de la période de référence (colonne 2 du tableau 21.3) par la durée (exprimée en années) de la période de référence (au cours de laquelle les décès sont déclarés), qui dans ce cas est égale à 1 an.

Les taux de mortalité corrigés pour tenir compte du sous-enregistrement des décès (colonne 5 du tableau 21.3) se calculent en divisant les décès corrigés par les personnes-années vécues. Par exemple, pour le groupe d'âge 15-19 ans, le taux de mortalité corrigé se calcule comme suit :

$$\frac{303,4}{125656} = 0,00241.$$

Étape 8 : Lissage par un modèle relationnel logit et une table-type de mortalité

Les estimations des quotients quinquennaux de mortalité des femmes, ${}_5q_x$, se calculent à partir des taux de mortalité selon la formule classique, comme cela apparaît en colonne 2 du tableau 21.4. Par exemple, le quotient de 15 à 19 ans se calcule comme suit :

TABLEAU 21.3 Calcul des taux de mortalité corrigés, Salvador, recensement de 1961

| Âge | Valeurs corrigées | | | |
|-------|------------------------------|--------------------|------------------------------------|-----------------------------------|
| | Population ${}_5N_x(t_m)$ | Décès ${}_5D_x$ | Personnes- années $PYL(x,5)$ | Taux de mortalité ${}_5m_x$ |
| 0-4 | | | | |
| 5-9 | 191 171 | 696 | 191 171 | 0,0036 |
| 10-14 | 150 274 | 244 | 150 274 | 0,0016 |
| 15-19 | 125 656 | 303 | 125 656 | 0,0024 |
| 20-24 | 114 049 | 332 | 114 049 | 0,0029 |
| 25-29 | 92 114 | 309 | 92 114 | 0,0033 |
| 30-34 | 78 094 | 359 | 78 094 | 0,0046 |
| 35-39 | 73 295 | 398 | 73 295 | 0,0054 |
| 40-44 | 57 222 | 385 | 57 222 | 0,0067 |
| 45-49 | 46 433 | 407 | 46 433 | 0,0087 |
| 50-54 | 38 806 | 439 | 38 806 | 0,0112 |
| 55-59 | 26 283 | 441 | 26 283 | 0,0166 |
| 60-64 | 29 417 | 738 | 29 417 | 0,0248 |
| 65-69 | 15 038 | 512 | 15 038 | 0,0337 |
| 70-74 | 11 260 | 572 | 11 260 | 0,0505 |
| 75+ | 16 273 | 1 551 | 16 273 | 0,0944 |

$${}_5q_{15} = \frac{5 \times 0,00241}{1 + 2,5 \times 0,00241} = 0,0120.$$

Les survivants de la table de la mortalité à l'âge $x + 5$ se calculent de proche en proche à partir des survivants à l'âge x , selon la formule classique, comme cela apparaît en colonne 3 du tableau 21.4. Par exemple, les survivants à 20 ans se calculent comme suit :

$$l_{20} = 0,9740(1 - 0,01200) = 0,9623.$$

La transformation logit des probabilités de survie figure en colonne 4 du tableau 21.4. Par exemple à 20 ans, la transformation logit de l_{20} se calcule comme suit :

$$\gamma_{20} = 0,5 \ln \left(\frac{1 - 0,9623}{0,9623} \right) = -1,6204.$$

La transformation de la table de mortalité conditionnelle pour les femmes, calculée à partir de la famille Ouest des tables-type de Princeton au niveau d'espérance de

TABLEAU 21.4 Calcul des taux de mortalité lissés à l'aide du modèle relationnel logit et d'une table-type de mortalité, Salvador, recensement de 1961

| Âge | Données observées | | | Lissage | | | Table de mortalité | | | |
|-----|-------------------|-----------|-------------|--------------------------------|-------------------------------|--------------------------|----------------------------|--------|--------|----------------------------|
| | ${}_5q_x$ | l_x/l_5 | Obs. $Y(x)$ | Princeton Ouest $l^s(x)$ | Logit standard $Y^s(x)$ | Logit lissé $Y(x)$ | Surviv. lissé $l(x)$ | $T(x)$ | $e(x)$ | Taux lissé ${}_5m_x$ |
| 0 | | | | | | | | | | |
| 5 | 0,0180 | 1 | | 1,0000 | | | 1 | 61,254 | 61,3 | 0,0025 |
| 10 | 0,0081 | 0,9820 | - 1,9987 | 0,9890 | - 2,2506 | - 2,1834 | 0,9875 | 56,285 | 57,0 | 0,0019 |
| 15 | 0,0120 | 0,9740 | - 1,8122 | 0,9805 | - 1,9585 | - 1,8980 | 0,9780 | 51,371 | 52,5 | 0,0028 |
| 20 | 0,0144 | 0,9623 | - 1,6204 | 0,9681 | - 1,7060 | - 1,6511 | 0,9645 | 46,515 | 48,2 | 0,0036 |
| 25 | 0,0166 | 0,9484 | - 1,4560 | 0,9519 | - 1,4928 | - 1,4428 | 0,9471 | 41,736 | 44,1 | 0,0041 |
| 30 | 0,0227 | 0,9327 | - 1,3141 | 0,9337 | - 1,3226 | - 1,2765 | 0,9278 | 37,049 | 39,9 | 0,0047 |
| 35 | 0,0268 | 0,9115 | - 1,1658 | 0,9132 | - 1,1766 | - 1,1337 | 0,9061 | 32,464 | 35,8 | 0,0054 |
| 40 | 0,0331 | 0,8870 | - 1,0304 | 0,8899 | - 1,0447 | - 1,0048 | 0,8818 | 27,994 | 31,7 | 0,0065 |
| 45 | 0,0429 | 0,8577 | - 0,8980 | 0,8628 | - 0,9194 | - 0,8824 | 0,8538 | 23,655 | 27,7 | 0,0081 |
| 50 | 0,0550 | 0,8209 | - 0,7611 | 0,8299 | - 0,7925 | - 0,7583 | 0,8200 | 19,470 | 23,7 | 0,0111 |
| 55 | 0,0806 | 0,7757 | - 0,6204 | 0,7863 | - 0,6514 | - 0,6205 | 0,7757 | 15,481 | 20,0 | 0,0155 |
| 60 | 0,1180 | 0,7132 | - 0,4555 | 0,7289 | - 0,4946 | - 0,4672 | 0,7180 | 11,747 | 16,4 | 0,0235 |
| 65 | 0,1569 | 0,6290 | - 0,2640 | 0,6490 | - 0,3074 | - 0,2842 | 0,6384 | 8,356 | 13,1 | 0,0357 |
| 70 | 0,2263 | 0,5303 | - 0,0607 | 0,5427 | - 0,0856 | - 0,0674 | 0,5337 | 5,426 | 10,2 | 0,0570 |
| 75 | | 0,4103 | 0,1814 | 0,4062 | 0,1898 | 0,2018 | 0,4005 | 3,090 | 7,7 | 0,0903 |
| 80 | | | | 0,2545 | 0,5373 | 0,5414 | 0,2530 | 1,457 | 5,8 | 0,1405 |
| 85 | | | | 0,1201 | 0,9956 | 0,9893 | 0,1215 | 0,521 | 4,3 | 0,2115 |

vie $e_0 = 60$ ans (colonne 5 du tableau 21.4) apparaît en colonne 6 du même tableau. Comme on peut le voir sur la figure 21.2, le modèle Ouest semble ajuster correctement les données, sauf peut-être aux très jeunes âges.

Les coefficients du modèle relationnel logit, α et β , se calculent comme la pente et l'ordonnée à l'origine de la droite de régression liant les transformations logit, qui figurent en colonne 4 et 6 du tableau 21.4, sur un intervalle d'âge choisi par l'utilisateur (entre 45 et 75 ans dans cet exemple), soit $\alpha = 0,0162$ et $\beta = 0,9774$. On a choisi ici les âges de 45 à 75 ans, car l'ajustement aux âges élevés est important pour estimer l'espérance de vie au début de l'intervalle ouvert.

On applique ensuite ces coefficients à la transformation logit des probabilités de survie conditionnelle de la table-type afin de calculer les logits lissés, comme en colonne 7 du tableau 21.4. Par exemple, le logit lissé de la survie à l'âge de 20 ans se calcule comme suit :

$$\gamma_{20}^{fitted} = 0,0162 + 0,9774 \times (-1,7060) = -1,6511.$$

Ces valeurs servent ensuite à calculer les survivants de la table de mortalité lissée, qui se trouvent en colonne 8 du tableau 21.4. Par exemple, la probabilité de survie à 20 ans se calcule comme suit :

$$l_{20}^{fitted} = \frac{1}{\exp(1 + 2 \times (-1,6511))} = 0,9645.$$

Les années de vie cumulées au-delà de l'âge x , notées T_x , apparaissent en colonne 9 du tableau 21.4. Elles se calculent à partir des survivants lissés de la table de mortalité, selon la formule classique. Ces valeurs sont aussi utilisées pour produire les taux de mortalité lissés, qui apparaissent en colonne 11 du tableau 21.4. Par exemple, à l'âge de 80 ans :

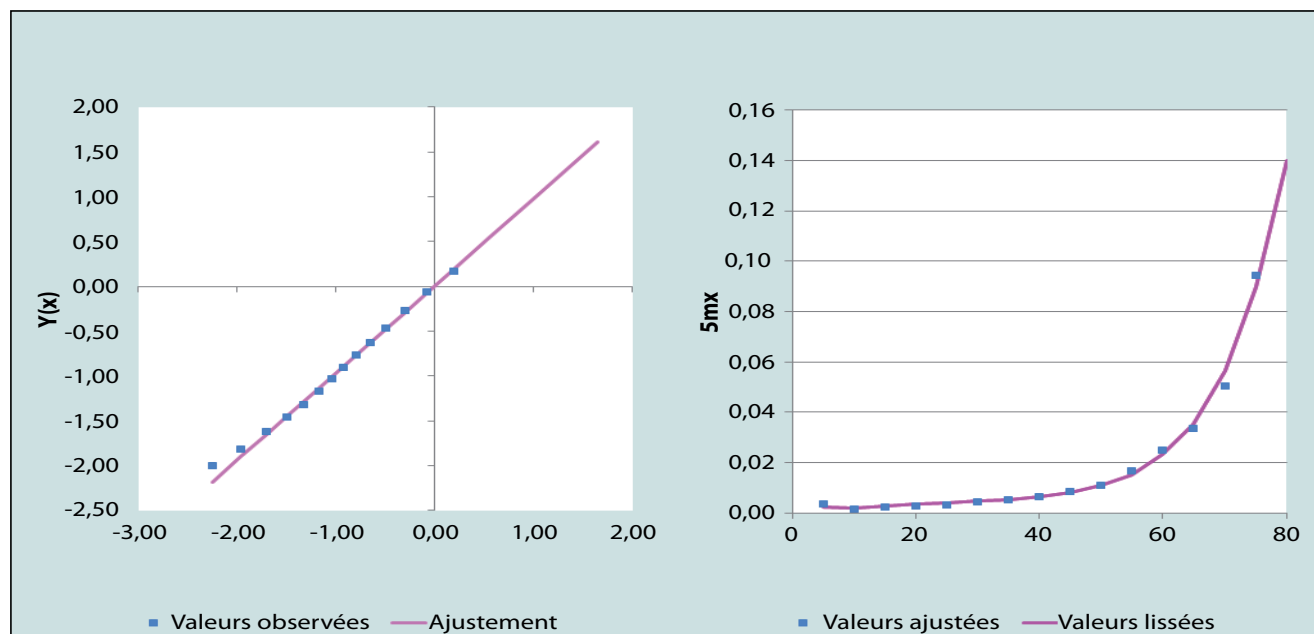
$$T_{80} = 0,521 + \frac{5}{2}(0,253 + 0,1215) = 1,457$$

$${}_5m_{80}^{fitted} = \frac{0,253 - 0,1215}{1,457 - 0,521} = 0,1405.$$

Et l'espérance de vie à l'âge x se calcule en divisant les nombres de la colonne 9 par ceux de la colonne 8, comme cela est montré en colonne 10 du tableau 21.4. Par exemple, l'espérance de vie à 65 ans se calcule comme :

$$\frac{8,356}{0,6384} = 13,089.$$

FIGURE 21.2 Logits des probabilités de survie, observés et lissés par la table-type du modèle Ouest de Princeton, Salvador, recensement de 1961



DIAGNOSTICS, ANALYSE ET INTERPRÉTATION

Contrôles et validation

L'exemple présenté ci-dessus était tiré du *Manuel X* (Division de la Population des NU 1984). Il donnait une estimation de la complétude d'environ 83 %, à la fois en appliquant cette méthode et la méthode de Brass de la balance d'accroissement démographique. La différence entre les deux estimations du *Manuel X* et celle produite par cette application (88 %) semble due essentiellement aux différences dans la méthode d'estimation de la population du groupe d'âge ouvert (A+). L'effet complet est contrebalancé, au moins dans une certaine limite, par une réduction de l'estimation (par rapport à l'estimation du *Manuel X*) due au fait que l'approche présentée ici trouve, par itération, un taux de croissance (3,02 %) qui est plus élevé que celui qui est utilisé dans le *Manuel X* (2,87 %). En conséquence, l'application de la méthode avec les estimations d'espérance de vie calculées à l'aide de la table-type du modèle Ouest de Princeton (colonne 3 du tableau 21.1) produit une estimation de la complétude de 85 %.

Interprétation

Du fait qu'il n'y ait pas de tendance visible (ni montante ni descendante) sur le graphique de la figure 21.1, il n'y a aucune raison d'abaisser l'âge de début de l'intervalle ouvert. Cependant, si cela avait été nécessaire, il aurait été difficile de déterminer quelle estimation de la complétude retenir, car celle-ci est de 85 % si on utilise l'intervalle ouvert de 70 ans et plus, mais de 76 % avec l'intervalle ouvert de 65 ans et plus. La feuille de calcul ne permet pas à cet âge de début de l'intervalle ouvert d'être inférieur à 65 ans. Mais si on avait utilisé un intervalle ouvert de 60 ans et plus, l'estimation de la complétude aurait été supérieure à 76 %. En règle générale, on recommande de ne pas s'arrêter à un âge finissant par zéro dans les populations qui font preuve d'une préférence marquée pour les chiffres ronds.

Dans l'ensemble, ces résultats indiquent que la complétude de la déclaration des décès est d'environ 85 %, un peu moins que les 92 % obtenus en appliquant la méthode de Brass de la balance de l'accroissement démographique avec les mêmes données. Il est intéressant de noter que si l'on avait utilisé une estimation de l'espérance de vie dans les tables-type du modèle Ouest de Princeton basées sur le rapport des décès à 10-39 ans et à 40-59 ans (${}_{30}D_{10} / {}_{20}D_{40}$) et un intervalle ouvert de 75 ans et plus, alors l'estimation de

la complétude serait tombée à 85 %, mais l'estimation de l'espérance de vie calculée à partir des taux lissés est plus proche de celle qui est obtenue par itération que de celles qui sont utilisées pour produire l'estimation de la complétude. Ceci suggère que la méthode n'est pas très sensible aux estimations de l'espérance de vie qui sont utilisées, surtout si l'intervalle ouvert commence à un âge élevé.

PROBLÈMES D'INTERPRÉTATION RELATIFS À LA MÉTHODE

Source des données sur les décès déclarés

En général, on trouve deux types de problèmes concernant les données sur les décès : la question de la complétude (inférieure ou supérieure à 1), ce qui est la base de la méthode ; et la question de la complétude différentielle selon l'âge, qui peut entraîner un biais dans les estimations. Quoique l'approche générale reste essentiellement la même quelle que soit la source des données sur les décès, les différentes sources disponibles ont chacune leurs propres biais, qui peuvent affecter l'interprétation des résultats. On illustrera ce point par différents exemples, mais en général il convient de prêter attention aux biais suivants pour ce qui concerne les données sur les décès.

1) Données sur les décès de l'état civil

Si la répartition proportionnelle de la population entre les zones urbaines et rurales (ou une autre catégorie similaire) diffère significativement selon l'âge, et que la complétude de la déclaration des décès dans les zones urbaines est significativement supérieure à celle des zones rurales, alors l'hypothèse que la complétude est indépendante de l'âge sera vraisemblablement violée, et l'on observera une baisse de la complétude selon l'âge au-delà de 50 ans si une proportion importante de personnes migrent de l'urbain vers le rural au moment de la retraite. Si on ignore ce biais, et si le taux de croissance est estimé à l'aide du module *Solveur*, ne pas respecter l'hypothèse conduira à sous-estimer le niveau moyen de complétude.

2) Données sur les décès déclarés dans les recensements et les enquêtes ménage

Ces données sont potentiellement sujettes à trois problèmes :

- Si une proportion importante de ménages sont dissous suite au décès d'une personne clé (par exemple la personne qui assure les ressources du ménage), alors les décès

de ces personnes risquent de n'être pas déclarés, ce qui induit une violation de l'hypothèse d'indépendance de la complétude avec l'âge. Si une proportion importante de décès de certains groupes d'âge se rapportent à des personnes qui ne vivent pas dans des ménages ordinaires (par exemple ceux qui vivent dans des maisons de retraite), la violation de l'hypothèse pourrait être encore plus forte. Cependant, ce problème ne se pose pas dans la plupart des pays en développement.

- Dans les cas où les jeunes adultes quittent le domicile familial quand ils grandissent, pour aller travailler en milieu urbain, il est possible qu'ils soient considérés comme membres de plusieurs ménages (ou bien d'aucun ménage), et que leurs décès soient déclarés plusieurs fois (ou jamais), ce qui conduit là encore à une violation de l'hypothèse de complétude de la déclaration uniforme selon l'âge. Dans ce cas, on pourrait en limiter l'impact en ignorant les données se rapportant à un groupe d'âge spécifique pour calculer la complétude.
- Erreurs sur la période de référence : lorsqu'il y a une confusion fréquente sur la période précise au cours de laquelle les décès doivent être déclarés, en plus des erreurs sur la date précise du décès, il peut arriver que l'on sous-estime, ou surestime, les décès. Si l'on peut supposer que ces biais sont indépendants de l'âge de la personne décédée, cette distorsion sera prise en compte dans l'estimation de la complétude, et ne posera pas de problème pour l'estimation finale des taux de mortalité corrigés.

3) Décès enregistrés dans les structures sanitaires

On ne sait que peu de choses sur la manière dont ces sources de données fonctionnent. Cependant, on peut supposer que la complétude dépendra de la répartition des structures sanitaires dans lesquelles ces données seront recueillies, et dans de nombreux pays en développement ce type de service sera vraisemblablement concentré en milieu urbain. Dans ce cas encore, si la proportion de la population vivant en milieu urbain, plutôt qu'en milieu rural, varie avec l'âge, alors on ne pourra pas faire l'hypothèse que la complétude est indépendante de l'âge. De plus, il est possible que certaines causes de décès soient plus fréquentes dans les structures sanitaires, et si ces causes sont importantes quantitativement et corrélées avec l'âge, on pourrait avoir une nouvelle violation de l'hypothèse de la complétude constante selon l'âge.

Dans tous les cas mentionnés ci-dessus, on devrait éviter d'ajuster le taux de croissance pour obtenir une séquence

de rapports bien horizontale. Au contraire, il conviendra plutôt de s'assurer que l'estimation de c est calculée sur un intervalle d'âge qui exclut ceux pour lesquels la complétude de la déclaration des décès est exceptionnellement forte ou exceptionnellement faible.

Interprétation du diagnostic général

En pratique, les séquences des deux rapports : ${}_5\widehat{N}_x / {}_5N_x$ et ${}_A\widehat{N}_x / {}_AN_x$ sont affectées par les violations des hypothèses qui sous-tendent la méthode. Toutefois, la puissance de cette technique vient en partie de ce que la plupart des violations courantes des hypothèses vont produire des écarts caractéristiques et distincts par rapport à la droite horizontale qui est attendue, et dans certaines circonstances ces écarts pourront être interprétés. Voici quelques exemples :

a) *taux de croissance erroné* : Si r est trop élevé, la séquence des points tend presque linéairement avec l'âge vers la valeur sous-jacente de la complétude, et vice versa, comme le montre l'analyse de l'équation (1) ci-dessous. Cet effet sera plus fort pour les rapports dans les groupes d'âge quinquennaux, ${}_5\widehat{N}_x / {}_5N_x$ que pour les âges cumulés ${}_A\widehat{N}_x / {}_AN_x$.

b) *Exagération de l'âge déclaré* : Il est courant que les parents qui déclarent le décès tendent à exagérer l'âge au décès du défunt, plus que les vivants ne le font lorsqu'ils déclarent leur propre âge au recensement. Ce biais produit une séquence de points dont la pente tend à augmenter, surtout après l'âge auquel cette exagération commence, et la séquence au-delà de ce point est marquée par une forte courbure vers le haut. Ce schéma peut se comprendre en examinant l'équation (1) ci-dessous : l'exagération de l'âge conduit non seulement à une augmentation du nombre de décès dans le groupe d'âge suivant, mais en plus le transfert d'un groupe à l'autre conduit à ce que ces décès soient multipliés par un coefficient exponentiel plus fort, quoique ce dernier biais soit de moindre importance. Un tel biais pourrait aussi être produit par une augmentation de la complétude de l'enregistrement des décès avec l'âge, au-delà d'un certain âge. Toutefois, en pratique ceci n'a pas été observé (Preston, Coale, Trussell *et al.* 1980).

c) *Erreurs sur l'âge dans les estimations de la population, et erreurs de décompte variant selon l'âge* : Ce type de biais se manifeste par une série erratique des rapports tout au long des âges de la vie. Comme la quantité ${}_5\widehat{N}_x$ est de forme cumulative, elle tend à suivre d'assez près la distribution par âge de la population. Donc, si l'on observe des zigzags, il est vraisemblable que les pics soient associés à des âges faisant

l'objet d'aversion ou de sous-enregistrement dans la population, et que les baisses soient associées à une préférence pour certains âge ou à un sur-enregistrement de la population. Si ces fluctuations sont indépendantes de l'âge, elles ne devraient pas particulièrement affecter les estimations de la complétude. John Blacker (Blacker 1988) a proposé d'utiliser d'autres groupes d'âge (comme 18-22, 23-27, etc.) pour éliminer ces zigzags, et a montré que ceci permettait de réduire les biais dans l'estimation de la pente lors de l'application de la méthode de Brass de la balance de l'accroissement démographique. Cependant, lorsque ces distorsions sont systématiques, par exemple produites par des migrations en-deçà d'un certain âge, il peut être préférable d'exclure ces points pour l'estimation de la complétude.

On peut, en général, éliminer l'essentiel de l'effet de la surestimation de l'âge en commençant l'intervalle ouvert à un âge suffisamment jeune pour y confiner l'essentiel de la surestimation.

Afin de distinguer les séquences des rapports ayant une pente déclinante et qui sont dues à la baisse de la mortalité de celles qui sont dues au choix d'un taux de croissance trop élevé, on doit considérer d'autres sources de données pour déterminer quelle est l'explication la plus plausible. Si la mortalité de la population tend à diminuer, la médiane des rapports des effectifs cumulés, par exemple de 10 à 45 ans, devrait quand même permettre de fournir une estimation raisonnable de la complétude de l'enregistrement des décès. Même si cette méthode a de sérieux atouts et qu'elle est plus robuste aux écarts à la stabilité que ne l'est la méthode de Brass de la balance de l'accroissement démographique, elle reste plus sensible que cette dernière à certains types d'erreurs sur l'âge. Il ne sera alors pas possible d'obtenir une seule estimation robuste de la complétude des données sur les décès, sauf si on peut confirmer par d'autres moyens que les hypothèses se vérifient, surtout l'hypothèse de la croissance démographique.

DESCRIPTION DÉTAILLÉE DE MÉTHODE

Exposé mathématique

La méthode de Preston et Coale est un cas particulier de la méthode de l'extinction des cohortes synthétiques, (chapitre 25) avec un taux de croissance $r(x+)$ de la population au-delà de l'âge x identique pour tous les âges. L'origine de la méthode remonte aux travaux pionniers de Paul Vincent (Vincent 1951) sur la méthode de l'extinction des cohortes,

et aux travaux de Samuel Preston et Kenneth Hill (Preston et Hill 1980), qui fut améliorée par la suite (Preston, Coale, Trussell *et al.* 1980). Elle repose sur l'idée simple que le nombre de personnes d'un certain âge, survivantes à un certain moment, doit être égal au nombre de décès qui se produiront dans cette cohorte, jusqu'au moment où le dernier survivant sera décédé. Dans une population stable et fermée aux migrations, cette relation s'écrit

$$N_x = \sum_{a=x}^{\infty} D_a e^{r(a-x)} \quad (1)$$

où D_a sont les décès d'âge a pris au même instant que la population N_x , car dans une population stable et fermée, $D_{a,t}$, le nombre de décès d'âge a qui vont se produire dans t années après le moment où on réalise l'enregistrement des décès sera égal à $D_a e^{rt}$.

Si, au lieu de D_a on connaît, D_a^r le nombre de décès enregistrés d'âge x au dernier anniversaire, et si on estime la population d'âge x , c'est-à-dire \widehat{N}_x , par la quantité :

$\widehat{N}_x = \sum_{a=x}^{\infty} D_a^r e^{r(a-x)}$, alors le rapport \widehat{N}_x / N_x (où N_x est la taille de la population au milieu de l'intervalle au cours duquel on enregistre les décès) fournira une indication sur le pourcentage de décès au-delà de l'âge x qui sont enregistrés, soit c_{x+} . Si, de plus, la population d'âge x , soit N_x , est disponible à deux points dans le temps, on peut affiner l'estimation en tenant compte du taux de croissance r au cours de la période. Cependant, si le niveau de complétude est estimé pour calculer les taux de mortalité, alors la même correction s'appliquera de fait au numérateur comme au dénominateur, si bien qu'elle pourra être ignorée.

En pratique, on rencontre un problème pour calculer \widehat{N}_x car les décès D_a^r ne sont pas toujours disponibles au-delà d'un certain âge (et même s'ils le sont, ils ont peu de chance d'être d'âge précis), c'est pourquoi on les regroupe au-delà de cet âge dans le groupe d'âge ouvert, soit D_{A+}^r où A désigne le début de cet intervalle ouvert. Plusieurs méthodes ont été proposées pour résoudre ce problème. Par exemple, dans le *Manuel X* (Division de la Population des NU 1984 : 136), on propose de faire l'hypothèse que le schéma de mortalité suit celui d'une des tables-types de mortalité de Princeton révisées (Coale, Demeny et Vaughan 1983). Dans ce cas, \widehat{N}_A se calcule comme suit

$$\widehat{N}_A = D_{A+}^r e^{rz(A)}$$

où

$$z(A) = a(A) + b(A) + c(A)e^{[D_{45+}/D_{10+}]}$$

Les coefficients correspondants ont été tabulés (voir le tableau 123 dans : Division de la Population des NU. 1984 : 137). On estime alors le rapport D_{45+} / D_{10+} par le rapport D_{45+}^r / D_{10+}^r .

Il existe une alternative (Bennett et Horiuchi 1984), qui recommande d'estimer la population d'âge A par la formule suivante :

$$\widehat{N}_A = {}_{\infty}D_A \left(\exp(r \times e_A) - (r \times e_A)^2 / 6 \right)$$

où l'espérance de vie à l'âge A (e_A) est interpolée au sein de la famille Ouest des tables-type de Princeton, à partir du rapport des décès déclarés entre les âges 10 et 40 ans et les âges 40 et 60 ans.

Puisque \widehat{N}_x peut être approché par :

$$N_{x+5}e^{5r} + {}_5D_x^r e^{2,5r}$$

lorsque \widehat{N}_A a été estimé, on peut aussi estimer les \widehat{N}_x à partir des décès ${}_5D_x^r$.

Limites de la méthode

Les principales limites de la méthode ont été décrites ci-dessus, et sont prises en compte dans les feuilles de calcul : il faut que la population soit stable et fermée aux migrations. La méthode ne devrait pas être utilisée lorsque ces conditions ne sont pas vérifiées et que l'on s'en écarte trop. Un exemple d'utilisation abusive de la méthode est donné dans le classeur Excel intitulé « SEG_South Africa_males » (voir le site web). L'exemple provient d'Afrique du Sud, pour les décès déclarés entre le recensement de 2001 et l'enquête de communauté de 2007, un substitut au recensement. En prenant pour population au milieu de la période la moyenne des deux populations, on trouve une estimation de la complétude de 84 % avec le même intervalle d'âge. Si on augmente à 35 ans l'âge minimal pour l'intervalle d'âge utilisé pour ajuster la ligne droite, on trouve une estimation de 86 %, ce qui est encore inférieur à l'estimation de 94 % que l'on trouve avec la Méthode de l'extinction des cohortes synthétiques.

Cette méthode est plus vulnérable aux erreurs sur l'âge que la méthode de Brass de la balance de l'accroissement démographique. En particulier, comme cela est mentionné ci-dessus, la tendance fréquente à exagérer les âges au décès déclarés (par rapport aux âges enregistrés au recensement) se manifestera par des points sur le graphique qui suivent une

pente positive au-delà de l'âge à partir duquel l'exagération est forte. Dans de telles situations, il est préférable d'utiliser le taux de croissance fourni par la méthode de Brass de la balance de l'accroissement démographique. De plus, comme cela a été montré ci-dessus, la méthode est aussi sensible au choix du groupe d'âge ouvert lorsque les données sont affectées par une forte préférence pour les âges ronds, ce qui est très souvent le cas avec les données de recensement.

Par contre, la méthode est moins vulnérable aux effets de déstabilisation résultants d'un changement rapide de la mortalité au cours du temps (Martin 1980). De plus, des simulations ont montré pour la méthode de Brass de la balance de l'accroissement démographique que les biais résultant d'une baisse lente et régulière de la mortalité (comme le connaissent certains pays en développement en l'absence d'épidémies, de famines et de guerres) sont assez faibles (Rashad 1978).

En ce qui concerne les variations de la fécondité, celles-ci tendent à n'avoir qu'un faible impact sur les performances de la méthode, car elles n'affectent guère que les groupes d'âge les plus jeunes, ce qui n'a qu'une influence modérée sur les estimations de la complétude. Si nécessaire, ces groupes d'âge peuvent d'ailleurs être exclus pour calculer les taux de croissance et estimer la complétude.

Les migrations peuvent affecter la population des jeunes adultes (surtout entre 20 et 35 ans), mais elles ont moins d'effet sur les décès, qui sont concentrés aux âges élevés. Ignorer une immigration nette tendra à faire baisser la pente, et donc à surestimer la complétude de l'enregistrement des décès, ce qui se traduit par une sous-estimation des taux de mortalité. Ignorer une émigration nette aura l'effet inverse. Certains démographes conseillent d'ajuster une ligne droite jusqu'à l'âge de 5 ans pour limiter l'effet des migrations inconnues, en supposant que les différences entre la complétude de l'enregistrement des décès à ces jeunes âges et celles qui prévaut aux âges élevés a peu de chance d'induire de fortes distorsions, car la mortalité reste très faible entre 5 et 14 ans. Cependant, on peut douter que cette adaptation puisse réduire le biais de manière importante.

Une alternative consiste à confiner l'ajustement des points au-delà de 35 ans, de manière à éliminer l'essentiel de l'effet perturbateur des migrations. Mais souvent les données concernant les âges élevés sont de qualité plus médiocre, ce qui rend l'estimation de la complétude moins fiable. Il est possible que ces adaptations puissent produire de meilleures estimations que celles obtenues en ignorant carrément les migrations, mais malheureusement, très peu de recherches

ont été conduites sur la précision des estimations de la complétude que produisent ces adaptations.

Techniquement parlant, si on dispose d'estimations fiables des migrations nettes par âge, on pourrait adapter la méthode en remplaçant le taux de croissance r par $r - {}_5i_x$, où ${}_5i_x$ désigne le taux net de migration pour le groupe d'âge x à $x + 4$, l'âge étant pris au dernier anniversaire, pour calculer \widehat{N}_x . En pratique, cependant, on ne dispose que rarement de ces taux de migrations nettes par âge dans les situations où l'on doit utiliser cette méthode, ce qui ne pousse pas à recommander cette adaptation.

AUTRES LECTURES ET RÉFÉRENCES

Puisque cette méthode est un cas particulier de la méthode plus générale intitulée méthode de l'extinction des cohortes synthétiques, on renverra le lecteur au chapitre 25 pour en savoir plus.

- Bennett NG and S Horiuchi. 1984. "Mortality estimation from registered deaths in less developed countries", *Demography* **21**(2):217-233. doi : <http://dx.doi.org/10.2307/2061041>
- Blacker J. 1988. *An Evaluation of the Pakistan Demographic Survey*. Karachi : Pakistan Federal Bureau of Statistics.
- Coale AJ, P Demeny and B Vaughan. 1983. *Regional Model Life Tables and Stable Populations*. New York : Academic Press.
- Division de la Population des NU. 1984. *Manuel X. Techniques indirectes d'estimation démographique*. New York : Nations Unies, Département des affaires économiques et sociales internationales, ST/ESA/SER.A/81. http://unstats.un.org/unsd/demographic/standmeth/handbooks/Manuel_X-fr.pdf
- Martin LG. 1980. "A modification for use in destabilized populations of Brass's Technique for estimating completeness of death

Les variations selon l'âge de la complétude de l'enregistrement des décès peuvent induire une courbure dans la série de points du graphique. En conséquence, une des forces de cette méthode est de faire une vérification graphique : lorsque les points correspondants aux âges successifs sont à peu près alignés, on peut alors raisonnablement supposer que la complétude est constante selon l'âge. Et si certains ne le sont pas, on peut alors choisir de limiter l'intervalle d'âge aux points alignés pour calculer l'estimation de la complétude.

- registration", *Population Studies* **34** : 381-395. doi : <http://dx.doi.org/10.2307/2175194>
- Preston SH, AJ Coale, J Trussell and M Weinstein. 1980. "Estimating the completeness of reporting of adult deaths in populations that are approximately stable", *Population Index* **46** : 179-202. doi : <http://dx.doi.org/10.2307/2736122>
- Preston SH and KH Hill. 1980. "Estimating the completeness of death registration", *Population Studies* **34** : 394-366. doi : <http://dx.doi.org/10.2307/2175192>
- Rashad HM. 1978. "The Estimation of Adult Mortality from Defective Registration Data." Unpublished PhD thesis, London : University of London.
- UN Population Division. 2011. *World Population Prospects : The 2010 Revision, Volume I : Comprehensive Tables*. New York : United Nations, Department of Economic and Social Affairs, ST/ESA/SER.A/313. http://esa.un.org/unpd/wpp/Documentation/pdf/WPP2010_Volume-I_Comprehensive-Tables.pdf
- Vincent P. 1951. "La mortalité des vieillards", *Population* **6** : 182-204. doi : <http://dx.doi.org/10.2307/1524149>

Chapitre 22 ESTIMATION INDIRECTE DE LA MORTALITÉ ADULTE À PARTIR DES PROPORTIONS D'ORPHELINS

Ian M Timæus

DESCRIPTION DE LA MÉTHODE

La méthode des proportions d'orphelins consiste à estimer la mortalité des hommes et femmes adultes à partir des données sur la survie des père et mère des répondants. Pour que la méthode puisse être appliquée, il est nécessaire qu'au moins un recensement ou une enquête auprès de la population ait inclus les questions « Votre mère est-elle en vie ? » et « Votre père est-il en vie ? ». La mortalité peut être estimée sans qu'il soit nécessaire de demander aux répondants de se rappeler les dates de décès ou les âges au décès des individus décédés.

Comme les mères doivent avoir été vivantes au moment où les répondants sont nés, la durée où elles ont été exposées au risque de décéder est égale à l'âge des répondants. En prenant en compte l'âge moyen auquel les mères ont leurs enfants dans la population concernée, on peut prévoir la probabilité de survie de la table de mortalité de 25 ans à 25 ans plus un nombre arrondi d'années (n) en fonction du groupe d'âge des répondants (l_{25+n}/l_{25}), en s'appuyant sur la proportion des répondants de chaque groupe d'âge dont la mère est en vie. De même, en considérant l'âge moyen auquel les pères ont leurs enfants, on peut prévoir la probabilité de survie des hommes adultes dans la table de mortalité à partir des proportions de répondants dont les pères sont encore en vie. Comme les hommes tendent à être plus âgés que leurs épouses et autres partenaires, leur survie est mesurée entre 35 ans et 35 + n ans, où n dépend à nouveau du groupe d'âge des répondants.

Si la mortalité a évolué au fil du temps, les rapports de survie estimés reflètent des taux de mortalité qui ont prévalu à divers âges et diverses périodes. Une méthode de « localisation dans le temps » a été développée pour estimer combien d'années avant le recensement chaque proportion de survivants dans les cohortes égale la proportion de sur-

vivants par période. Ces intervalles s'accroissent avec l'âge des répondants, s'étageant entre 4 et 14 ans avant la collecte des données. Ainsi, si les rapports de survie estimées à partir des déclarations des répondants des différents groupes d'âge sont traduits en indice courant de la mortalité des adultes (par exemple $_{45}q_{15}$) grâce à un système de tables types de mortalité à 1 paramètre, ces statistiques se rapporteront à diverses dates et pourront être utilisées pour retracer l'évolution de la mortalité dans le temps.

Les méthodes des proportions d'orphelins ont un avantage sur les questions concernant les décès survenus dans le ménage, qui nécessitent des recensements ou des enquêtes exceptionnellement larges pour recueillir des informations sur un nombre suffisant de décès dans les ménages dans l'année précédant l'enquête, afin que les estimations de la mortalité soient assez précises pour être utiles. La méthode des proportions d'orphelins peut être utilisée dans des enquêtes beaucoup plus petites, même si toutes les méthodes d'estimation de la mortalité adulte nécessitent des données sur des milliers de ménages. En outre, la méthode ne nécessite pas que la population soit fermée à la migration. Les résultats ne seront toutefois pas représentatifs concernant des petits états ou des régions dans lesquels une proportion substantielle de la population aura émigré ou sera composée d'immigrés.

DONNÉES NÉCESSAIRES ET HYPOTHÈSES

Tableaux de données nécessaires

Pour estimer la mortalité des femmes adultes :

- La proportion de répondants dont la mère est vivante par groupe quinquennal d'âge du répondant. (Ceux qui ne savent pas si leur mère est vivante ou qui n'ont pas répondu à la question doivent être exclus des calculs.)

- Le nombre de naissances dans l'année précédant le recensement ou l'enquête classé par groupe quinquennal d'âge de la mère.

Pour estimer la mortalité des hommes adultes :

- La proportion de répondants dont le père est vivant par groupe quinquennal d'âge du répondant. (Ceux qui ne savent pas si leur père est vivant ou qui n'ont pas répondu à la question doivent être exclus des calculs.)
- Le nombre de naissances dans l'année précédant le recensement ou l'enquête classé par groupe quinquennal d'âge de la mère.
- Une estimation de la différence d'âge entre les hommes et les femmes ayant des enfants, par exemple la différence entre les âges médians des hommes et des femmes actuellement mariés.

Ces tableaux doivent généralement être établis séparément pour les répondants hommes et femmes et les estimations sont faites à partir des deux séries de proportions et pour les deux sexes ensemble.

Pour estimer la mortalité, le mieux est que les questions sur la survie des père et mère soient posées à toutes les personnes de moins de 50 ans (le plus simple peut être de poser les questions à tout le monde). Si toutefois elles n'ont concerné que les enfants orphelins, les données peuvent être utilisées pour une estimation ponctuelle de la mortalité des hommes et des femmes quelques années plus tôt.

Si des pondérations d'échantillonnage vous ont été fournies en même temps que les données, n'oubliez pas de les incorporer de manière appropriée dans votre logiciel quand vous établirez les tableaux de base.

Hypothèses importantes

Une limitation inhérente à la méthode des proportions d'orphelins est que les données relatives à la survie des parents ne peuvent être recueillies qu'auprès de ceux de leurs enfants eux-mêmes survivants. La survie des adultes sans enfants vivants n'est pas représentée dans les déclarations concernant la survie des parents. En outre, les parents ayant plus d'un enfant survivant sont surreprésentés par rapport à ceux n'en ayant qu'un, en proportion du nombre de leurs enfants survivants. La méthode ne produit donc des résultats non biaisés que si la mortalité des parents n'est pas liée au nombre de leurs enfants vivants au moment où les données sont recueillies. En général, le biais de sélection dû au non respect de cette hypothèse est modeste (Palloni, Massagli and Marcotte 1984). Dans des populations touchées par

une épidémie généralisée de Sida, il peut cependant être plus sévère. Les méthodes qui traitent au moins en partie les biais spécifiques à ce type de population sont présentées dans une section particulière de ce chapitre sur les orphelins dans les populations touchées par le Sida.

La méthode des proportions d'orphelins permet d'estimer l'évolution de la mortalité à partir des données fournies par les répondants de différents groupes d'âge : plus les répondants sont âgés, plus est ancien en moyenne le décès de leurs parents. Pour convertir les séries de mesures de la survie obtenues à partir de différents groupes d'âge en un indicateur unique qui puisse être comparé au fil du temps, on doit supposer que le schéma de mortalité aux âges adultes est correctement représenté par la table de mortalité standard choisie. Pour estimer la localisation dans le temps de ces mesures, il faut en outre supposer que la mortalité a reculé linéairement en termes de ce standard au long de la période considérée. Ces deux hypothèses risquent de poser problème dans des populations touchées par une épidémie généralisée de Sida.

TRAVAUX PRÉPARATOIRES ET RECHERCHES PRÉLIMINAIRES

Avant de débiter l'analyse, il convient de vérifier combien de répondants ont déclaré ne pas savoir si leur mère ou, plus fréquemment, leur père était vivant ou n'ont pas répondu du tout aux questions. Le taux de réponse à ces questions est généralement très élevé et on peut se contenter d'exclure de l'analyse ceux qui ont répondu « ne sait pas » ou qui n'ont pas répondu à la question. En effet, ceci revient à supposer que la proportion des parents de ces répondants qui sont décédés est la même que pour ceux qui ont répondu à la question. Toutefois, quelques enquêtes ont recueilli des données suffisamment incomplètes pour suggérer que le biais de non-réponse pourrait être un problème important. Il se peut par exemple que la plupart des personnes qui ne répondent pas à la question aient des parents décédés. Si c'est le cas, ces orphelins non déclarés comme tels peuvent représenter une proportion importante de l'ensemble des orphelins, en particulier aux jeunes âges et les estimations finales de la mortalité peuvent se trouver fortement biaisées vers le bas.

Un contrôle utile sur la qualité des données concernant les orphelins peut consister à comparer les réponses des hommes et des femmes du même âge. On ne s'attend pas à ce que la proportion de parents décédés diffère significativement entre les hommes et les femmes du même âge. Si

les proportions divergent parmi les répondants âgés, ce peut être dû à des différences entre hommes et femmes dans les erreurs de déclaration des âges ou cela peut indiquer que le genre qui déclare moins de décès de parents (en général les hommes) a sans doute davantage perdu contact avec la famille et suppose à tort que certain parent est encore vivant.

Précautions et mises en garde

- Les estimations obtenues par la méthode des proportions d'orphelins sont des probabilités conditionnelles de survie, c'est-à-dire des probabilités de survivre au cours d'un intervalle de l'âge adulte conditionnées par le fait d'être en vie au début de l'intervalle. Pour avoir une table de mortalité complète, il faut insérer des estimations de la survie depuis la naissance jusqu'à l'âge adulte tirées d'une autre source de données sur la mortalité des jeunes enfants.
- Les décès des parents ne surviennent pas en un point du temps mais peuvent être survenus à n'importe quel moment entre la naissance des répondants et la date de leur interview. La méthode des proportions d'orphelins ne peut indiquer qu'une évolution lissée de la mortalité ; elle est incapable de saisir des crises de mortalité de court terme ou des renversements abrupts de tendance comme ceux résultant du Sida après l'apparition d'une épidémie généralisée.
- Les données sur les proportions d'orphelins à la base des estimations de mortalité les plus récentes sont celles relatives aux enfants de 5-14 ans. En général, ceux-ci sont seulement quelques pour cents à être orphelins de mère et les estimations qui en résultent peuvent avoir de larges intervalles de confiance, même quand les enquêtes sont faites auprès de larges échantillons.
- Dans diverses applications en Afrique de l'Est et ailleurs, la méthode des proportions d'orphelins a donné des résultats qui indiquent des baisses rapides de mortalité peu vraisemblables et de fortes incohérences entre les estimations tirées d'enquêtes successives. Ceci apparaît dû à « l'effet d'adoption », c'est-à-dire une sous-déclaration de la perte des parents lorsque celle-ci est survenue alors que les enfants étaient très jeunes (Blacker 1984 ; Blacker and Gapere 1988 ; Hill 1984 ; Timæus 1986). Dans ces cas, les enfants sont souvent élevés par d'autres parents, dont ils apparaissent fréquemment comme les propres enfants dans les données de recensement. Ils sont alors considérés comme ayant un père ou une mère vivant, ce qui peut se traduire par des estimations très basses de la mortalité.

Cette erreur apparaît particulièrement fréquente en cas de décès de la mère. À mesure que les répondants avancent en âge, la probabilité s'accroît que, en plus du parent biologique, le parent d'accueil, le parent adoptif ou le beau-parent soit également décédé. Il en résulte que le biais est surtout fort pour les jeunes enfants, car le parent substitué a de fortes chances d'être vivant. S'il n'est pas reconnu, l'effet d'adoption conduit à une sous-estimation de la mortalité adulte récente et à une surestimation de l'ampleur de la baisse de la mortalité.

- Des estimations peuvent être faites à partir des données de répondants d'une quarantaine d'années, dont les parents sont souvent âgés et ont une mortalité très élevée. La précision avec laquelle on peut estimer à la fois la mortalité à partir des données sur la survie des parents et les dates de référence des estimations est donc beaucoup moindre que pour les répondants jeunes.
- La méthode des proportions d'orphelins n'inclut pas d'hypothèse sur l'absence de migration. Il peut néanmoins être difficile d'interpréter des estimations de la mortalité adulte tirées des proportions d'orphelins pour des unités géographiques infranationales, comme des fractions urbaines et rurales, ou pour des répondants ayant des caractéristiques socioéconomiques spécifiques. Certains des parents de répondants vivront ailleurs que les répondants eux-mêmes et leurs caractéristiques socioéconomiques différeront de celles de leurs enfants. Mais parents et enfants partagent généralement la même identité ethnique.

APPLICATION DE LA MÉTHODE

Les orphelins de mère

Étape 1a : Calculer \bar{M}^f , l'âge moyen des femmes à la naissance de leurs enfants

Pour appliquer la méthode des proportions d'orphelins, on doit disposer d'une estimation de l'âge moyen auquel les parents ont eu leurs enfants afin de prendre en compte la variation dans l'éventail des âges où ils ont été exposés au risque de décéder. L'âge moyen des femmes à la naissance de leurs enfants est généralement calculé à partir des données de recensement ou d'enquête sur les naissances récentes par groupe quinquennal d'âge à l'interview des femmes ayant eu un enfant. La mesure est simplement l'âge moyen des femmes ayant eu une naissance, sans ajustement prenant en

compte la structure par âge de la population, selon la formule suivante :

$$\bar{M}^f = \frac{\sum_{x=15}^{45} {}_5B_x(x+2)}{\sum_{x=15}^{45} {}_5B_x}$$

Dans cette équation, ${}_5B_x$ représente les naissances des femmes du groupe d'âge entre x et $x+4$ ans révolus et $(x+2)$ représente le point médian du groupe d'âge décalé vers le bas d'une demi année, pour tenir compte du fait que les femmes ayant une naissance au cours de l'année précédant l'interview l'ont eu en moyenne 6 mois plus tôt et qu'elles étaient donc alors 6 mois plus jeunes. Ce calcul peut être réalisé dans la feuille de calcul associée à ce chapitre (voir le site web). Si les données utilisées pour calculer \bar{M}^f sont classées par âge des femmes au moment de l'accouchement, le point médian de chaque groupe d'âge est égal à $x+2,5$.

Il n'est pas nécessaire d'ajuster les données relatives aux naissances pour corriger les erreurs sur la période de référence avant de calculer \bar{M}^f . En outre, les estimations de mortalité sont peu sensibles aux biais dans cet indicateur. Toutefois, s'il apparaît que la répartition par âge des naissances a été sévèrement déformée, parce que les femmes ont exagéré leur âge, le nombre de naissances par âge peut être recalculé à partir d'une distribution des âges et d'une distribution de la fécondité ajustées avant le calcul de \bar{M}^f .

En principe, l'âge moyen à la maternité doit se référer au moment où les répondants sont nés, ce qui peut se situer entre 5 et 45 ans avant le recueil des données sur la survie des parents. Une estimation s'appuyant sur les données de fécondité recueillies dans l'enquête même où sont collectées les informations sur les orphelins peut convenir dans des populations qui n'ont pas encore connu de baisse importante de la fécondité. Si on pense que la fécondité a reculé et qu'il existe des données de recensement ou d'enquête antérieures, \bar{M}^f peut aussi être calculé à partir de ces données pour voir s'il a évolué. Si c'est le cas, la meilleure façon de décider des valeurs finales de \bar{M}^f pour l'estimation de la mortalité adulte dépendra des données disponibles et du schéma d'évolution de la fécondité. Une option peut être de calculer \bar{M}^f à partir de données recueillies environ au moment où la fécondité a commencé à baisser, d'utiliser cette valeur pour les groupes d'âge des répondants nés à cette époque ou anté-

rieurement, et d'interpoler linéairement entre cette valeur et la valeur actuelle pour estimer pour des femmes plus jeunes.

Étape 2a : Calculer les rapports conditionnels de survie des tables de mortalité pour les femmes

Le fichier Excel (voir le site web) contient des feuilles séparées pour le calcul des rapports de survie des femmes adultes (**Orphelins de mère**) et des hommes adultes (**Orphelins de père**). Les feuilles sont établies à la fois pour analyser les données des répondants masculins et féminins séparément et pour les combiner afin de produire des estimations fondées sur les données fournies simultanément par les deux sexes. Sur la feuille des orphelins de mère, il faut entrer soit le nombre de répondants de chaque sexe par groupe quinquennal d'âge dont la mère survit et le nombre de ceux qui ont répondu à la question, soit les proportions de ceux dont la mère survit calculées à partir de ces nombres.

La survie des femmes est estimée entre une limite inférieure de 25 ans et $25+n$ ans, où n est la limite supérieure de chacun des groupes d'âge successifs des répondants. Nous utilisons l'équation de régression suivante et les coefficients figurant au tableau 22.1 :

$${}_n p_{25} = a(n) + b(n)\bar{M}^f + c(n){}_5S_{n-5}$$

où ${}_5S_{n-5}$ est la proportion de répondants dans le groupe d'âge entre les anniversaires $n-5$ et n ayant leur mère vivante. Par exemple, quand n est égal à 10, la survie dans les tables de mortalité est estimée sur un intervalle de 10 ans entre les 25^e et 35^e anniversaires, à partir des données sur la survie des mères fournies par les répondants de 5-9 ans.

TABLEAU 22.1 Coefficients pour l'estimation de la survie des femmes à partir des proportions de répondants dont la mère est vivante

| n | $a(n)$ | $b(n)$ | $c(n)$ |
|-----|---------|---------|--------|
| 10 | -0,2894 | 0,00125 | 1,2559 |
| 15 | -0,1718 | 0,00222 | 1,1123 |
| 20 | -0,1513 | 0,00372 | 1,0525 |
| 25 | -0,1808 | 0,00586 | 1,0267 |
| 30 | -0,2511 | 0,00885 | 1,0219 |
| 35 | -0,3644 | 0,01287 | 1,0380 |
| 40 | -0,5181 | 0,01795 | 1,0753 |
| 45 | -0,6880 | 0,02343 | 1,1276 |
| 50 | -0,8054 | 0,02721 | 1,1678 |

Source : Timæus (1992)

Orphelins de père

Étape 1b : Calculer \bar{M}^m , l'âge moyen des hommes à la naissance de leurs enfants

En principe, la meilleure façon de calculer l'âge moyen auquel les hommes ont des enfants consiste à calculer l'âge moyen des partenaires des femmes qui ont eu une naissance une année donnée. Mais la méthode nécessite qu'existe un lien entre les données des femmes et celles de leur partenaire, ce qui n'est pas le cas la plupart du temps.

L'âge moyen auquel les hommes ont leurs enfants est donc généralement estimé en ajoutant à l'âge moyen des femmes à la maternité une mesure de la différence d'âge entre les hommes et les femmes ayant des enfants : $\bar{M}^m = \bar{M}^f + d$.

Une estimation de cette différence qui peut être calculée aisément à partir des données de recensement est la différence entre les âges médians des hommes et des femmes actuellement mariés. C'est une solution plus appropriée que la différence entre les âges moyens au premier mariage des hommes et des femmes, dans les populations où les dissolutions d'union et les mariages polygames sont courants. La médiane est préférée à la moyenne de sorte que la surestimation différentielle des âges par les répondants âgés, dont la fécondité est vraisemblablement nulle de toute façon, n'affecte pas l'estimation.

Cette procédure d'estimation de l'âge moyen des hommes à la naissance de leurs enfants revient à supposer que les âges des pères des enfants nés de mères non mariées sont égaux, en moyenne, aux âges des pères des enfants nés de mères mariées. Ce n'est pas nécessairement le cas et cela peut introduire un biais significatif dans l'estimation de \bar{M}^m dans les populations où la fécondité hors mariage est importante. Ce problème est difficile à résoudre mais, heureusement, les estimations de la mortalité sont peu sensibles aux erreurs dans la mesure de \bar{M}^m .

Étape 2b : Calculer les rapports conditionnels de survie des tables de mortalité pour les hommes

Le fichier Excel (voir le site web) contient des feuilles séparées pour le calcul des rapports de survie pour les femmes adultes et les hommes adultes. Sur **la feuille des orphelins de père**, il faut entrer soit le nombre de répondants de chaque sexe par groupe quinquennal d'âge dont le père survit et le nombre de ceux qui ont répondu à la question, soit les proportions de ceux dont le père survit calculées à partir de ces nombres. Les données ne sont utilisées que pour les répondants âgés de 5 à 44 ans.

Chaque estimation de la survie des hommes est obtenue à partir des données de deux groupes quinquennaux d'âge adjacents. Pour chaque paire de groupes d'âge, sauf la paire la plus jeune analysée, la survie des hommes est mesurée de 35 ans à $35 + n$, où n est le point médian de la paire de groupes d'âge, en utilisant l'équation de régression suivante et les coefficients figurant au tableau 22.2 :

$${}_n p_{35} = a(n) + b(n)\bar{M}^m + c(n){}_5 S_{n-5} + d(n){}_5 S_n$$

Par exemple, quand n est égal à 10, la survie de la table de mortalité est estimée sur un intervalle de 10 ans du 35^e au 45^e anniversaire, à partir des données sur la survie des pères fournies par les répondants des deux groupes d'âge 5-9 et 10-14 ans.

Pour la paire de groupes d'âge la plus jeune, ce rapport de survie est nettement supérieur aux proportions de pères vivants, ce qui rend les estimations très sensibles au schéma exact de mortalité par âge dans la population. Des estimations plus robustes sont obtenues grâce à l'équation :

$${}_{15} p_{35} = a(10) + b(10)\bar{M}^m + c(10){}_5 S_5 + d(n){}_5 S_{10}$$

En d'autres termes, le même rapport de survie, ${}_{15} q_{35}$, est estimé à partir des données des répondants âgés à la fois de 5-14 et de 10-19 ans.

TABLEAU 22.2 Coefficients pour l'estimation de la survie des hommes à partir des proportions de répondants dont le père est vivant

| n | $a(n)$ | $b(n)$ | $c(n)$ | $d(n)$ |
|-----|---------|---------|--------|---------|
| 10 | -0,5578 | 0,00040 | 1,4708 | 0,0698 |
| 15 | -0,4013 | 0,00576 | 1,5602 | -0,3522 |
| 20 | -0,3329 | 0,01031 | 0,6656 | 0,3419 |
| 25 | -0,4726 | 0,01559 | 0,2161 | 0,7896 |
| 30 | -0,7056 | 0,02076 | 0,1997 | 0,9066 |
| 35 | -0,9153 | 0,02493 | 0,3484 | 0,8631 |
| 40 | -0,9950 | 0,02635 | 0,4269 | 0,8263 |

Source : Timæus (1992)

Orphelins de mère et de père

Étape 3 : Convertir les rapports de survie en estimations du niveau de mortalité

Pour établir une évolution de la mortalité à partir d'une série de rapports de survie, ${}_n p_b$, obtenus à partir de différents groupes d'âge de répondants et se rapportant à différentes

dates, il convient de les convertir en un indice commun de mortalité qui puisse être comparé dans le temps. Nous le faisons en ajustant une table-type de mortalité à 1 paramètre à chaque mesure et en obtenant l'indice commun de mortalité de la table-type.

Des indices très variés ont ainsi été utilisés, incluant les paramètres de niveau de divers systèmes de tables-types de mortalité, des rapports de survie comme ${}_{35}p_{30}$, qui correspondent bien à l'intervalle d'âge couvert par la méthode des proportions d'orphelins, l'espérance de vie à divers âges entre 5 et 30 ans et l'espérance de vie temporaire entre 25 et 70 ans, ${}_{45}e_{25}$. L'utilisation des paramètres des tables-types a l'avantage de souligner que la table de mortalité complète est estimée par l'ajustement d'un modèle, plutôt que d'être mesurée directement. Les mesures de l'espérance de vie résument la survie sur l'ensemble des âges adultes, alors que l'utilisation des rapports de survie ou des espérances de vie temporaires évite l'extrapolation aux âges avancés à partir de mesures faites à des âges plus jeunes. Récemment, les estimations ont été présentées de plus en plus souvent en termes de probabilité de décéder entre le 15^e et le 60^e anniversaire, ${}_{45}q_{15}$, car cette mesure a la faveur de plusieurs agences internationales comme indicateur résumé de la mortalité des adultes jeunes et d'âge moyen.

Dans les applications de la méthode des proportions d'orphelins présentées ici, les rapports de survie sont convertis en paramètre α d'un système à 1 paramètre de tables-types de mortalité logit relationnelles puis en estimations de la probabilité de décéder entre le 15^e et le 60^e anniversaire, ${}_{45}q_{15}$, ou en probabilité de décéder entre le 30^e et le 60^e anniversaire, ${}_{30}q_{30}$. Cette dernière mesure est utile quand il s'agit de comparer entre elles des séries d'estimations tirées des proportions d'orphelins sans extrapoler de la mortalité sur un intervalle d'âge à la mortalité sur un autre intervalle davantage qu'il n'est nécessaire. Les paramètres des modèles à 1 paramètre sont calculés à partir des estimations de ${}_np_b$:

$$\alpha = -\frac{1}{2} \ln \left(1 + \frac{\frac{{}_np_b}{l_{b+n}^s} - \frac{1}{l_b^s}}{1 - {}_np_b} \right)$$

où les estimations de ${}_np_b$ proviennent de l'étape 2, avec $b = 25$ pour les estimations de la survie des femmes et $b = 35$ pour celles de la survie des hommes, et où les valeurs de l_x^s proviennent de la table de mortalité standard. On obtient

donc une série de valeurs de α correspondant aux mesures de la survie conditionnelle prises à partir des données de différents groupes d'âge de répondants. Des valeurs élevées de α correspondent à une mortalité forte. Et pour chaque α :

$${}_{45}q_{15} = 1 - \frac{1 + e^{2(\alpha + Y_s(15))}}{1 + e^{2(\alpha + Y_s(60))}}$$

ou

$${}_{30}q_{30} = 1 - \frac{1 + e^{2(\alpha + Y_s(30))}}{1 + e^{2(\alpha + Y_s(60))}}$$

Le fichier Excel joint permet de calculer α , ${}_{45}q_{15}$ et ${}_{30}q_{30}$ en utilisant un standard tiré soit du réseau Général des tables-types de Nations Unies, soit d'un des quatre réseaux des tables-types de Princeton. La table standard doit être choisie parce qu'elle a un schéma de mortalité par âge aux âges adultes qui ressemble à celui de la population étudiée. Une autre table peut être utilisée comme standard, s'il y a des raisons de penser qu'elle ressemble davantage au schéma de mortalité adulte de la population étudiée. La table de mortalité la mieux adaptée n'appartient pas nécessairement au réseau de tables-types qui reproduit le mieux la relation entre la mortalité aux jeunes âges et à l'âge adulte. Si on ne sait rien du schéma par âge de la mortalité des adultes, il est recommandé d'utiliser le réseau Général des Nations Unies ou le réseau Ouest de Princeton.

Étape 4 : Calculer la localisation dans le temps des estimations

Chaque rapport de survie obtenu par la méthode des proportions d'orphelins représente une moyenne de la mortalité ayant eu cours pendant les N années où les parents ont été exposés au risque de décéder. Le point du temps où le rapport est égal à la survie du moment dépend du niveau de la mortalité et peut être estimé à partir de la proportion de pères ou de mères en vie, de l'âge des répondants et de l'âge moyen des parents à la naissance des enfants.

La seule information supplémentaire nécessaire pour calculer les dates auxquelles s'applique chacune des estimations de la mortalité adulte est la date où les répondants ont été interrogés sur la survie de leurs père et mère. Cette information peut être calculée comme la moyenne des dates où les interviews ont eu lieu ou peut être prise comme le point médian de la période de travail sur le terrain si les dates précises des interviews ne sont pas disponibles.

Les estimations sont égales à la date du travail de terrain moins T , où T est calculé comme suit pour les femmes :

$$T = \frac{N}{2} \left(1 - \frac{1}{3} \ln \left({}_5S_{n-5} \right) + \frac{1}{3} \ln \left(\frac{80 - \bar{M}^f - N}{80 - \bar{M}^f} \right) \right)$$

Comme la mortalité des mères est estimée à partir des données des répondants d'un seul groupe quinquennal d'âge, leur durée moyenne d'exposition au risque de décéder, N , est le point médian de ce groupe d'âge (c'est-à-dire 7,5 ans, 12,5 ans, etc.). Mais comme des paires de groupes d'âge sont utilisées pour estimer la mortalité des hommes, N est cette fois l'âge qui sépare les deux groupes d'âge (c'est-à-dire 10 ans, 15 ans, etc.). Comme en outre les pères peuvent être décédés pendant les 9 mois avant que les répondants soient nés, T devient

$$T = \frac{(N + 0,75)}{2} \left(1 - \frac{1}{3} \ln \left(\sqrt{{}_5S_{n-5} \cdot {}_5S_n} \right) + \frac{1}{3} \ln \left(\frac{80 - \bar{M}^m - N}{80 - (\bar{M}^m - 0,75)} \right) \right).$$

Pour la mortalité des hommes aussi bien que des femmes, la localisation dans le temps des estimations obtenues à partir des proportions d'orphelins parmi les répondants des différents groupes d'âge au-delà de 30 ans est généralement la même. La date à laquelle peuvent avoir lieu les premiers décès de parents devient plus précoce à mesure qu'augmente l'âge des répondants, mais ceci est contrebalancé par l'accroissement de la proportion de décès survenus peu avant l'enquête à mesure que les parents sont plus âgés. Il arrive que la localisation dans le temps des indicateurs de mortalité tirés des données des groupes d'âge les plus avancés soit plus récente que la localisation des indicateurs tirés des données de répondants cinq ans plus jeunes. Étant donné la nature approximative de l'équivalence posée entre mesures de mortalité des cohortes et estimations du moment, il est souvent recommandé d'omettre les estimations tirées des groupes d'âge les plus avancés ou de faire la moyenne des deux ou trois estimations finales et des dates auxquelles elles se réfèrent pour n'en tirer qu'une seule mesure.

EXEMPLE

Nous utilisons les données sur la survie des mères et des pères recueillies lors du recensement de 1997 en Irak disponibles grâce à IPUMS. Nous ne présentons les résultats que pour les deux sexes réunis.

Étape 1a : Calculer \bar{M}^f , l'âge moyen des femmes à la naissance de leurs enfants

L'âge moyen à la naissance des enfants est une moyenne simple des âges des femmes à l'accouchement. Il peut être calculé à partir des données individuelles ou estimé approximativement à partir d'un tableau des naissances par groupe quinquennal d'âge des mères (tableau 22.3).

$$\bar{M}^f = \frac{173338360}{613080} = 28,28$$

TABLEAU 22.3 Calcul de l'âge moyen à la naissance des enfants, Irak, 1997

| Groupe d'âge | Naissances de la dernière année $B(i)$ | Âge au point médian N | $B(i) * N$ |
|--------------|--|-------------------------|------------|
| 15-19 | 50 720 | 17 | 862 240 |
| 20-24 | 147 040 | 22 | 3 234 880 |
| 25-29 | 174 350 | 27 | 4 707 450 |
| 30-34 | 129 860 | 32 | 4 155 520 |
| 35-39 | 66 440 | 37 | 2 458 280 |
| 40-44 | 35 900 | 42 | 1 507 800 |
| 45-49 | 8 770 | 47 | 412 190 |
| Totaux | 613 080 | | 17 338 360 |

Étape 2a : Calculer les rapports conditionnels de survie des tables de mortalité pour les femmes

Les rapports de survie sont dans la sixième colonne du tableau 22.4 ; ils sont calculés à partir des proportions de la quatrième colonne, en utilisant les coefficients de régression du tableau 22.1 et l'estimation de \bar{M}^f , égale à 28,28 ans à l'étape 1a. Par exemple, pour les répondants âgés de 5-9 ans :

$$\begin{aligned} {}_{10}P_{25} &= -0,2894 + 0,00125 \times 28,28 + 1,2559 \times 0,9916 \\ &= 0,9913. \end{aligned}$$

Notez que les mesures de la table de mortalité sont du même ordre que les proportions à partir desquelles elles ont été calculées.

Étape 1b : Calculer \bar{M}^m , l'âge moyen des hommes à la naissance de leurs enfants

L'âge moyen des hommes à la naissance de leurs enfants est obtenu en ajoutant la différence entre les âges médians des hommes et des femmes actuellement mariés à l'âge

TABEAU 22.4 Estimation de la survie des femmes, des dates auxquelles ces estimations sont localisées dans le temps, et estimations correspondantes de α et ${}_{30}q_{30}$ tirées des proportions d'orphelins de mère, Irak, Recensement de 1997, Sexes réunis

| Groupe d'âge | Total des répondants | Mère vivante | Proportion mère vivante | Âge n | $\frac{l(25+n)}{l(25)}$ | Niveau (α) | Probabilité de décéder (${}_{30}q_{30}$) | Date |
|--------------|----------------------|--------------|-------------------------|---------|-------------------------|---------------------|--|--------|
| 5-9 | 2 834 210 | 2 810 420 | 0,9916 | 10 | 0,9913 | - 0,756 | 0,072 | 1994,2 |
| 10-14 | 2 391 580 | 2 347 190 | 0,9814 | 15 | 0,9826 | - 0,670 | 0,084 | 1992,1 |
| 15-19 | 2 181 840 | 2 108 280 | 0,9663 | 20 | 0,9709 | - 0,625 | 0,090 | 1990,1 |
| 20-24 | 1 827 030 | 1 719 880 | 0,9414 | 25 | 0,9514 | - 0,561 | 0,101 | 1988,5 |
| 25-29 | 1 615 160 | 1 455 150 | 0,9009 | 30 | 0,9198 | - 0,497 | 0,112 | 1987,0 |
| 30-34 | 1 235 120 | 1 035 480 | 0,8384 | 35 | 0,8698 | - 0,438 | 0,124 | 1985,9 |
| 35-39 | 788 170 | 601 590 | 0,7633 | 40 | 0,8103 | - 0,447 | 0,122 | 1985,4 |
| 40-44 | 817 460 | 519 580 | 0,6356 | 45 | 0,6910 | - 0,375 | 0,137 | 1985,5 |
| 45-49 | 603 840 | 298 760 | 0,4948 | 50 | 0,5419 | - 0,346 | 0,144 | 1988,3 |

TABEAU 22.5 Âges des hommes et des femmes actuellement mariés, Irak, 1997

| Groupe d'âge | Hommes mariés | Femmes mariées | Proportion cumulée d'hommes | Proportion cumulée de femmes |
|--------------|------------------|------------------|-----------------------------|------------------------------|
| 10-14 | 1 380 | 7 300 | 0,0005 | 0,0024 |
| 15-19 | 34 140 | 171 320 | 0,0122 | 0,0583 |
| 20-24 | 192 160 | 429 420 | 0,0782 | 0,1984 |
| 25-29 | 471 780 | 550 700 | 0,2401 | 0,3780 |
| 30-34 | 496 440 | 488 090 | 0,4105 | 0,5372 |
| 35-39 | 321 400 | 345 680 | 0,5208 | 0,6500 |
| 40-44 | 375 650 | 336 270 | 0,6497 | 0,7597 |
| 45-49 | 279 350 | 245 010 | 0,7456 | 0,8396 |
| 50-54 | 224 610 | 164 470 | 0,8227 | 0,8933 |
| 55-59 | 166 800 | 115 380 | 0,8800 | 0,9309 |
| 60-64 | 105 760 | 75 470 | 0,9163 | 0,9555 |
| 65-69 | 99 360 | 63 890 | 0,9504 | 0,9764 |
| 70-74 | 59 000 | 37 220 | 0,9706 | 0,9885 |
| 75-79 | 41 110 | 19 700 | 0,9847 | 0,9949 |
| 80+ | 44 430 | 15 570 | 1,0000 | 1,0000 |
| Total | 2,913,370 | 3,065,490 | | |

moyen des femmes à la naissance de leurs enfants. On voit au tableau 22.5 que l'âge médian des hommes actuellement mariés est compris entre le point médian du groupe 30-34 ans et le point médian du groupe 35-39 ans. Par interpolation linéaire,

Âge médian des hommes

$$= 32,5 + 5 \times \frac{0,5 - 0,4105}{0,5208 - 0,4105} = 36,56$$

et

Âge médian des femmes

$$= 27,5 + 5 \times \frac{0,5 - 0,3780}{0,5372 - 0,3780} = 31,33$$

L'âge moyen des hommes à la naissance de leurs enfants est donc :

$$M^m = M^f + d = 28,28 + (36,56 - 31,33) = 33,51$$

Étape 2b : Calculer les rapports conditionnels de survie des tables de mortalité pour les hommes

Les rapports de survie sont dans la sixième colonne du tableau 22.6 ; ils sont calculés à partir des proportions de la quatrième colonne en utilisant les coefficients de régression du tableau 22.2 et l'estimation de \bar{M}^m , égale à 33,51 ans à l'étape 1b. Par exemple, pour les estimations finales du tableau 22.6 :

$$\begin{aligned} {}_{40}p_{35} &= -0,9950 + 0,02635 \times 33,51 + 0,4269 \times 0,4818 \\ &\quad + 0,8263 \times 0,3432 = 0,3772 \end{aligned}$$

L'estimation de 0,9336 (à partir des proportions de 0,9622 et 0,8999) et l'estimation de 0,9000 (à partir des proportions de 0,8999 et 0,8394) sont toutes deux des mesures de ${}_{15}p_{35}$, alors que l'estimation suivante, 0,8398, mesure ${}_{20}p_{35}$, et ainsi de suite.

TABLEAU 22.6 Estimation de la survie des hommes, des dates auxquelles ces estimations sont localisées dans le temps, et estimations correspondantes de α et ${}_{30}q_{30}$ tirées des proportions d'orphelins de père, Irak, Recensement de 1997, Sexes réunis

| Groupe d'âge | Total des répondants | Père vivant | Proportion père vivant | Âge n | $\frac{l(35+n)}{l(35)}$ | Niveau (α) | Probabilité de décéder (${}_{30}q_{30}$) | Date |
|--------------|----------------------|-------------|------------------------|---------|-------------------------|---------------------|--|--------|
| 5-9 | 2 834 210 | 2 727 110 | 0,9622 | 15 | 0,9336 | -0,216 | 0,175 | 1992,7 |
| 10-14 | 2 391 580 | 2 152 090 | 0,8999 | 15 | 0,9000 | 0,045 | 0,251 | 1990,6 |
| 15-19 | 2 181 840 | 1 831 390 | 0,8394 | 20 | 0,8398 | 0,057 | 0,254 | 1988,7 |
| 20-24 | 1 827 030 | 1 435 320 | 0,7856 | 25 | 0,7707 | 0,026 | 0,245 | 1987,0 |
| 25-29 | 1 615 160 | 1 127 370 | 0,6980 | 30 | 0,6629 | 0,043 | 0,250 | 1985,5 |
| 30-34 | 1 235 120 | 726 870 | 0,5885 | 35 | 0,5409 | 0,025 | 0,244 | 1984,6 |
| 35-39 | 788 170 | 379 770 | 0,4818 | 40 | 0,3772 | 0,065 | 0,257 | 1984,8 |
| 40-44 | 817 460 | 280 540 | 0,3432 | | | | | |

Étape 3 : Convertir les rapports de survie en estimations du niveau de mortalité

Les septièmes colonnes des tableaux 4 et 6 contiennent α , le paramètre de niveau du système des tables-types de mortalité logit relationnelles avec comme standard une table-type de mortalité des Nations Unies, réseau Général, pour les deux sexes avec $e_0 = 60$. Par exemple, α est calculé à partir de l'estimation de ${}_{10}p_{25}$ pour les femmes :

$$\alpha = -\frac{1}{2} \ln \left(1 + \frac{\frac{0,9913}{0,8335} - \frac{1}{0,86313}}{1 - 0,9913} \right) = -0,756.$$

Les huitièmes colonnes des tableaux 4 et 6 contiennent ${}_{30}q_{30}$, la probabilité de décéder entre le 30^e et le 60^e anniversaire. Ayant calculé $\alpha = -0,756$ à partir de ${}_{10}p_{25}$, la mesure correspondante de ${}_{30}q_{30}$ est :

$${}_{30}q_{30} = 1 - \frac{1 + e^{2(-0,756-0,8655)}}{1 + e^{2(-0,756-0,3064)}} = 0,072.$$

Étape 4 : Calculer la localisation dans le temps des estimations

Les dates dans les neuvièmes colonnes du tableau 22.4 et du tableau 22.6 sont calculées en soustrayant la localisation des estimations dans le temps à la date du recensement de 1997 en Irak, soit le 16 octobre. En numération décimale, cette date devient 1997,79.

La localisation dans le temps de chaque estimation est calculée à partir des âges des répondants, l'estimation appropriée de

$$T = \frac{(10 + 0,75)}{2} \left(1 - \frac{1}{3} \ln(\sqrt{0,9622 \cdot 0,8999}) + \frac{1}{3} \ln \left(\frac{80 - 33,51 - 10}{80 - (33,51 - 0,75)} \right) \right) = \frac{10,75}{2} (1 - 0,0621) = 5,04.$$

La date à laquelle se réfère cette première estimation est donc $1997,79 - 5,04 = 1992,75$.

La localisation dans le temps de l'estimation de la mortalité des femmes calculée à partir des données fournies par les répondants âgés de 45-49 ans est près de trois ans postérieure à la localisation dans le temps de l'estimation calculée à partir des données des répondants âgés de 40-44 ans. Comme dans de nombreuses applications de la méthode des proportions d'orphelins, les données du groupe d'âge le plus élevé doivent être éliminées.

DIAGNOSTICS, ANALYSE ET INTERPRÉTATION

Contrôles et validation

Le nombre de répondants qui ont déclaré ne pas savoir si leur mère ou leur père est vivant ou qui n'ont pas répondu à la question doit être vérifié, avant que les données soient retirées en vue de l'analyse. Si ce nombre est important, il se peut que l'information donnée par ceux qui ont répondu à ces questions ne

soit pas représentative de l'ensemble de la population. En outre, un taux élevé de non réponse peut indiquer que les enquêteurs ou les répondants ont eu des difficultés avec les questions. Il est alors possible que les réponses, même quand elles ont été fournies, ne soient pas fiables. S'il y a un taux élevé de non réponse à une question, il peut être intéressant de savoir si le défaut est concentré sur une minorité d'enquêteurs ou sur un certain type de répondants.

Si l'information sur la survie des mères et des pères a été obtenue à la fois de *répondants* hommes et femmes dans un recensement ou une grande enquête avec des marges d'erreur étroites, il est possible d'établir les proportions de pères et de mères encore en vie séparément pour les répondants de chaque sexe en vue de comparer la cohérence des déclarations. La cohérence des déclarations ne garantit pas leur exactitude, mais des différences statistiquement significatives entre les proportions calculées pour les répondants hommes et femmes impliquent qu'au moins un des deux sexes, et peut-être les deux, donnent des réponses inexactes.

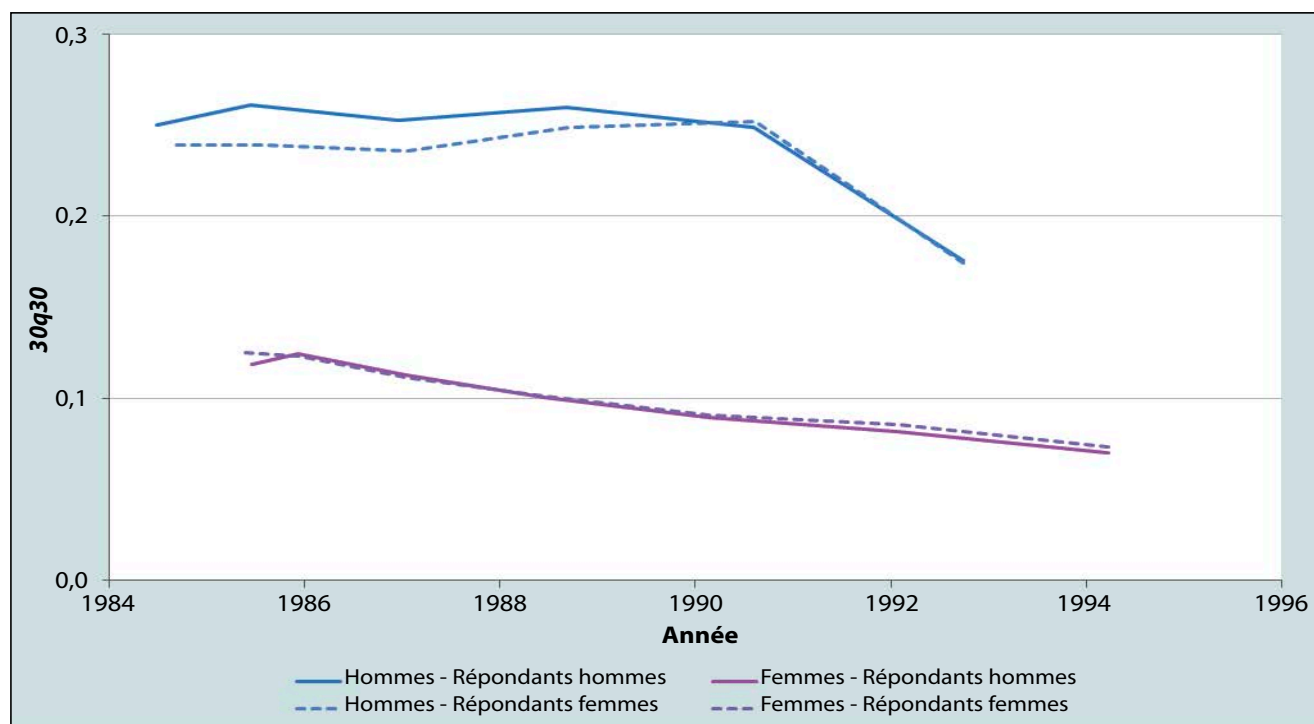
Il est courant de constater que les femmes déclarent des proportions de parents vivants plus basses que les hommes. Certains analystes pensent que c'est parce que les femmes

restent en contact plus étroit que les hommes avec leurs parents et que certains hommes disent que leurs parents sont en vie parce qu'ils ne savent pas qu'ils sont en fait décédés. Si c'est vrai, les données des femmes sont plus exactes que celles des hommes. Mais il n'y a pas de preuve convaincante qui confirme cette interprétation et d'autres facteurs, en particulier les différences entre hommes et femmes dans les erreurs sur la déclaration de leur âge, peuvent aussi être à l'origine des incohérences dans les proportions déclarées par les hommes et les femmes. De plus, dans les enquêtes ne portant que sur quelques milliers de ménages, les marges d'erreur des estimations pour les premiers groupes d'âge peuvent être larges par rapport aux proportions de mères ou de pères décédés. S'il n'y pas de raison particulière pour privilégier les réponses des femmes, nous conseillons donc d'établir les estimations finales de la mortalité adulte sur les réponses combinées des deux sexes.

Interprétation

Les résultats de l'analyse des données sur les proportions d'orphelins tirées du recensement de 1997 en Irak sont représentés graphiquement sur la figure 22.1. Il ne s'agit

FIGURE 22.1 Evolution de la probabilité de décéder entre le 30^e et le 60 anniversaire estimée à partir des proportions d'orphelins, Irak, recensement de 1997



pas simplement d'interpréter ces courbes en lisant des tendances, mais de faire le partage entre ce qui reflète des évolutions réelles de la mortalité et ce qui tient à des problèmes de données ou de méthode.

Dans cette application de la méthode, les estimations tirées des données fournies par les répondants hommes et femmes sont semblables, ce qui rassure sur la qualité à la fois de la déclaration des âges et des déclarations sur la survie des parents en Irak. Néanmoins, les femmes adultes sont un peu moins enclines à déclarer leur père vivant que les hommes adultes du même âge et l'ampleur de cette incohérence dans les déclarations s'accroît à mesure que l'âge des répondants augmente. Cet écart est à l'inverse de ce qu'on observe dans la plupart des populations. On pourrait l'expliquer par une tendance à surestimer leur âge plus forte chez les femmes que chez les hommes, ou une tendance à le sous-estimer plus forte chez les hommes que chez les femmes, mais ces deux schémas d'erreur semblent peu vraisemblables. Autre explication possible : certains hommes adultes pourraient avoir déclaré décédé leur père survivant ou certaines femmes pourraient avoir déclaré vivant leur père décédé. Cette dernière possibilité est peut-être la plus plausible des quatre explications, mais il n'y a pas de preuve solide pour la confirmer. Le plus raisonnable consiste peut-être à supposer que les déclarations des hommes et des femmes sont légèrement biaisées dans des sens opposés et à utiliser les estimations calculées sur les répondants des deux sexes réunis.

Quelle que soit la série de données qu'on choisisse, les estimations des proportions d'orphelins suggèrent un recul constant de la mortalité des femmes adultes en Irak entre le milieu des années 1980 et le milieu des années 1990, le niveau atteint à cette dernière date étant à peine supérieur à celui dans la plupart des pays occidentaux. La régularité de la tendance est rassurante mais le bas niveau de la mortalité est étonnant. Sans autre donnée pour corroborer ces résultats, on accueillera ceux-ci avec réserve car ils pourraient être sous-estimés, les décès de certaines mères pouvant avoir été omis, en particulier ceux qui se sont produits lorsque les répondants étaient de jeunes enfants. Ceci tend à exagérer l'impression de baisse de la mortalité.

La mortalité des hommes adultes est estimée à un niveau bien supérieur à celle des femmes. Les résultats suggèrent en outre une stagnation de la mortalité masculine pendant la plupart des années 1980, avant une baisse rapide au début de la décennie suivante. Il est vraisemblable que ce schéma reflète l'impact sur les proportions d'orphelins des

pertes durant la guerre Iran-Irak de 1980-1988, au cours de laquelle un quart à un demi million de soldats irakiens sont sans doute décédés. Les statistiques d'état civil font généralement la distinction entre les décès civils et militaires, mais pas les données de recensement et d'enquête, qui les font apparaître ensemble. L'estimation la plus récente sous-estime sans doute la mortalité des hommes du fait d'une sous-déclaration des pères décédés, mais il se peut qu'elle soit largement inférieure aux autres car près de la moitié des répondants sur la déclaration desquels elle s'appuie sont nés après la fin de la guerre. C'est pourquoi leur père a de grandes chances d'avoir survécu. En revanche, toutes les autres estimations sont fondées sur les déclarations de répondants nés avant 1988. Les pères des répondants âgés de 30-39 ans, qui ont fourni l'information utilisée dans l'estimation la plus ancienne, avaient environ 50 ans en moyenne au début de la guerre. La plupart d'entre eux étaient trop âgés pour être envoyés au front.

Il faut cependant rappeler que la méthode des proportions d'orphelins tend à lisser les fluctuations de la mortalité, puisque chaque estimation est fondée sur les décès survenus sur l'ensemble des années séparant la naissance des répondants de la date à laquelle les données sont recueillies. En outre, la tendance estimée peut être quelque peu déformée si la table-type de mortalité adoptée pour le calcul de ${}_xq_{30}$ est inappropriée. Ces résultats ne peuvent donc pas être interprétés comme s'ils reflétaient précisément l'évolution de la mortalité masculine en Irak entre le début des années 1980 et le début des années 1990. Il se peut par exemple que la mortalité des hommes ait augmenté quelque peu avant de baisser, ou qu'elle ait fluctué fortement d'une année à l'autre pendant la guerre, en fonction du nombre des victimes.

DESCRIPTION DÉTAILLÉE DE LA MÉTHODE

Introduction

L'usage de l'information sur les parents décédés pour une estimation de la mortalité adulte a d'abord été envisagé par Louis Henry (1960). Si la survie peut être considérée comme une fonction linéaire de l'âge, la proportion d'enfants orphelins à tout âge x , $O(x)$, dépend, à très peu près, de la probabilité de survivre entre l'âge moyen (\bar{M}) des parents à la naissance de leurs enfants et l'âge $\bar{M} + x$. Cependant, cette relation sous-estime le rapport de survie (${}_x p_{\bar{M}}$) car l'accroissement des taux de mortalité avec l'âge est rapide et la courbe de survie aux âges adultes, convexe. Le faible risque

de décéder de ceux qui ont leurs enfants jeunes ne suffit pas à compenser le risque élevé de ceux qui ont leurs enfants relativement tard. Pour en tenir compte, Henry a proposé un ajustement fondé sur la variance des âges à la naissance des enfants.

La mise au point ultérieure de méthodes simples et robustes pour estimer la mortalité à partir des proportions d'orphelins doit beaucoup à William Brass. Hill et Blacker ont aussi joué un rôle important en établissant des moyens d'estimer la mortalité des hommes à partir des proportions d'orphelins de père. Le développement de la procédure est passé par un certain nombre d'étapes, plusieurs variantes ont circulé de manière informelle et ont finalement été publiées dans Brass (1975). La version ultime de la méthode de Brass est, toutefois, celle publiée dans Brass et Hill (1973). Elle consiste à calculer des mesures de la survie dans la table de mortalité comme une moyenne pondérée des proportions de répondants ayant des parents encore en vie dans deux groupes d'âge contigus, les pondérations étant déterminées par l'âge moyen à la naissance des enfants (\bar{M}) dans la population, et permettant d'introduire un ajustement qui en tient compte.

Par la suite, plusieurs chercheurs ont proposé des méthodes de régression pour prédire la mortalité des femmes à partir des proportions de répondants dont la mère survit (Hill and Trussell 1977 ; Palloni and Heligman 1985). Timæus (1992) a démontré que les facteurs de pondération tendent à sous-estimer la mortalité quand celle-ci est supérieure au standard utilisé pour les calculer et à surestimer la mortalité quand elle est inférieure au standard. Une approche par la régression doit généralement donner des estimations plus précises, en particulier de la mortalité des hommes, car le terme constant introduit une flexibilité supplémentaire permettant de modéliser plus précisément la relation entre la survie des parents et la fonction de survie des tables de mortalité à différents niveaux d'espérance de vie. C'est pourquoi Timæus (1992) a proposé des coefficients de régression pour la mortalité des hommes, ainsi que des coefficients obtenus à partir d'hypothèses cohérentes pour l'estimation de la mortalité des femmes. C'est cette variante de la méthode des proportions d'orphelins qui est présentée ici.

Exposé mathématique de la méthode des proportions d'orphelins

Soit ${}_aV_y$ le nombre d'enfants nés a années avant un recensement ou une enquête, issus de femmes âgées de y au

moment de la naissance de l'enfant. La probabilité que les enfants soient encore en vie au moment d'être interrogés est $l(a)$ et la probabilité que leur mère soit encore en vie est $l(y + a) / l(y)$. Si on suppose que la mortalité des orphelins est égale à celle des enfants ayant encore leurs parents, la proportion de répondants âgés de a ayant encore leur mère, $S(a)$, est

$$S(a) = \frac{l(a) \int_a^{\omega} V_y l(y + a) / l(y) dy}{l(a) \int_a^{\omega} V_y dy} \quad (1)$$

où les intégrales sont prises sur l'ensemble des âges à la naissance des enfants, de s à ω .

Le nombre d'enfants nés de femmes âgées de y est une fonction du nombre de femmes âgées de y et du taux de fécondité à l'âge y . Si nous supposons une structure par âge stable, on a

$${}_aV_y = B(0)e^{-r(y+a)}l(y)f(y)$$

où r est le taux d'accroissement naturel et $f(y)$ est le taux de fécondité des mères à l'âge y . C'est pourquoi, la proportion de répondants dans un groupe quinquennal d'âge de x à $x + 5$ ayant leur mère encore en vie est

$${}_5S_x = \frac{\int_x^{x+5} e^{-ra} l(a) \int_s^{\omega} e^{-ry} f(y) l(y + a) dy da}{\int_x^{x+5} e^{-ra} l(a) \int_s^{\omega} e^{-ry} f(y) l(y) dy da}$$

L'exposition des pères va de la conception des répondants à leur âge actuel. La prise en compte du fait qu'un père peut décéder entre la conception et la naissance de son enfant rend un peu plus compliquée l'équation de la proportion de répondants dans un groupe quinquennal d'âge de x à $x + 5$ ayant leur père encore en vie :

$${}_5S_x = \frac{\int_x^{x+5} e^{-ra} l(a) \int_s^{\omega} e^{-ry} f(y) l(y) l(y + a) / l(y - 0.75) dy da}{\int_x^{x+5} e^{-ra} l(a) \int_s^{\omega} e^{-ry} f(y) l(y) dy da}$$

Dans cette équation, $f(y)$ représente le schéma de fécondité par âge, et $l(a)$ la fonction de survie de la table de mortalité, pour les hommes au lieu des femmes et les âges entre les-

quels la naissance des enfants survient, s et ω , sont aussi les âges des hommes. Il n'y a pas de façon simple d'intégrer le numérateur ou le dénominateur de ces expressions de ${}_5S_x$, qui doivent donc être évaluées numériquement.

Mise en œuvre de la méthode

En principe, rien n'empêche de résoudre l'équation 1 directement pour déterminer la fonction de survie de la table de mortalité, en utilisant le solveur d'Excel ou un outil équivalent et une structure par âge appropriée pour ceux qui donnent naissance aux enfants ${}_aV_y$. Pour obtenir une solution unique, il faut encore faire une hypothèse sur le schéma par âge de la mortalité adulte, par exemple en adoptant un système à 1 paramètre de tables-types de mortalité relationnelles. En pratique, les estimations sont généralement obtenues à partir d'un modèle de régression qui prend en compte \bar{M} , lequel a été ajusté à des données simulées sur la survie des parents, générées pour des populations ayant un large éventail de structures par âge et de profils de fécondité et de mortalité (Timæus 1992).

Les proportions de mères encore en vie dans les groupes d'âge de x à $x + 5$ sont plus étroitement corrélées à la survie des femmes dans l'intervalle d'âge entre 25 ans et $x + 5$ qu'à tout autre rapport de survie, aussi cette mesure peut elle être estimée plus précisément que toute autre à partir des proportions de mères survivantes :

$${}_n p_{25} = a(n) + b(n)\bar{M}^f + c(n){}_5S_{n-5}$$

En particulier dans les populations de fécondité basse et d'âge moyen élevé à la naissance des enfants, la précision des estimations est augmentée significativement si on estime la survie à partir des proportions de pères encore en vie dans deux groupes d'âge contigus. Le groupe d'âge supplémentaire saisit les effets des variations du taux auquel la mortalité augmente avec l'âge. À la plupart des âges, les proportions de pères encore en vie dans les groupes d'âge de x à $x + 10$ sont plus étroitement corrélées à la survie des hommes dans l'intervalle d'âge entre 35 ans et $x + 5$ qu'à tout autre rapport de survie. L'équation de régression est donc

$${}_n p_{35} = a(n) + b(n)\bar{M}^m + c(n){}_5S_{n-5} + d(n){}_5S_n$$

Les deux premiers groupes d'âge, les répondants âgés de 5-9 et 10-14 ans, prédisent la survie de 35 à 50 ans plus précisément que de 35 à 45 ans. L'équation devient donc

$${}_{15} p_{35} = a(10) + b(10)\bar{M}^m + c(10){}_5S_5 + d(n){}_5S_{10}$$

Exposé mathématique – localisation des estimations dans le temps

L'équation 1 peut être réécrite comme une moyenne pondérée des rapports de survie dans les cohortes :

$$S(a) = \int_s^{\omega} {}_a v_y {}_a p_y^c dy$$

où

$${}_a v_y = \frac{{}_a V_y}{\int_s^{\omega} {}_a V_y dy}$$

Ainsi, ${}_a v_y$ représente la contribution faite à $S(a)$, proportion de parents adultes survivant a années, par les parents qui ont commencé leur exposition au risque à l'âge y . Les méthodes de localisation dans le temps visent à estimer le temps T , où les mesures de la survie des cohortes qui ont permis de calculer la proportion de parents survivants, ${}_a p_y^c$, sont égales aux mesures du moment équivalentes, ${}_a p_y(T)$, de sorte que

$$S(a) = \int_s^{\omega} {}_a v_y {}_a p_y(T) dy$$

Si nous notons ${}_a g_y$ la durée moyenne écoulée depuis le décès des personnes décédées entre y et $y + a$, par définition

$${}_a g_y = \frac{\int_y^{y+a} (z - y - a) \mu(z) l(z) dz}{\int_y^{y+a} \mu(z) l(z) dz} \quad (2)$$

où $\mu(z)l(z)$ désigne les décès de la table de mortalité à l'âge z . Brass et Bangboye (1981) montrent que, si les profils de mortalité sont conformes à un système de tables-types de mortalité logit relationnelles à 1 paramètre (Brass 1971), et si on peut supposer l'évolution de la mortalité linéaire en α , le paramètre de ce système de tables-types, alors la date à laquelle la survie des adultes par cohorte est égale à la survie du moment est une moyenne pondérée des durées écoulées depuis le décès des parents du répondant.

$$T = \frac{\int_{x+5}^{x+5} l(a) \int_s^{\omega} {}_a v_y l(y + a, T) (1 - {}_a p_y(T)) {}_a g_y dy da}{\int_x^{x+5} l(a) \int_s^{\omega} {}_a v_y l(y + a, T) (1 - {}_a p_y(T)) dy da} \quad (3)$$

Cette localisation dans le temps dépend du niveau de la mortalité et des âges des parents, mais elle est indépendante du taux de variation en α . Pour établir l'équation 3, Brass et Bamgboye tirent parti d'une relation entre les variations de la mortalité au fil des âges et au fil du temps qui est propre au système logit relationnel de tables-types, mais il est possible d'arriver à des formules similaires pour T sur la base d'autres hypothèses raisonnables sur l'évolution dans le temps de la mortalité par âge (Palloni, Massagli and Marcotte 1984).

L'équation 3 peut être évaluée numériquement, en utilisant des valeurs de ${}_a v_y$ et pour des mesures de la table de mortalité choisies sur la base de données observées. Pour développer une procédure aisée d'estimation de T à partir des caractéristiques observées d'une population, il faut supposer une relation beaucoup plus simple que celle décrite par l'équation 3. Brass et Bamgboye (1981) font valoir que la variation de T avec a dans des intervalles d'âge limités est suffisamment proche de la linéarité pour que tous les répondants dans un groupe quinquennal d'âge puissent être traités comme regroupés à l'âge central N . Ils font valoir ensuite qu'aux âges et aux niveaux de mortalité où les méthodes indirectes sont utilisées pour estimer la mortalité adulte, la force de mortalité s'accroît à peu près exponentiellement avec l'âge. En conséquence, pour de telles applications, la variation de ${}_a g_y$ avec y est insignifiante. C'est pourquoi les facteurs de pondération de ${}_a g_y$ dans l'équation 2 ont peu d'effet et tous les parents adultes peuvent être traités comme si leur exposition au risque débutait à leur âge moyen à l'entrée, M . D'où l'approximation satisfaisante, $T = {}_N g_M$.

Si la survie des adultes recule linéairement avec l'âge, de sorte que le même nombre de décès survienne à chaque âge, alors ${}_N g_M$ est égal à $N/2$ quelle que soit la valeur de M . Dans des tables de mortalité moins extrêmes, la mortalité s'accroît avec l'âge plus rapidement que cela et les décès des parents sont concentrés aux âges avancés et, en conséquence, dans la portion récente de la période de N années. Ceci signifie que la localisation dans le temps des estimations est plus proche de l'enquête que $N/2$. En remplaçant $\mu(z)l(z)$ par $e^{kz}\mu(y)l(y)$ dans l'équation 2 et en développant la partie droite en puissances de N , Brass et Bamgboye (1981) démontrent que l'ajustement approprié pour ${}_N g_M$ est une fonction du niveau de mortalité tel que mesuré par k et de l'âge des répondants mesuré par N :

$${}_N g_M \approx \frac{N}{2} \left(1 - \frac{kN}{6} \right) \quad (4)$$

Brass et Bamgboye (1981) démontrent aussi que l'hypothèse d'une augmentation exponentielle de la mortalité avec l'âge implique que, dans un système de tables de mortalité logit relationnel :

$$\frac{e^{kN}}{e^{k_s N}} = \frac{({}_N p_M)^2}{(l_s(M+N) / l_s(M))^2}$$

En résolvant pour kN et en remplaçant cette expression dans l'équation 4, on obtient une estimation de ${}_N g_M$ et donc de T :

$$T = \frac{N}{2} \left(1 - \frac{\ln({}_N p_M)}{3} + \frac{1}{3} \ln \left(\frac{l_s(M+N)}{l_s(M)} \right) - \frac{k_s N}{6} \right) \quad (5)$$

Dans cette formulation, les localisations dans le temps des mesures de survie conditionnelle tirées des données sur les parents adultes sont estimées à la moitié de la durée d'exposition, N , réduite par un facteur qui dépend du niveau de la survie conditionnelle comparé au niveau de la survie dans une table de mortalité standard.

Ayant obtenu cette expression de T sur une base théorique, Brass (1985) approxime ${}_N p_M$ par ${}_5 S_x$ et adopte comme standard une table de mortalité où $l_s(x)$ est linéaire aux âges adultes et est pris égal à $(1 - x/80)/2$. Comme $l_s(x)$ est linéaire, $T = \frac{1}{2}N$ et k_s devient nul. T est donc estimé à partir des données observées grâce à

$$T = \frac{N}{2} \left(1 - \frac{\ln({}_5 S_x)}{3} + \frac{1}{3} \ln \left(\frac{80 - M - N}{80 - M} \right) \right) \quad (6)$$

L'accord entre les estimations de T obtenues par cette procédure simplifiée et celles obtenues en calculant l'équation 2 est très étroit jusqu'à ce que l'âge des parents, $M+N$, atteigne 75 ans. Mais à mesure que $M+N$ s'accroît, le schéma par âge de la mortalité commence à avoir un effet important sur la durée moyenne écoulée depuis les décès des parents adultes, ${}_N g_M$. Par exemple, si la force de mortalité s'accroît relativement lentement avec l'âge par rapport au standard, une proportion élevée de parents décédés seront morts il y a relativement longtemps. S'il y a des signes d'existence d'un tel schéma de mortalité par âge ou d'un schéma opposé, il est possible d'en tenir compte en définissant un standard plus approprié en modifiant β , le second paramètre dans un système logit relationnel de tables de mortalité (Brass 1971). Dans ce cas, $l_s(x)$ n'est plus linéaire avec l'âge dans la table

retranchant une mortalité forte utilisée comme standard et le terme en k_i dans l'équation 5 doit être retranché des estimations de T tirées de l'équation 6. En faisant les mêmes hypothèses que précédemment sur la forme de, l'ajustement est :

$$\delta T = -\frac{N}{12}(\beta - 1) \ln \left(\frac{(80 - M)(80 + M + N)}{(80 + M)(80 - M - N)} \right)$$

Quand on utilise l'équation 6 pour calculer la localisation dans le temps des estimations de la mortalité féminine à partir de données sur la survie de la mère, ${}_5S_x$ devient ${}_5S_{n-5}$, la proportion de répondants dont la mère est encore en vie, N est le point médian de ce groupe d'âge, $n - 2,5$; et $M = \bar{M}$ est l'âge moyen des femmes à la naissance de leurs enfants. Quand on calcule la localisation dans le temps des estimations de la mortalité masculine à partir des données sur la survie des pères, ${}_5S_x$ est la proportion des répondants dont le père est encore en vie, M devient l'âge moyen des pères à la conception de leurs enfants ; et $N = n + 0,75$ où n est l'âge qui sépare les deux groupes d'âge, ${}_5S_{n-5}$ et ${}_5S_n$, qui contribuent à l'estimation de la survie des hommes. Afin de centrer ${}_5S_x$ sur l'âge n , il est calculé comme ${}_5S_x = \sqrt{{}_5S_{n-5} \cdot {}_5S_n}$.

EXTENSIONS DE LA MÉTHODE

Proportions d'orphelins

dans des populations affectées par le sida

Les estimations de mortalité obtenues par la méthode des proportions d'orphelins peuvent être affectées par des biais importants dans les populations soumises à une épidémie généralisée de VIH. Premièrement, les femmes séropositives ont une fécondité moindre que les femmes non infectées, pour des raisons à la fois comportementales et biologiques. Deuxièmement, les femmes séropositives peuvent transmettre le virus à leur enfant durant la grossesse, à l'accouchement et pendant qu'elles l'allaitent. Il en résulte que les femmes infectées par le VIH, qui ont un risque de décéder élevé, tendent à avoir moins d'enfants vivants que la population en général. La proportion de mères encore en vie se trouve de ce fait surestimée. Comme les femmes risquent d'avoir été infectées par ou d'infecter leur mari, les estimations de la mortalité masculine fondée sur les proportions d'orphelins dans les populations affectées par le VIH sont également biaisées vers le bas, quoique dans une moindre mesure que les estimations de la mortalité féminine. Un troisième biais important lié au VIH

dans les estimations par les proportions d'orphelins tient à ce que les coefficients utilisés pour convertir des proportions de parents encore en vie en indices conventionnels des tables de mortalité postulent un schéma par âge de la mortalité adulte très différent de celui qu'on observe dans les populations atteintes par une épidémie de Sida.

Ajustement pour tenir compte

du biais de sélection lié au VIH

Si on dispose d'estimations des caractéristiques pertinentes de l'épidémie de VIH dans une population au moment de la naissance des répondants, on peut tenir compte des deux premiers biais en ajustant à la baisse les proportions déclarées de mères encore en vie (Timæus et Nunn 1997). La procédure d'ajustement inclut l'hypothèse selon laquelle la plupart des individus déjà infectés par le VIH au moment de la naissance de leurs enfants vont décéder en quelques années. Cette hypothèse était raisonnable avant que des traitements soient répandus à grande échelle, mais elle a rapidement cessé de l'être après leur diffusion massive. En fait, si ces programmes se révèlent capables de réduire la mortalité des adultes séropositifs à un niveau proche de celui de la population générale, il n'y aura plus besoin d'ajustement.

Pour ajuster les données recueillies dans les recensements réalisés vers 2010 ou antérieurement, on doit disposer d'estimations de la prévalence de l'infection par VIH parmi les femmes fréquentant les consultations prénatales (P), le taux de transmission verticale du VIH (h), c'est-à-dire la proportion d'enfants qui reçoivent le VIH de leur mère, et le niveau relatif de la fécondité des femmes séropositives comparée à celle des femmes séronégatives (F). De ces mesures, Timæus et Nunn (1997) tirent un facteur correctif pour ajuster les proportions déclarées de répondants ayant leur mère encore en vie (${}_5S_x^*$) et obtenir les proportions qui seraient déclarées si les femmes séropositives avaient autant d'enfants que les autres femmes (${}_5S_x$) :

$${}_5S_x = \frac{1 - hP}{1 + \frac{1 - F}{F}P} {}_5S_x^*$$

Le facteur correctif est fondé sur l'idée que le biais dans la proportion déclarée de répondants ayant leur mère encore en vie s'accroît avec la prévalence du VIH à un taux déterminé par F et h . Comme on peut s'y attendre, plus le taux de transmission verticale est élevée et plus est faible la fécondité des femmes infectées, plus est important le biais dans les déclarations de la survie des mères. Quoique F , P et peut-

être h varient avec l'âge de la mère, la comparaison des résultats de cet ajustement brut avec ce qu'on obtient par des calculs détaillés par âge suggère que la précision de l'ajustement brut est suffisante (Timæus and Nunn 1997).

Si le traitement par les antirétroviraux ne bénéficie qu'à très peu, voire aucune des personnes infectées par le VIH, comme c'était le cas dans les populations africaines au xx^e siècle, la transmission verticale est généralement de l'ordre de 25 à 40 %, en moyenne un tiers (De Cock, Fowler, Mercier *et al.* 2000). Mais les programmes de prévention de la transmission mère-enfant (PTME) ont déjà réduit ce taux d'environ un quart dans les pays de forte prévalence et pourraient même l'amener à moins de 8 % dans les pays pauvres (Mahy, Stover, Kiragu *et al.* 2010). La réduction de la fécondité des femmes infectées est généralement d'environ 25 % dans les populations où peu d'adultes reçoivent des antirétroviraux, mais on ne sait pas bien comment cela pourrait évoluer à l'avenir (Chen and Walker 2010). Pour les déclarations des répondants nés avant la diffusion des programmes PTME, un ajustement adéquat pourrait être¹ :

$${}_5S_x = \frac{1 - \frac{P}{3}}{1 + \frac{P}{3}} {}_5S_x^*$$

Si la prévalence de l'infection VIH était à l'époque de 7 % parmi les femmes fréquentant des consultations prénatales, ceci devient

$${}_5S_x = \frac{1 - 0,0233}{1 + 0,0233} {}_5S_x^* = 0,954 {}_5S_x^*$$

La correction des proportions de répondants déclarant que leur père est encore en vie est plus difficile. L'ajustement à la baisse dépend de la proportion d'hommes infectés avec des partenaires infectées. Malheureusement, cette statistique est rarement connue et les observations disponibles suggèrent qu'elle varie fortement d'une population à l'autre, de sorte que l'utilisation d'une valeur moyenne donne des résultats plutôt imprécis (de Walque 2007 ; Freeman and Glynn 2004 ; Kaiser, Bunnell, Hightower *et al.* 2011). En outre,

1 Ce facteur correctif est un peu plus élevé que celui estimé par défaut par (Timæus and Nunn 1997) puisque, sur la base des données disponibles à l'époque, ceux-ci ont adopté comme valeurs des paramètres $F = 0,8$ et $h = 0,25$.

il n'a pas été démontré qu'un facteur correctif brut ignorant des différences des paramètres liées à l'âge fonctionne correctement pour les données relatives aux pères. Néanmoins, les proportions ajustées sont généralement moins biaisées que les proportions non ajustées dans les populations où l'épidémie est sévère. Si w est la proportion d'hommes avec des partenaires infectées, le facteur d'ajustement approprié est :

$${}_5S_x = (1 - (1 - (1 - h)F)(1 - w)P^*) {}_5S_x^*$$

où P^* est une estimation de la prévalence du VIH chez les hommes dans la population générale, par opposition à l'estimation chez les femmes, établie dans les consultations prénatales. Par exemple, pour $h = 1/3$, $F = 0,75$, $w = 0,5$ et $P^* = 0,05$ on obtient

$$\begin{aligned} {}_5S_x &= (1 - (1 - (1 - 0,3333)0,75)(1 - 0,5)0,05) {}_5S_x^* \\ &= 0,9875 {}_5S_x^* \end{aligned}$$

Pour les répondants âgés de moins de 15 ans, l'hypothèse que tous les parents infectés au moment de la naissance des répondants sont déjà décédés est irréaliste, même dans les données recueillies avant la mise en place de programmes de traitement de masse. Environ la moitié de ce groupe de parents est encore en vie chez les répondants de 5-9 ans et peut-être un quart chez les 10-14 ans. S'il en va ainsi, la correction (c'est-à-dire 1 moins le facteur correctif) faite sur les données des répondants de 5-9 ans doit être réduite de moitié et celle des répondants de 10-14 ans, d'un quart².

Coefficients révisés pour l'estimation de la mortalité féminine

Timæus et Nunn (1997) ont aussi développé une nouvelle série de coefficients de régression pour estimer les survivants de la table de mortalité à partir des proportions de répondants dont

2 Si l'on poursuit l'idée de Timæus et Nunn (1997), on arrive à la conclusion qu'il faudrait réduire l'ampleur de l'ajustement des données sur les deux premiers groupes d'âges en multipliant le facteur de correction initial par :

$$(1 + P(d - F)/F)/(1 - (-h)dP)$$

où d représente le rapport estimé des proportions des parents infectés à celles des parents non infectés encore vivants, selon leur statut HIV défini au moment de la naissance du répondant. L'effet de cette procédure diffère peu du résultat que l'on obtiendrait en réduisant simplement le complément du facteur de correction par la proportion estimée de parents encore en vie pour les répondants du groupe d'âges.

la mère est vivante, dans des populations subissant une épidémie sévère de VIH (voir tableau 22.7). Ces coefficients sont destinés à des populations où la prévalence de l'infection VIH chez les adultes atteint ou dépasse 5 %. Le modèle de régression est le même que dans le cas standard et les coefficients ne diffèrent que pour les répondants âgés de moins de 30 ans :

$${}_nP_{25} = a(n) + b(n)\bar{M}^f + c(n){}_5S_{n-5}$$

Des coefficients équivalents n'ont pas été développés pour l'estimation de la mortalité masculine, car on sait peu de choses sur les interrelations entre les schémas par âge de l'incidence du VIH et de la fécondité.

TABLEAU 22.7 Coefficients pour l'estimation de la survie des femmes à partir des proportions de répondants dont la mère survit dans des populations subissant une épidémie généralisée de VIH

| <i>n</i> | <i>a(n)</i> | <i>b(n)</i> | <i>c(n)</i> |
|----------|-------------|-------------|-------------|
| 10 | – 0,3611 | 0,00125 | 1,2974 |
| 15 | – 0,4030 | 0,00222 | 1,3732 |
| 20 | – 0,2120 | 0,00372 | 1,1342 |
| 25 | – 0,2389 | 0,00586 | 1,1131 |
| 30 | – 0,2513 | 0,00885 | 1,0223 |

Source : Timæus and Nunn (1997)

On en sait davantage aujourd'hui que dans les années 1990 sur les schémas par âge de l'incidence du VIH, sur la survie post-infection et sur le schéma par âge des réductions de la fécondité chez les personnes infectées. Les coefficients du tableau 22.7 pourraient sûrement être améliorés. Mais ils ne sont sans doute pas fortement biaisés. Ils sont néanmoins probablement inappropriés dans des populations où le traitement par les antirétroviraux s'est largement diffusé auprès des adultes infectés par le VIH.

Estimation du niveau et de l'évolution de la mortalité

Le défi le plus important dans l'estimation de la mortalité adulte à partir des proportions d'orphelins, dans des populations subissant une épidémie généralisée de sida, est l'absence de système adéquat de tables-types de mortalité reproduisant les schémas par âge de mortalité caractéristiques de séries de populations subissant des épidémies de sida à divers degrés de sévérité. Comme les décès dus au sida sont concentrés chez les adultes entre 25 et 55 ans environ, les variations du niveau de mortalité résultant de la diffusion du VIH modifient aussi radicalement le schéma par âge de la mortalité. Ceci réduit grande-

ment la confiance avec laquelle on peut passer de mesures de la survie dans différents intervalles d'âge à un indice commun de mortalité. De plus, comme l'incidence de l'infection par VIH dans la plupart des populations touchées s'est d'abord accrue rapidement, avant de diminuer aujourd'hui, il peut être approprié d'utiliser différents modèles pour différentes cohortes de répondants. Le fichier Excel (voir le site web) comprend des standards de mortalité élaborés pour des populations subissant une épidémie de VIH modérée. Mais ceux-ci ne seront pas adaptés ni aux pays où la prévalence de l'infection par VIH chez les adultes est seulement de quelques pourcents, ni aux pays où l'épidémie est sévère. Il se peut même que les standards soient mal adaptés aux pays où l'épidémie est modérée.

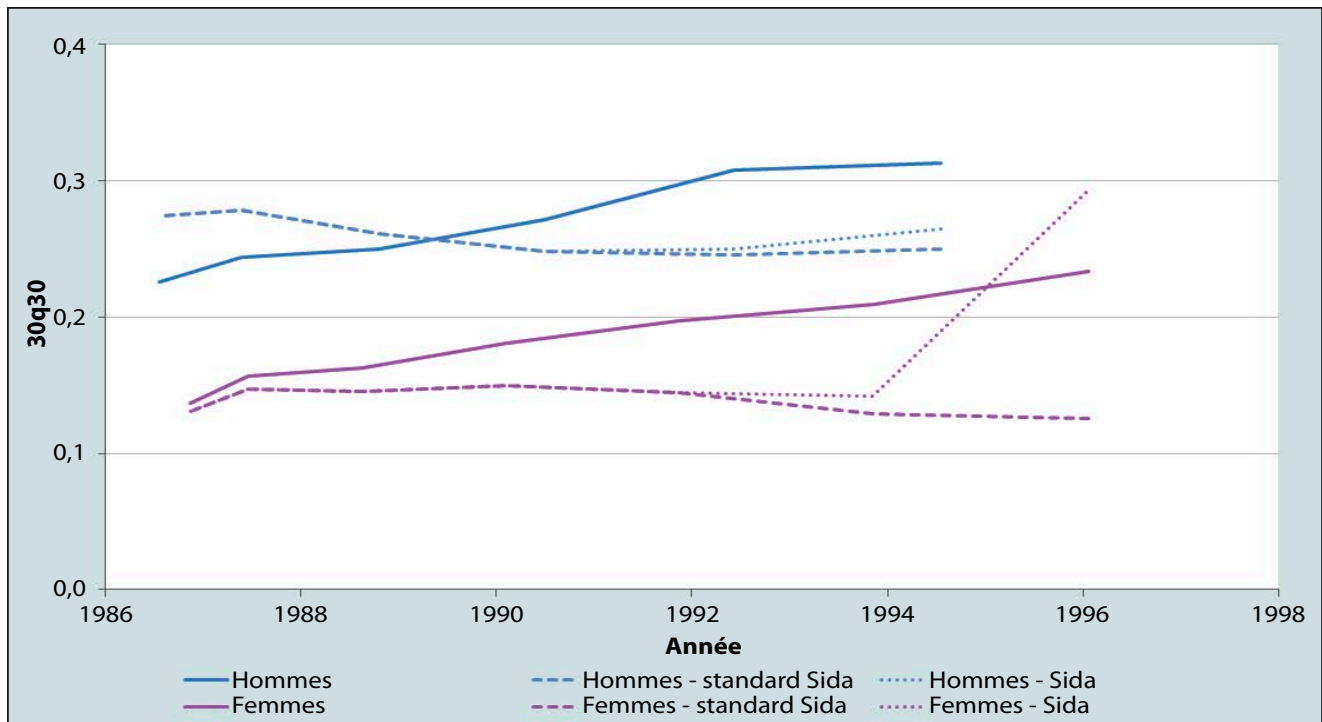
Illustration

Le problème que rencontre l'analyste est illustré par les estimations de la survie adulte faites pour le Kenya, à partir des données du recensement de 1999 sur les proportions d'orphelins représentées figure 22.2. Les estimations standards, faites sans recours à aucun des ajustements qu'on vient de présenter et traduites en un indice commun de mortalité grâce au réseau Général des tables-types de Nations Unies, sont représentées par la courbe en traits continus. Elles font apparaître une hausse continue, entre environ 1987 et 1996, aussi bien pour les hommes que les femmes, de la probabilité de décéder entre 30 et 60 ans (${}_{30}q_{30}$), une hausse d'environ 10 points de pourcentage dans chaque cas au cours de la décennie.

Les estimations dites « standard sida » ont été obtenues exactement de la même façon, sauf qu'elles ont été converties en estimations de ${}_{30}q_{30}$ grâce à des standards incorporant la mortalité par Sida. Les résultats sont très différents : la mortalité des hommes apparaît comme continuant à décroître lentement au Kenya jusqu'au début des années 1990, alors que le reste de la mortalité adulte stagne.

Les estimations intitulées « Sida » révèlent l'effet additionnel quand on ajuste les proportions pour tenir compte du biais de sélection dû au Sida et, pour l'estimation féminine la plus récente, quand on utilise les coefficients destinés aux populations où la prévalence de l'infection par VIH chez les adultes dépasse 5 %. Pour les répondants âgés de 5-9 ans, l'ajustement pour tenir compte du biais de sélection et l'utilisation des nouveaux coefficients entraînent tous les deux une hausse substantielle de l'estimation de la mortalité féminine. L'épidémie du VIH au Kenya dans les années 1990 n'était cependant pas assez sévère pour affecter les autres estimations.

FIGURE 22.2 Évolution de la probabilité de décéder entre le 30^e et le 60^e anniversaire estimée à partir des proportions d'orphelins, Kenya, recensement de 1999



Utiliser le réseau Général des tables-types des Nations Unies pour estimer $_{30}q_{30}$ au Kenya pour les années 1990, revient en effet à supposer que les hausses de la mortalité par Sida affectant les jeunes adultes se sont accompagnées de hausses comparables de la mortalité chez les individus d'âge moyen. Ceci semble peu vraisemblable. Les estimations de $_{30}q_{30}$ s'appuyant sur les standards Sida impliquent, au contraire, que toute hausse de la mortalité des jeunes adultes a été plus que compensée par un recul continu des taux de mortalité des adultes d'âge moyen, à la fin des années 1980, et compensée plus ou moins exactement par ces reculs, jusque vers 1993. Rares sont cependant les parents de répondants eux-mêmes dans la trentaine en 1999 à risquer d'avoir contracté le VIH, aussi l'utilisation du standard Sida pour faire les premières estimations de $_{30}q_{30}$ semble-t-elle également inappropriée. Le scénario peut-être le plus vraisemblable est que la probabilité de décéder entre 30 et 60 ans a stagné entre les fin des années 1980 et le milieu des années 1990 et a commencé ensuite à augmenter, mais sans doute moins abruptement que ne l'indique la courbe « Sida » pour les femmes.

Un point crucial est à souligner : à partir de ces données, il est impossible de déterminer exactement ce qui est arrivé à la mortalité adulte au Kenya pendant les années 1990. Les seules estimations qui soient suffisamment fiables sont relatives à la fin des années 1980. Les résultats sont fondés sur des groupes d'âge ne nécessitant qu'un ajustement limité des proportions des parents encore en vie afin d'estimer $_{30}q_{30}$. Les séries s'appuyant sur différents standards se recoupent à cette période.

Devenir orphelin avant ou après son mariage

Il existe des méthodes pour estimer la mortalité adulte à partir des proportions d'orphelins, quand des questions supplémentaires sont posées dans une enquête unique sur la position des décès des parents par rapport au premier mariage des répondants (Timæus 1991). Deux objectifs sous-tendent ces méthodes. Le premier est de fournir des méthodes où la localisation des résultats dans le temps est plus précise qu'avec les méthodes originelles. Le second est de développer des méthodes qui soient moins sujettes au biais de sous-déclaration des proportions d'orphelins aux

jeunes âges. Le mariage est un événement qui distingue, pour les répondants de chaque groupe d'âge, entre les décès parentaux récents et ceux survenus il y a longtemps. L'information sur le calendrier des décès est moins précise que celle résultant de questions directes sur la date de décès des parents, mais elle peut aussi être plus fiable. Même si les répondants ne peuvent pas se souvenir précisément de la date de décès de leurs parents, sans doute que presque tous sont capables de situer les décès de leurs parents par rapport à leur premier mariage, qui est un autre événement majeur de leur vie.

Quelque 15 enquêtes de la première phase du programme des Enquêtes démographiques et de santé ont recueilli des données sur la position des décès des parents par rapport au premier mariage. Malheureusement, très peu l'ont fait plus récemment. Des données sur la position des décès parentaux par rapport à la naissance du premier né des répondants pourraient être analysées exactement de la même manière et pourraient être plus représentatives de la mortalité de l'ensemble des parents, dans les populations où nombreux sont ceux qui ne se marient jamais ou lorsque le mariage est davantage considéré comme un processus que comme un événement survenant à une date bien définie.

Comme les hommes et les femmes se marient (et ont leur première naissance) à des âges différents, les réponses des hommes et des femmes interrogés doivent être analysées séparément. Les coefficients d'estimation présentés ici ont été développés en premier lieu pour l'analyse des données fournies par les réponses des femmes (Timæus 1991), mais ils peuvent aussi être utilisés pour l'analyse de données fournies par des hommes, si leur âge moyen au mariage est inférieur à 25 ans. Pour les répondants dont l'âge dépasse l'âge moyen au mariage, les estimations sont robustes vis-à-vis de toute caractéristique de la distribution des âges au mariage autre que sa moyenne.

Devenir orphelin après son premier mariage

La proportion de mères qui sont restées en vie depuis que les répondants sont mariés est étroitement reliée à la probabilité de survivre depuis un âge égal à la somme de l'âge moyen du moment à la naissance des enfants et l'âge moyen au mariage dans les cohortes jusqu'à un âge égal à la somme de l'âge moyen à la naissance des enfants et l'âge actuel des répondants. Les estimations faites à partir des données sur les proportions de répondants devenus orphelins depuis leur mariage mesurent une mortalité plus

récente que celles fondées les proportions de répondants devenus orphelins à un moment quelconque de leur vie. De plus, comme les décès parentaux survenus depuis le mariage ont eu lieu alors que les répondants étaient suffisamment âgés pour bien s'en souvenir, ces données sont moins sujettes à erreur de déclaration que celles portant sur la perte des parents en général.

L'âge central des répondants (n) le plus jeune pour lequel on puisse estimer un rapport de survie, ${}_n p_b$, à partir des données sur les proportions de répondants devenus orphelins depuis leur mariage est 30 ans. Pour les femmes, afin de préserver une relation étroite entre la survie des mères et la fonction de survie des tables de mortalité, cette dernière est mesurée à partir d'un âge de base, b , de 45 ans. Le modèle utilisé pour estimer les mesures des tables de mortalité à partir des proportions de mères survivantes depuis que les répondants sont mariés est donc

$${}_{n-20} p_{45} = a(n) + b(n)\bar{M}^f + c(n)\bar{m} + d(n){}_5 S_{n-5} / {}_5 S_{n-5}^m + e(n){}_5 S_n / {}_5 S_n^m$$

où \bar{m} est l'âge moyen au mariage dans les cohortes et ${}_5 S_n^m$ est la proportion de répondants dans le groupe d'âge n à $n+5$ dont la mère était en vie au moment de leur premier mariage.

La seule différence dans l'équation pour l'estimation de la mortalité masculine à partir des proportions de répondants devenus orphelins de père depuis leur mariage tient au fait que les hommes sont généralement plus âgés que les femmes à la naissance de leurs enfants. C'est pourquoi, la fonction de survie est estimée à partir d'un âge plus élevé de 10 ans. Les estimations sont faites à partir de

$${}_{n-20} p_{55} = a(n) + b(n)\bar{M}^m + c(n)\bar{m} + d(n){}_5 S_{n-5} / {}_5 S_{n-5}^m + e(n){}_5 S_n / {}_5 S_n^m$$

où l'âge au mariage continue d'être celui des répondants, qui sont généralement des femmes, pas celui des parents, qui sont cette fois des pères.

Jusqu'à ce que la plupart des répondants soient mariés, la relation entre les survivants de la table de mortalité et les proportions de parents encore en vie est sensible à la forme de la distribution des âges au premier mariage. Des coefficients existent donc pour estimer la survie sur des intervalles d'âge de ${}_{10} p_{45}$ à ${}_{30} p_{45}$ pour les femmes adultes et de ${}_{10} p_{55}$ à ${}_{20} p_{55}$ pour les hommes adultes (voir tableau 22.8).

TABLEAU 22.8 Coefficients pour l'estimation de la mortalité des hommes et des femmes adultes à partir des proportions de répondants devenus orphelins depuis leur premier mariage

| <i>n</i> | <i>a(n)</i> | <i>b(n)</i> | <i>c(n)</i> | <i>d(n)</i> | <i>e(n)</i> |
|----------------|-------------|-------------|-------------|-------------|-------------|
| Femmes adultes | | | | | |
| 30 | 0,5617 | 0,00836 | -0,00261 | -1,1231 | 1,4199 |
| 35 | 0,0476 | 0,01396 | -0,00536 | -0,3916 | 1,1354 |
| 40 | -0,3715 | 0,01966 | -0,00744 | 0,5394 | 0,5286 |
| 45 | -0,6562 | 0,02587 | -0,00716 | 1,0208 | 0,1789 |
| 50 | -0,8341 | 0,03045 | -0,00561 | 1,1898 | 0,0541 |
| Hommes adultes | | | | | |
| 30 | 0,0676 | 0,01588 | -0,00633 | -1,2070 | 1,8284 |
| 35 | -0,5459 | 0,02273 | -0,01083 | -0,2509 | 1,3867 |
| 40 | -0,8674 | 0,02622 | -0,01135 | 0,6057 | 0,7198 |

Source : Timæus (1991)

Devenir orphelin avant son premier mariage

La proportion de femmes dont la mère est vivante au moment de leur mariage est approximativement égale à la probabilité dans la table de mortalité de survivre entre l'âge moyen des mères à la naissance de leurs enfants et cet âge plus l'âge moyen de leurs filles à leur premier mariage. Si les données ne sont pas biaisées par l'effet d'adoption, cette variante de la méthode des proportions d'orphelins a deux caractéristiques intéressantes. Premièrement, elle mesure la mortalité sur un intervalle de temps et un éventail d'âges limités et assez clairement définis. Deuxièmement, elle permet d'étendre rétrospectivement la série temporelle d'estimations de la mortalité fournie par la méthode des proportions d'orphelins jusqu'au moins 30 ou 35 ans avant la collecte des données.

Dans cette variante, un terme d'interaction entre l'âge moyen au mariage et la proportion de parents décédés améliore l'ajustement du modèle. La probabilité de survivre des femmes entre 25 et 45 ans peut être estimée à partir des proportions de répondants devenus orphelins avant leur mariage grâce à l'équation de régression :

$${}_{20}p_{25} = a(n) + b(n)\bar{M}^f + c(n)\bar{m} + d(n){}_5S_n^m + e(n)\bar{m}_5S_n^m$$

Les coefficients de cette équation pour les différents groupes d'âge définis par *n* figurent au tableau 22.9.

Les mêmes considérations s'appliquent à l'estimation de la mortalité adulte masculine à partir des proportions de répondants devenus orphelins de père avant leur mariage.

TABLEAU 22.9 Coefficients pour l'estimation de la mortalité des hommes et des femmes adultes à partir des proportions de répondants devenus orphelins avant leur premier mariage

| <i>n</i> | <i>a(n)</i> | <i>b(n)</i> | <i>c(n)</i> | <i>d(n)</i> | <i>e(n)</i> |
|----------------|-------------|-------------|-------------|-------------|-------------|
| Femmes adultes | | | | | |
| 30 | -0,9607 | 0,00418 | 0,04466 | -0,04291 | 1,8178 |
| 35 | -0,9921 | 0,00429 | 0,04700 | -0,04501 | 1,8428 |
| 40 | -1,0129 | 0,00433 | 0,04822 | -0,04611 | 1,8607 |
| 45 | -1,0206 | 0,00434 | 0,04861 | -0,04648 | 1,8680 |
| Hommes adultes | | | | | |
| 30 | -1,2719 | 0,01060 | 0,04480 | -0,04007 | 1,8383 |
| 35 | -1,2977 | 0,01068 | 0,04652 | -0,04124 | 1,8530 |
| 40 | -1,3203 | 0,01070 | 0,04769 | -0,04225 | 1,8726 |
| 45 | -1,3232 | 0,01070 | 0,04783 | -0,04238 | 1,8753 |

Source : Timæus (1991)

Les estimations de la probabilité de survie entre 35 et 55 ans sont faites grâce à

$${}_{20}p_{35} = a(n) + b(n)\bar{M}^f + c(n)\bar{m} + d(n){}_5S_n^m + e(n)\bar{m}_5S_n^m$$

Ces coefficients figurent également au tableau 22.9.

Les séries finales de coefficients du tableau 22.9 pour les répondants devenus orphelins de mère et de père avant leur mariage se réfèrent à un groupe d'âge dans lequel à peu près tous les premiers mariages ont été conclus. Ces coefficients peuvent servir à estimer la mortalité à partir des déclarations des répondants de tous les groupes d'âge au-delà de 40 ans. Seule la qualité des déclarations sur les décès des parents et l'âge des répondants eux-mêmes impose un âge limite aux données qui peuvent être utilisées pour estimer la mortalité à partir des proportions de ceux qui sont devenus orphelins avant leur mariage.

Localisation dans le temps des estimations

Comme celles tirées des proportions d'orphelins en général, les estimations à partir des données fournies par les répondants des différentes cohortes, devenus orphelins avant et après le mariage, reflètent la mortalité sur des périodes de temps variables et mal définies. Pour l'avant mariage, les mesures de mortalité se réfèrent à une période qui est le produit de la distribution des intervalles entre le décès des parents et le mariage et de la distribution des intervalles entre le mariage et l'interview. Au degré de précision recherché ici, l'ancienneté de référence des estimations des proportions d'orphelins de mère est égale à l'intervalle moyen entre le décès de la mère

et le premier mariage plus l'intervalle moyen du premier mariage à l'interview. Pour les cohortes féminines presque toutes mariées, les âges auxquels les parents sont exposés au risque de décéder sont concentrés entre leur âge moyen à la naissance des enfants et la somme de cet âge et de l'âge moyen de leurs filles à leur premier mariage. L'ancienneté de référence des mesures de mortalité est égale à :

$$T = (N - \bar{m}) + \bar{m} g_{\bar{M}}$$

Le second terme à droite de l'équation 5 peut être estimé par la procédure déjà développée lors de la présentation de la méthode de base (voir l'étape 4). Comme pour les proportions d'orphelins en général, la date de référence des estimations de la mortalité masculine doit prendre en compte les éventuels décès entre la conception et la naissance et est donc calculée en utilisant $\bar{m} + 0,75 g_{\bar{M} - 0,75}$.

Pour les répondants devenus orphelins après leur mariage, l'âge auquel les parents débutent leur exposition au risque de décéder est le produit des distributions de leurs âges à la naissance de leurs enfants et des âges de leurs filles au mariage. Pour les groupes d'âge ayant à peu près complété leur nuptialité, on peut estimer l'âge moyen auquel débute

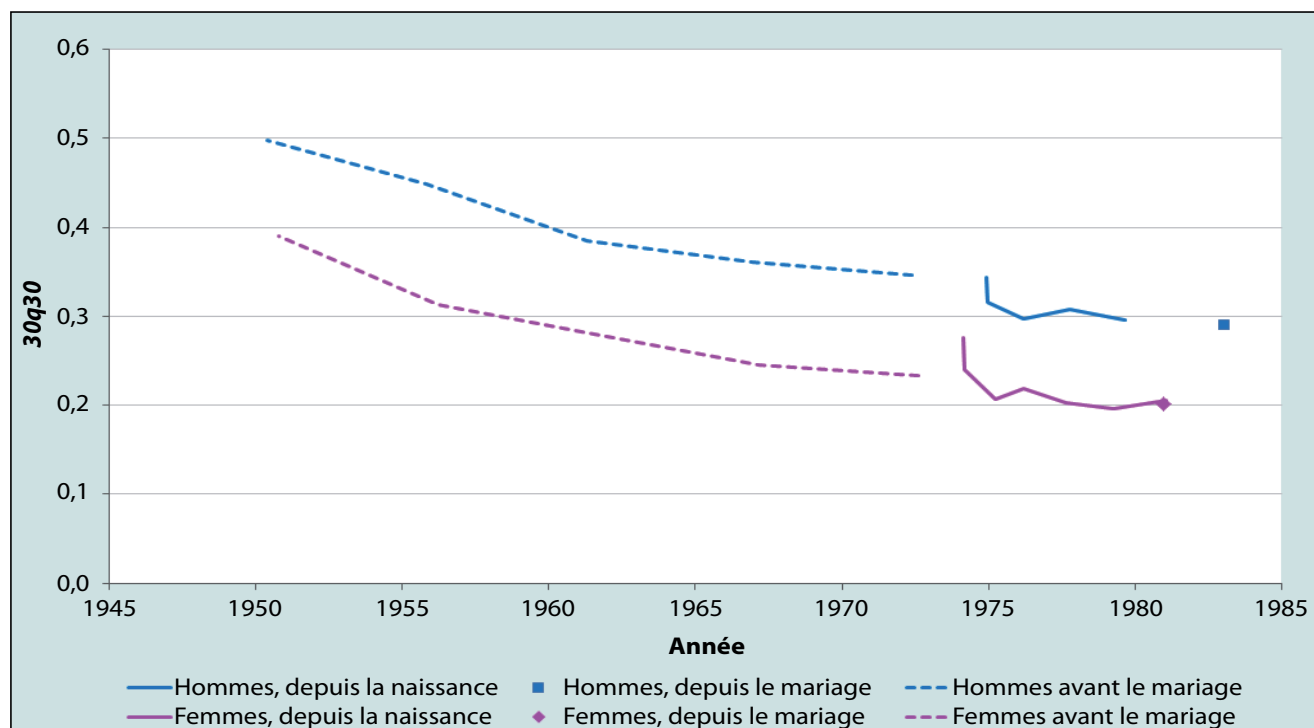
l'exposition au risque comme la somme des moyennes de ces deux distributions. L'exposition des parents se prolonge jusqu'au moment de l'interview, $N - \bar{m}$ années plus tard. Les anciennetés de référence des mesures de la mortalité sont donc $T = N - \bar{m} g_{\bar{M} + \bar{m}}$.

Comme les âges auxquels les pères sont exposés au risque de décès débutent bien après la naissance de leurs filles, cette équation vaut à la fois pour les estimations de la mortalité des hommes et des femmes.

Illustration

La figure 22.3 illustre les résultats de l'application de la méthode des proportions d'orphelins avant et après leur mariage pour l'estimation de la mortalité adulte aux données recueillies auprès des femmes de 15 à 49 ans dans l'Enquête démographique et de santé (EDS) de l'Égypte en 1988. Ce qui frappe d'abord dans ces résultats, c'est que l'information relative aux répondantes âgées devenues orphelines avant leur premier mariage, recueillie en 1988, permet des estimations de la mortalité adulte qui remonte au début des années 1950. Deuxièmement, les séries supplémentaires d'estimations obtenues en distinguant les femmes orphelines, entre celles

FIGURE 22.3 Évolution de la probabilité de décéder entre le 30^e et le 60^e anniversaire estimée à partir des proportions d'orphelins, Égypte, EDS 1988



qui l'étaient avant leur premier mariage et celles qui le sont devenues depuis, concordent bien avec celles obtenues grâce aux données sur les proportions d'orphelins en général. Ces résultats ne révèlent pas de biais vers le bas résultant d'une sous-déclaration des décès parentaux due à l'effet d'adoption, ni dans les estimations les plus récentes fondées sur les proportions d'orphelins en général, ni dans celles fondées sur les proportions d'orphelins avant leur premier mariage.

D'après ces résultats, il s'est produit une baisse continue, graduellement ralentie, de la mortalité adulte, en Égypte du début des années 1950 au milieu des années 1980. La probabilité de décéder entre le 30^e et le 60^e anniversaire a reculé d'environ 200 pour mille au cours de cette période, pour les hommes comme pour les femmes, à partir d'un niveau très élevé au début des années 1950 jusque vers 200 pour mille pour les femmes et 300 pour mille pour les hommes au début des années 1980. Mêmes ces estimations les plus récentes représentent une mortalité plutôt élevée. L'écart de 100 pour mille entre les probabilités des décéder masculines et féminines n'a guère varié tout au long de la période. Les estimations par les proportions d'orphelins de la mortalité adulte des

hommes, en particulier, sont nettement plus élevées que celles fondées sur d'autres sources (UN Population Division 2011).

Questions sur le calendrier des décès parentaux

Une autre extension de la méthode des proportions d'orphelins consiste à demander aux répondants qui ont perdu leurs parents quand est survenu le décès (Chackiel and Orellana 1985). Si les dates de décès des parents sont déclarées avec assez de précision, cela permet à l'analyste de distinguer entre les décès parentaux récents et les plus anciens et d'obtenir des estimations de la mortalité mieux à jour. La meilleure façon d'analyser ces données consiste à utiliser l'information sur les date de décès pour reconstruire la proportion de répondants qui avaient encore leurs parents cinq ans et dix ans plus tôt. De ces coupes transversales successives, on peut tirer des mesures de la survie des parents dans des cohortes synthétiques formellement identiques à celles fournies par les données collectées dans une série d'enquêtes séparées. C'est pourquoi ces méthodes d'analyse sont présentées plus loin, en même temps que l'analyse des proportions d'orphelins tirées d'enquêtes multiples.

AUTRES LECTURES ET RÉFÉRENCES

La méthode des proportions d'orphelins est présentée dans tous les manuels classiques d'estimation indirecte (Sloggett, Brass, Eldridge *et al.* 1994 ; Division de la Population des NU 1984). Mais, à l'exception de celui des Nations unies sur l'estimation de la mortalité adulte (UN Population Division 2002), ces manuels s'attachent à l'ancienne variante de la méthode, qui recourt à des facteurs de pondération pour obtenir des indices de tables de mortalité, plutôt qu'à la méthode fondée sur la régression, couramment utilisée aujourd'hui. Les méthodes fondées sur la régression ont été proposées précédemment pour les femmes (Hill and Trussell 1977 ; Palloni and Heligman 1985), mais elles ont été développées pour les hommes plus récemment par Timæus (1992). Ce dernier article passe aussi en revue des contributions plus anciennes sur le sujet et présente les bases théoriques de la méthode.

Les méthodes d'estimation de la mortalité adulte à partir des proportions de répondants devenus orphelins avant et après leur mariage ont été présentées par Timæus (1991). Les procédures d'estimation de la mortalité féminine dans les populations subissant une épidémie généralisée de VIH ont été développées par Timæus and Nunn (1997).

Luy (2012) a proposé une méthode modifiée des proportions d'orphelins pour étudier les différentiels socioéconomiques de mortalité adulte, dans les pays qui disposent déjà d'une information agrégée précise sur la mortalité adulte. Sa méthode consiste à déterminer la relation entre la proportion de répondants à l'enquête qui déclarent des parents encore en vie et les survivants de la table de mortalité, à partir de statistiques empiriques sur la population concernée, plutôt que de recourir à des modèles démographiques. La méthode de Luy adapte donc le processus d'estimation en fonction du contexte. Rares sont les pays à revenu bas ou moyen qui disposent de données nationales suffisamment précises pour que la démarche puisse s'appliquer, bien que certains aspects puissent se révéler pertinents dans certaines circonstances.

Les méthodes qui exploitent les possibilités analytiques supplémentaires, lorsque des questions sur la survie des parents ont été posées dans au moins deux enquêtes successives auprès de la même population, sont présentées dans le chapitre 26.

Blacker JGC. 1984. "Experiences in the use of special mortality questions in multi-purpose surveys: the single-round approach," in *Data Bases for Mortality Measurement*. New York: United

- Nations, pp. 79-89. http://www.un.org/esa/population/publications/UN_1984_Data_Bases_for_Mortality_Measurement/UN1984_Data_Bases.htm
- Blacker JGC and JM Gapere. 1988. "The indirect measurement of adult mortality in Africa: results and prospects," in *African Population Conference, Dakar, 1988*. Liège: International Union for the Scientific Study of Population, Vol. 2:3.2.23-38.
- Brass W. 1971. "On the scale of mortality," in Brass, W (ed). *Biological Aspects of Demography*. London: Taylor and Francis, pp. 69-110.
- Brass W. 1975. *Methods for Estimating Fertility and Mortality from Limited and Defective Data*. Chapel Hill: International Program of Laboratories for Population Statistics.
- Brass W. 1985. *Advances in Methods for Estimating Fertility and Mortality from Limited and Defective Data*. London: London School of Hygiene & Tropical Medicine.
- Brass W and EA Bamgboye. 1981. *The Time Location of Reports of Survivorship: Estimates for Maternal and Paternal Orphanhood and the Ever-widowed*. London: London School of Hygiene & Tropical Medicine.
- Brass W and K Hill. 1973. "Estimating adult mortality from orphanhood," in *International Population Conference, Liège, 1973*. Liège: International Union for the Scientific Study of Population, Vol. 3:111-123.
- Chackiel J and H Orellana. 1985. "Adult female mortality trends from retrospective questions about maternal orphanhood included in censuses and surveys," in *International Population Conference, Florence, 1985*. Liège: International Union for the Scientific Study of Population, Vol. 4:39-51.
- Chen W-J and N Walker. 2010. "Fertility of HIV-infected women: insights from Demographic and Health Surveys", *Sexually Transmitted Infections* **86**(Suppl 2):ii22-ii27. doi: <http://dx.doi.org/10.1136/sti.2010.043620>
- De Cock KM, MG Fowler, E Mercier, I de Vincenzi *et al.* 2000. "Prevention of Mother-to-Child HIV Transmission in Resource-Poor Countries", *Journal of the American Medical Association* **283**(9):1175-1182. doi: <http://dx.doi.org/10.1001/jama.283.9.1175>
- de Walque D. 2007. "Sero-discordant couples in five African countries: implications for prevention strategies", *Population and Development Review* **33**(3):501-523. doi: <http://dx.doi.org/10.1111/j.1728-4457.2007.00182.x>
- Division de la Population des NU. 1984. *Manuel X. Techniques indirectes d'estimation démographique*. New York : Nations Unies, Département des affaires économiques et sociales internationales, ST/ESA/SER.A/81. http://unstats.un.org/unsd/demographic/standmeth/handbooks/Manuel_X-fr.pdf
- Freeman EE and JR Glynn. 2004. "Factors affecting HIV concordancy in married couples in four African cities", *AIDS* **18**(12):1715-1721. doi: <http://dx.doi.org/10.1097/01.aids.0000139075.13906.2f>
- Henry L. 1960. «Mesure indirecte de la mortalité des adultes», *Population* **15**:457-466.
- Hill K. 1984. "An evaluation of indirect methods for estimating mortality," in Vallin, J, Pollard John H and L Heligman (eds). *Methodologies for the Collection and Analysis of Mortality Data*. Liège: Ordina, pp. 145-176.
- Hill K and TJ Trussell. 1977. "Further developments in indirect mortality estimation", *Population Studies* **31**(2):313-334. doi: <http://dx.doi.org/10.2307/2173920>
- Kaiser R, R Bunnell, A Hightower, AA Kim *et al.* 2011. "Factors Associated with HIV Infection in Married or Cohabiting Couples in Kenya: Results from a Nationally Representative Study", *PLoS ONE* **6**(3):e17842. doi: <http://dx.doi.org/10.1371/journal.pone.0017842>
- Luy M. 2012. "Estimating mortality differences in developed countries from survey information on maternal and paternal orphanhood", *Demography* **49**(2):607-627. doi: <http://dx.doi.org/10.1007/s13524-012-0101-4>
- Mahy M, J Stover, K Kiragu, C Hayashi *et al.* 2010. "What will it take to achieve virtual elimination of mother-to-child transmission of HIV? An assessment of current progress and future needs", *Sexually Transmitted Infections* **86**(Suppl 2):ii48-ii55. doi: <http://dx.doi.org/10.1136/sti.2010.045989>
- Palloni A and L Heligman. 1985. "Re-estimation of structural parameters to obtain estimates of mortality in developing countries", *Population Bulletin of The United Nations* **18**:10-33.
- Palloni A, M Massagli and J Marcotte. 1984. "Estimating adult mortality with maternal orphanhood data: analysis of sensitivity of the techniques", *Population Studies* **38**(2):255-279. doi: <http://dx.doi.org/10.1080/00324728.1984.10410289>
- Sloggett A, W Brass, SM Eldridge, IM Timæus, P Ward and B Zaba. 1994. *Estimation of Demographic Parameters from Census Data*. Tokyo, Japan: United Nations Statistical Institute for Asia and the Pacific.
- Timæus I. 1986. "An assessment of methods for estimating adult mortality from two sets of data on maternal orphanhood", *Demography* **23**(3):435-450. doi: <http://dx.doi.org/10.2307/2061440>
- Timæus IM. 1991. "Estimation of adult mortality from orphanhood before and since marriage", *Population Studies* **45**(3):455-472. doi: <http://dx.doi.org/10.1080/0032472031000145636>
- Timæus IM. 1992. "Estimation of adult mortality from paternal orphanhood: a reassessment and a new approach", *Population Bulletin of The United Nations* **33**:47-63.
- Timæus IM and AJ Nunn. 1997. "Measurement of adult mortality in populations affected by AIDS: an assessment of the orphanhood method", *Health Transition Review* **7**(Suppl 2):23-43. <http://htc.anu.edu.au/pdfs/Timaeus2.pdf>
- UN Population Division. 2002. *Methods for Estimating Adult Mortality*. New York: United Nations, Department of Economic and Social Affairs, ESA/P/WP.175. http://www.un.org/esa/population/techcoop/DemEst/methods_adultmort/methods_adultmort.html
- UN Population Division. 2011. *World Mortality Report 2007*. New York: United Nations, Department of Economic and Social Affairs, ST/ESA/SER.A/289. <http://www.un.org/esa/population/publications/worldmortalityreport2007/WORLD%20MORTALITY%20REPORT.PDF>

Chapitre 23 ESTIMATION INDIRECTE DE LA MORTALITÉ ADULTE À PARTIR DES DONNÉES SUR LES FRATRIES

Ian M Timæus

DESCRIPTION DE LA MÉTHODE

Cette méthode estime la mortalité adulte indirectement à partir des données fournies par les adultes sur la survie de leurs frères et sœurs adultes (c'est-à-dire leur fratrie). Ces données sont classées par groupe d'âge des répondants. La mortalité peut ainsi être estimée sans que les répondants doivent se rappeler les dates où les décès sont survenus ou les âges au décès des personnes décédées. L'information sur la survie des frères est utilisée pour estimer la mortalité des hommes, celle sur les sœurs pour estimer la mortalité des femmes.

Les répondants sont souvent incapables de déclarer – et peuvent ne rien savoir sur – leurs frères et sœurs décédés avant ou peu après leur propre naissance. Mais on réduit largement l'impact de ce biais en n'incluant dans l'analyse que la fratrie ayant survécu jusqu'à 15 ans. Pour que la méthode puisse être appliquée, il faut qu'un recensement ou une enquête ait demandé aux répondants adultes (par exemple ceux âgés de 15 à 49 ans) combien de leurs frères et sœurs ont survécu jusqu'à 15 ans et combien d'entre eux sont encore en vie. De nombreuses enquêtes ne recueillent l'information qu'après des femmes, mais les données fournies par les hommes peuvent être analysées exactement selon les mêmes méthodes.

Les répondants ont des frères et sœurs approximativement du même âge qu'eux en moyenne. La proportion des frères et sœurs ayant survécu jusqu'à 15 ans qui sont encore en vie à la date d'enquête est un bon estimateur de la probabilité conditionnelle de la table de mortalité de survivre de 15 ans jusqu'à l'âge actuel des répondants.

Si la mortalité a évolué au fil du temps, les rapports de survie estimés reflètent des taux de mortalité qui ont affecté chaque cohorte à divers âges et diverses périodes. Une méthode de « localisation dans le temps » a été développée pour estimer combien d'années avant le recensement ou l'enquête chaque

proportion de survivants dans les cohortes égale la proportion de survivants par période. Ces intervalles s'accroissent avec l'âge des répondants, s'étageant entre 3 et 13 ans avant la collecte des données. Ainsi, si les rapports de survie estimés à partir des déclarations des répondants des différents groupes d'âge sont traduits en indice courant de la mortalité des adultes, grâce à un système de tables types de mortalité à 1 paramètre, ces statistiques se rapporteront à diverses dates et pourront être utilisées pour retracer l'évolution de la mortalité dans le temps.

Les méthodes des fratries ont un avantage sur les questions concernant les décès survenus dans le ménage, qui nécessitent des recensements ou des enquêtes exceptionnellement larges pour recueillir des informations sur un nombre suffisant de décès dans les ménages dans l'année précédant l'enquête, afin que les estimations de la mortalité soient assez précises pour être utiles. Comme les réponses des enquêtés portent en moyenne sur plusieurs frères et sœurs, et que les estimations sont fondées sur l'ensemble de l'exposition au risque au-delà de 15 ans, la méthode des fratries peut être utilisée dans des enquêtes relativement petites, même si toutes les méthodes d'estimation de la mortalité adulte nécessitent des données sur des milliers de ménages. En outre, il n'est pas indispensable que la population soit fermée à la migration. Les résultats ne seront toutefois pas représentatifs concernant des petits États ou des régions dans lesquels une proportion substantielle de la population aura émigré ou sera composée d'immigrés.

DONNÉES NÉCESSAIRES ET HYPOTHÈSES

Tableaux de données nécessaires

Pour l'estimation de la mortalité des femmes adultes, les répondants âgés de 15 à 49 ans doivent être interrogés sur

le nombre de leurs sœurs ayant vécu jusqu'à 15 ans et le nombre de celles-ci encore en vie. À partir des tableaux des réponses à ces deux questions par groupe d'âge des répondants, on peut calculer :

- La proportion de sœurs encore en vie parmi celles ayant vécu au moins jusqu'à leur 15^e anniversaire, par groupe quinquennal d'âge des répondants. (Ceux qui n'ont pas répondu aux deux questions doivent être exclus du calcul.)

Pour l'estimation de la mortalité des hommes adultes, les répondants âgés de 15 à 49 ans doivent être interrogés sur le nombre de leurs frères ayant vécu jusqu'à 15 ans et le nombre de ceux-ci encore en vie. À partir des tableaux des réponses à ces deux questions par groupe d'âge des répondants, on peut calculer :

- La proportion de frères encore en vie parmi ceux ayant vécu au moins jusqu'à leur 15^e anniversaire, par groupe quinquennal d'âge des répondants. (Ceux qui n'ont pas répondu aux deux questions doivent être exclus du calcul.)

Les tableaux des nombres de frères et sœurs atteignant 15 ans et encore en vie doivent *exclure* le répondant en personne. (C'est évidemment toujours le cas quand les répondants sont interrogés sur leur fratrie de sexe opposé). Cette nécessité sera expliquée dans la discussion sur les hypothèses importantes faites par la méthode.

Les tableaux sur la fratrie de même sexe que le répondant (c'est-à-dire les sœurs des femmes et les frères des hommes) doivent être pondérés seulement par les pondérations d'échantillonnage fournies avec les données. Les tableaux sur la fratrie de sexe opposé (c'est-à-dire les frères des femmes et les sœurs des hommes) doivent en outre être pondérés par l'inverse de la taille de la fratrie survivante de même sexe que le répondant. Cette nécessité sera aussi expliquée dans la discussion sur les hypothèses importantes faites par la méthode.

Pour éliminer les ambiguïtés résultant des mariages polygames et des remariages, il est généralement indiqué aux enquêteurs que la « fratrie » concerne les enfants nés de la même mère. Que cela ait été fait ou non, les réponses peuvent généralement être acceptées telles qu'elles sont. Aussi longtemps que les répondants ont le même groupe de parents à l'esprit quand ils répondent aux deux questions sur le nombre de frères et sœurs en vie à 15 ans, d'une part, et encore en vie aujourd'hui, d'autre part, qui sont leurs parents importe peu.

Si on a posé les bonnes questions à la fois aux hommes et aux femmes, les réponses doivent généralement être tabulées

séparément de sorte que les deux séries de données fassent l'objet d'un contrôle mutuel.

Des nombreuses Enquêtes démographiques et de santé recueillent des histoires complètes de fratrie auprès des femmes de 15-49 ans. Il est demandé à chaque répondante le nom, le sexe, l'âge de chacun de ses frères et sœurs nés de la même mère, ainsi que son statut de survie et, s'il est décédé, son âge au décès et l'année du décès. Les taux de mortalité toutes causes des hommes et des femmes doivent généralement être calculés à partir de ces historiques par la méthode directe de la fratrie. On pense que la déclaration des décès récents dans la fratrie est plus complète que celle des décès plus anciens et la méthode directe permet de limiter l'analyse aux données des années qui ont précédé immédiatement l'enquête. Il est cependant facile d'effectuer, à partir des historiques complets, les décomptes sommaires que nécessite la méthode indirecte. Si la déclaration des frères et sœurs est à peu près complète, mais que les déclarations des âges au décès et des dates de décès sont très imparfaites, la procédure indirecte peut donner des résultats plus fiables que la procédure directe.

Hypothèses importantes

Une limitation des méthodes de mesure de la mortalité adulte fondées sur les fratries est qu'elles sous-estiment la mortalité dans la mesure où celle-ci forme des grappes au sein des fratries (c'est-à-dire des ensembles de frères et/ou de sœurs nés de la même mère). Ce regroupement en grappes survient chaque fois que les décès sont plus concentrés dans une petite fraction des fratries que sous le seul effet du hasard et il résulte de l'hétérogénéité entre fratries du risque de décès des individus (Zaba and David 1996). Il en résulte un biais vers le bas des estimations de mortalité, simplement parce que les membres des fratries à forte mortalité sont moins nombreux que les membres de fratries à faible mortalité à survivre pour répondre aux questions sur leurs frères et sœurs. Ce biais ne peut pas être complètement corrigé car, à la limite, les ensembles de fratries dont les membres sont tous décédés ne sont déclarés par personne. Il n'est pas possible de savoir combien de ces fratries ont existé ou quelles étaient leurs tailles.

Les estimations de l'évolution de la mortalité seront en conséquence biaisées si l'ampleur des regroupements en grappes parmi les fratries varie avec l'âge. Par exemple, si des caractéristiques communes aux frères et sœurs (par exemple, les facteurs génétiques, le vécu de l'enfance, le statut socioé-

conomique, les styles de vie et la localisation) influencent fortement la mortalité des adultes d'âge moyen, alors que la mortalité avant 40 ans a une large composante aléatoire, les estimations faites pour les répondants âgés sous-estimeront la mortalité davantage que les données fournies par les répondants plus jeunes, créant ainsi une impression fallacieuse d'augmentation de la mortalité au fil du temps.

La question du biais résultant des réponses multiples aux questions sur les fratries a donné lieu à une abondante littérature. Le problème existe aussi bien dans les données d'enquête que de recensement, car plus un individu est cité dans un recensement, plus il a de chances d'avoir un frère ou une sœur *qui le cite dans un échantillon probabiliste*.¹ En outre, même dans les enquêtes, il peut arriver que des réponses multiples soient données concernant un même individu. Par exemple, si deux sœurs de la même mère sont interrogées dans le même ménage, il y aura des déclarations multiples concernant les autres membres de la fratrie. La procédure d'analyse standard utilisée, par exemple, dans les rapports EDS s'appuie sur les événements et les durées d'exposition des frères et sœurs cités, en laissant de côté la durée d'exposition de la répondante (survivante) elle-même. Les événements et la durée d'exposition sont pondérés seulement par les pondérations d'échantillonnage du répondant, sans prendre en compte les nombres de répondants survivants potentiels dans la fratrie.

Trussell et Rodriguez (1990) ont démontré mathématiquement que, pour des groupes de fratries de même sexe ayant un risque sous-jacent identique de décéder, le calcul standard qui exclut aussi le répondant du dénominateur des mesures donne des estimations non biaisées de la mortalité. En effet, la réduction du nombre de citations des personnes décédées au numérateur, du fait que les personnes décédées ne peuvent pas se citer l'une l'autre, et l'exclusion des répondants vivants du dénominateur se compensent mutuellement et donnent le risque correct concernant les fratries en tant que groupe.

La question des biais qui pourraient résulter de différences de mortalité en fonction de la taille de la fratrie est liée à la question de la citation multiple. Elle a suscité

beaucoup d'intérêt chez les chercheurs car, à la différence d'autres facteurs qui jouent sur le risque au sein des fratries classées par sexe et âge du répondant, la taille de la fratrie de chaque répondant est connue. Si la mortalité ne varie pas avec la taille de la fratrie, les estimations standard sont les mêmes à la fois pour chaque taille de fratrie, y compris les fratries d'une personne qui sont exclues de l'analyse parce que le répondant n'a personne à citer, et pour la population dans son ensemble. Même si la mortalité varie avec la taille de la fratrie, les estimations standard restent sans biais pour chaque taille de la fratrie, comme l'a montré Masquelier (2013). Mais pour obtenir des estimations de la mortalité pour la population, il convient de repondérer les estimations relatives aux fratries de différentes tailles par la prévalence de celles-ci dans la population. Quand les répondants citent les personnes de même sexe, on peut le faire en divisant la proportion de frères ou sœurs survivant depuis 15 ans jusqu'au groupe d'âge à l'enquête par l'ensemble des fratries de même taille. Pour les fratries d'une seule personne, leur mortalité doit être extrapolée à partir de la mortalité dans les fratries plus grandes.

Gakidou et King (2006) soutiennent que, à l'inverse de la démarche standard, les proportions de décédés doivent être estimées pour des fratries incluant le répondant survivant, mais doivent toujours être pondérées en plus par la probabilité qu'elles soient citées – c'est-à-dire par l'inverse du nombre de répondants potentiels qui survivent dans la fratrie. Comme dans la démarche de Masquelier, un ajustement supplémentaire doit aussi être fait pour les fratries non déclarées du fait du décès de tous leurs membres. Dans une analyse des historiques complets de fratries tirés d'enquêtes EDS, Obermeyer, Rajaratnam, Park *et al.* (2010) estiment que l'absence d'ajustement tenant compte de la probabilité de citation peut biaiser de – 20 % les estimations de la mortalité d'ensemble. Toutefois, Masquelier (2013) prétend que Obermeyer *et alii* ont mal repondéré les données et qu'il en est résulté une surestimation du biais. Il souligne que, si on doit repondérer, il importe d'ajuster seulement pour tenir compte des citations multiples par les frères et sœurs qui ont survécu à l'âge initial à partir duquel la mortalité est mesurée. Mais il s'interroge en outre sur la réalité de la variation de la mortalité par taille de la fratrie. Il pourrait s'agir d'un artefact dû à l'omission plus fréquente des frères et sœurs décédés dans les historiques relatifs à des fratries plus grandes. C'est pourquoi Masquelier recommande d'utiliser l'approche standard, sans essayer de repondérer les données

¹ La question discutée ici est celle du biais dans l'estimation centrale de la mortalité. Elle ne doit pas être confondue avec la question du calcul correct de l'erreur d'échantillonnage des estimations de mortalité, dans les enquêtes où certains individus sont cités plus d'une fois et où la mortalité constitue des grappes parmi les fratries.

pour tenir compte de la mortalité différentielle par taille de la fratrie. C'est la démarche que nous adopterons ici.

Les questions sont différentes quand on analyse les citations concernant le sexe opposé (par exemple, les réponses faites par des femmes au sujet de leurs frères). Dans ce cas, le répondant ne fait pas partie du groupe exposé au risque de décéder. Mais le calcul standard donnera encore des résultats biaisés pour l'ensemble de la population si la mortalité des frères est associée au nombre de sœurs, dans l'exemple où c'est l'une d'entre elles qui est interrogée. Concernant les citations par le sexe opposé, il est donc clair qu'il faut pondérer chaque citation par l'inverse du nombre de frères ou sœurs du même sexe que le répondant, comme suggéré par Gakidou and King (2006). Les questions sur la fratrie de sexe opposé ne peuvent évidemment donner aucune information sur les fratries dont les membres vivants n'incluent personne de même sexe que le répondant. L'adoption de cette démarche revient donc à supposer que la mortalité des individus dans de telles fratries est égale à la mortalité du reste de la population. Mais dans les enquêtes qui recueillent des données auprès des deux sexes, chacun d'entre eux fournit cette information sur l'autre, de sorte qu'on peut pondérer les décès et l'exposition au risque déclarés par les répondants par l'inverse de la probabilité que les frères et sœurs dans chaque groupe d'âge soient cités.

La méthode des frères et sœurs adultes permet d'estimer l'évolution de la mortalité à partir des données fournies par les répondants de différents groupes d'âge : plus le répondant est âgé, plus les décès de ses frères et sœurs sont anciens en moyenne. Pour convertir les séries de mesures de la survie obtenues indirectement à partir des données des différents groupes d'âge en un indicateur unique qui puisse être comparé au fil du temps, il faut supposer que le profil de la mortalité adulte par âge est représenté par la table de mortalité standard choisie. Pour estimer la localisation dans le temps de ces mesures, il est en outre supposé que la mortalité a reculé linéairement, en termes de ce standard, au long de la période considérée.

TRAVAUX PRÉPARATOIRES ET RECHERCHES PRÉLIMINAIRES

Avant de débiter l'analyse, il faut vérifier le nombre de répondants ayant déclaré ne pas savoir combien de leurs frères et sœurs avaient atteint 15 ans ou combien d'entre eux étaient encore en vie, ou n'ayant pas répondu du tout aux questions. Le taux de réponse à ces questions est généra-

lement élevé, mais quelques enquêtes ont recueilli des données assez incomplètes pour qu'un biais dû aux non réponses puisse créer un problème important.

Si les questions ont été posées aussi bien aux hommes qu'aux femmes, un contrôle utile de la qualité des données sur la fratrie consiste à établir combien de frères et sœurs sont déclarés, en moyenne, par les répondants de l'autre sexe et à vérifier si le rapport de masculinité à la naissance varie fortement à mesure que l'âge des répondants augmente.

On doit aussi comparer les proportions de frères et sœurs décédés dans les déclarations des répondants hommes et femmes du même âge. La mortalité des individus d'un sexe telle que déclarée par leurs frères doit être égale à la mortalité des mêmes individus telle que déclarée par leurs sœurs. Si ce n'est pas le cas, ce peut être le signe d'un biais important dans les proportions de décédés de l'un ou des deux sexes, du fait de la citation multiple de certaines fratries et du fait qu'aucun des frères et sœurs n'a survécu pour déclarer les autres. Par ailleurs, si les proportions déclarées par les répondants hommes et femmes divergent à mesure que leur âge augmente, ce peut être dû à des différences de genre dans les erreurs de déclaration des âges ou au fait que le sexe qui déclare moins de frères ou sœurs décédés (généralement les hommes) risque davantage d'avoir perdu contact avec sa famille d'origine et tend à supposer à tort que certains de ses frères et sœurs décédés sont encore en vie.

PRÉCAUTIONS ET MISES EN GARDE

- La méthode indirecte de la fratrie a été initialement développée pour estimer la survie des frères et sœurs depuis la naissance (Hill and Trussell 1977). Malheureusement, ces déclarations sont souvent très incomplètes, en particulier pour les frères et sœurs décédés avant ou peu après la naissance du répondant. À la place, on recommande la méthode indirecte de la fratrie adulte, mais elle ne peut fournir que des probabilités de survie de 15 ans à des âges plus avancés de la vie adulte, sous condition que les personnes aient été en vie à leur 15^e anniversaire. Pour établir une table de mortalité complète, il faut estimer la survie de la naissance à 15 ans, en utilisant des estimations fondées sur une autre source de données.
- Les décès des frères et sœurs n'ont pas lieu en un point du temps mais peuvent être survenus à tout moment entre le 15^e anniversaire des répondants et la date de leur interview. C'est pourquoi les applications de la méthode indirecte de la fratrie ne peuvent indiquer qu'une évolution

lissée de la mortalité adulte et ne sauraient saisir les crises de mortalité de court terme ou les changements brutaux de tendance de la mortalité tels que ceux dus au Sida après le déclenchement d'une épidémie généralisée de VIH.

- Les estimations de mortalité les plus récentes qu'on puisse tirer des données sur les frères et sœurs adultes sont celles faites à partir des informations fournies par les répondants de 20-24 ans. Mais seuls quelques pourcents de leurs frères et sœurs sont décédés et les estimations qu'on peut en tirer ont de larges intervalles de confiance, même dans les enquêtes sur de grands échantillons.
- La procédure indirecte d'estimation de la mortalité adulte à partir de l'information sur la fratrie adulte n'implique pas l'hypothèse d'une population fermée à la migration. Il peut néanmoins être difficile d'interpréter des estimations fondées sur la fratrie concernant la mortalité d'unités géographiques infranationales, comme des zones urbaines et rurales ou des districts, ou des groupes de répondants ayant des caractéristiques socioéconomiques spécifiques. En effet, si les frères et sœurs partagent généralement la même identité ethnique, il arrive fréquemment que les frères et sœurs vivent ailleurs que les répondants et aient des caractéristiques socioéconomiques différentes d'eux. Les estimations relatives aux populations infranationales risquent aussi d'avoir de très larges intervalles de confiance.

APPLICATION DE LA MÉTHODE

Étape 1 : Calculer les rapports de survie conditionnelle de la table de mortalité

La procédure d'estimation des indices de la table de mortalité à partir des proportions de frères et sœurs encore en vie est la même que l'on analyse les données relatives aux frères, aux sœurs ou aux uns et aux autres et que les répondants soient les hommes, les femmes ou les deux conjointement. La feuille de calcul Excel associée à ce chapitre (voir le site web) contient des tableaux pour l'estimation de la survie des hommes et celle des femmes à partir de données fournies par des répondants hommes, femmes et des deux sexes. Il convient de saisir dans les tableaux appropriés le nombre de frères ou de sœurs en vie à 15 ans et le nombre de frères ou de sœurs encore en vie à l'enquête par groupe quinquennal d'âge du répondant, ou les proportions de frères/sœurs en vie à 15 ans qui le sont encore à l'enquête. Les estimations sont établies à partir des données des répondants âgés de 20 à 49 ans.

La survie est estimée entre le 15^e anniversaire et 15 + n ans, où n est la limite supérieure de chacun des groupes d'âge successifs des répondants. L'équation de régression est la suivante et les coefficients du tableau 23.1 sont utilisés :

$${}_{n-15}p_{15} = a(n) + b(n){}_5S_{n-5}$$

où ${}_5S_{n-5}$ représente, pour les répondants âgés de $n - 5$ à n , la proportion encore en vie des frères et sœurs qui étaient en vie à leur 15^e anniversaire. Par exemple, si $n = 25$, la survie de la table de mortalité est estimée sur l'intervalle de 10 ans entre le 15^e et le 25^e anniversaire, à partir des données sur la survie des frères et sœurs fournies par les répondants de 20-24 ans.

TABLEAU 23.1 Coefficients pour l'estimation de la survie adulte à partir des proportions de répondants ayant des frères ou des sœurs vivants

| n | $a(n)$ | $b(n)$ |
|-----|----------|--------|
| 25 | - 0,0003 | 1,0011 |
| 30 | - 0,1546 | 1,1560 |
| 35 | - 0,1645 | 1,1660 |
| 40 | - 0,1388 | 1,1406 |
| 45 | - 0,1140 | 1,1168 |
| 50 | - 0,1018 | 1,1066 |

Source : Timæus, Zaba and Ali (2001)

Étape 2 : Convertir les rapports de survie en estimations du niveau de mortalité

Pour déduire les tendances de la mortalité à partir d'une série de rapports de survie, ${}_{n-15}p_{15}$, établis pour les répondants de différents groupes d'âge et se rapportant à différentes dates, il faut convertir ceux-ci en un indice commun de mortalité qui puisse être comparé au fil du temps. On le fait en ajustant à chaque mesure une table-type de mortalité logit relationnelle à 1 paramètre, afin de tirer du modèle l'indice de mortalité commun.

Une large gamme d'indices a été utilisée dans ce but, incluant les paramètres de niveau de divers systèmes de tables-types de mortalité, des rapports de survie comme ${}_{35}p_{15}$, qui correspond à l'intervalle d'âge le plus large pour lequel la méthode des fratries adultes fournit des mesures, et l'espérance de vie à 15 ans. Le recours aux paramètres des modèles permet de souligner que la table de mortalité complète est estimée par l'ajustement d'un modèle, plutôt que mesurée directement. La mesure de l'espérance de vie

résume la survie sur l'ensemble des âges adultes, alors que le recours à la fonction de survie évite l'extrapolation aux âges avancés à partir des mesures relatives aux adultes plus jeunes. Dans les années récentes, les estimations ont été présentées de plus en plus souvent en termes de probabilité pour une personne de 15 ans de décéder avant 60 ans, ${}_{45}q_{15}$, cette mesure ayant la faveur de plusieurs agences internationales comme indicateur de la mortalité des adultes jeunes et d'âge moyen.

Dans les applications de la méthode de la fratrie adulte présentées ici, les rapports de survie sont convertis en valeurs de α , le paramètre de niveau d'un système de tables-types de mortalité logit relationnelles, puis en estimations dans ces modèles ajustés de la probabilité conditionnelle qu'une personne âgée de 15 ans décède avant son 60^e anniversaire (${}_{45}q_{15}$), son 50^e anniversaire (${}_{35}q_{15}$), ou son 40^e anniversaire (${}_{25}q_{15}$). Le dernier de ces indices se situe au milieu de la série des valeurs de ${}_{n-15}q_{15}$ estimées grâce aux modèles de régression. Le paramètre de la table-type de mortalité est calculé ainsi :

$$\alpha = -\frac{1}{2} \ln \left(1 + \frac{\frac{{}_{n-15}p_{15} - 1}{l_s(n)}}{1 - {}_{n-15}p_{15}} \right)$$

où les estimations de ${}_{n-15}p_{15}$ proviennent de l'étape 1 et les $l_s(x)$ valeurs de sont tirées d'une table de mortalité standard. On obtient donc une série de valeurs de α correspondant aux estimations de la fonction de survie obtenues à partir des données des répondants des différents groupes d'âge. Des valeurs plus élevées de α correspondent à une mortalité plus élevée. Ensuite pour chaque α ,

$${}_xq_{15} = 1 - \frac{1 + e^{2(\alpha + Y_i(15))}}{1 + e^{2(\alpha + Y_i(15+x))}}$$

pour $x = 25, 35$, et 45 .

Le fichier Excel (voir le site web) permet de calculer α et les trois probabilités de décès en prenant comme standard une des tables-types de mortalité du réseau Général des Nations Unies (1982) ou de l'un des quatre réseaux de Princeton (Coale, Demeny and Vaughan 1983). La table de mortalité standard doit être choisie de sorte qu'elle ait un profil par âge de mortalité adulte qui ressemble à celui de la population étudiée. Une autre table de mortalité peut être utilisée comme standard s'il y a des raisons de croire qu'elle ressemble davantage au profil de la mortalité adulte

de la population étudiée. Le standard le mieux adapté peut ne pas appartenir au réseau des tables types qui saisit le mieux la relation entre la mortalité des adultes et celle des jeunes enfants. Si on ne sait rien du profil par âge de la mortalité adulte, nous recommandons d'utiliser le modèle Général des Nations Unies ou le modèle Ouest de Princeton.

Étape 3 : Calculer la localisation dans le temps des estimations

Chaque rapport de survie obtenu par la méthode de la survie des fratries adultes représente une moyenne de la mortalité depuis les 15^e anniversaires des frères et sœurs. La durée d'exposition est plus longue pour les plus âgés que pour les plus jeunes, mais c'est en moyenne environ $n - 18,3$ ans (comme expliqué plus loin). Le point du temps où le rapport est égal à la survie du moment dépend du niveau de la mortalité et peut être estimé à partir de la proportion de frères et sœurs encore en vie et de l'âge des répondants.

La seule information supplémentaire dont on ait besoin pour calculer les dates auxquelles se réfèrent les estimations de la mortalité adulte est la date de l'enquête où les répondants ont été interrogés sur la survie de leurs frères et sœurs. Elle peut être calculée comme la moyenne des dates des interviews ou comme le point médian de la période de travail sur le terrain, si les dates exactes des interviews ne sont pas connues.

La localisation estimée est égale à la date du travail de terrain moins T , où T peut être calculé pour chaque estimation grâce à l'équation suivante et les coefficients du tableau 23.2.

$$T = a(n) - b(n) \ln({}_5S_{n-5})$$

TABLEAU 23.2 Coefficients pour calculer la localisation dans le temps des estimations de la survie des adultes à partir des proportions de répondants dont les frères et sœurs sont encore en vie

| n | $a(n)$ | $b(n)$ |
|-----|--------|--------|
| 25 | 3,23 | 1,12 |
| 30 | 5,46 | 1,95 |
| 35 | 7,52 | 2,78 |
| 40 | 9,38 | 3,62 |
| 45 | 11,00 | 4,45 |
| 50 | 12,32 | 5,28 |

Source : Timæus, Zaba and Ali (2001)

EXEMPLE

Cet exemple s'appuie sur les données concernant la survie des frères et sœurs recueillies lors de l'Enquête mondiale sur la santé en 2003 au Bangladesh (consultation du 17/11/2012). Cette enquête a collecté des données auprès des hommes et des femmes adultes sur à la fois leurs frères et leurs sœurs. Les réponses des femmes concernant leurs sœurs sont utilisées pour illustrer les calculs nécessaires à la méthode.

Étape 1 : Calculer les rapports

de survie conditionnelle de la table de mortalité

Le nombre de sœurs que les répondantes ont déclarées comme survivantes à leur 15^e anniversaire et le nombre de ces sœurs déclarées comme encore vivantes à la date d'enquête figurent aux deuxième et troisième colonnes du tableau 23.3. Ces nombres ont été obtenus selon la démarche standard, en ne pondérant les réponses que par les pondérations de l'enquête. Les proportions de sœurs encore en vie à la quatrième colonne sont calculées en divisant les effectifs de la troisième colonne par ceux de la deuxième. Les rapports de survie figurent à la sixième colonne du tableau 23.3. Ils ont été estimés à partir des proportions de la quatrième colonne et des coefficients de régression du tableau 23.1. Ainsi, pour les répondantes de 25-29 ans,

$${}_{10}p_{15} = -0,0003 + 1,0011 \times 0,9533 = 0,9541.$$

Étape 2 : Convertir les rapports

de survie en estimations du niveau de mortalité

La septième colonne du tableau 23.3 contient les valeurs de α , le paramètre de niveau du système de tables types de mortalité logit relationnelles pour une table-type de Princeton, Sud, sexes réunis, $e_0 = 60$. La huitième colonne du tableau contient les estimations de ${}_{25}q_{15}$ (la probabilité qu'une personne de 15 ans décède avant son 40^e anniversaire). Par exemple, α est calculé ainsi à partir de l'estimation de ${}_{10}p_{15}$ pour les femmes :

$$\alpha = -\frac{1}{2} \ln \left(1 + \frac{\frac{0,9541}{0,8376} - \frac{1}{0,8557}}{1 - 0,9541} \right) = 0,517.$$

Ayant calculé α , la valeur correspondante de ${}_{25}q_{15}$ est donc

$${}_{25}q_{15} = 1 - \frac{1 + e^{2(0,517-0,8899)}}{1 + e^{2(0,517-0,6902)}} = 0,136.$$

Étape 3 : Calculer la localisation

dans le temps des estimations

Les dates dans la neuvième colonne du tableau 23.3 ont été calculées en retranchant la localisation dans le temps des estimations de la date de l'enquête en 2003 au Bangladesh,

TABLEAU 23.3 Estimation de la survie des femmes, des dates où ces estimations sont localisées dans le temps, et estimations correspondantes de α et ${}_{25}q_{15}$, partir de la survie de sœurs adultes, Bangladesh, 2003

| Groupe d'âge | Sœurs en vie à 15 ans | Sœurs encore en vie | Proportion en vie (${}_5S_{n-5}$) | n | $\frac{l(n)}{l(15)}$ | Niveau (α) | Probabilité de décéder (${}_{25}q_{15}$) | Date |
|--------------|-----------------------|---------------------|-------------------------------------|-----|----------------------|---------------------|--|--------|
| 15-19 | 871,6 | 851,8 | 0,9773 | | | | | |
| 20-24 | 858,8 | 818,7 | 0,9533 | 25 | 0,9541 | 0,517 | 0,136 | 2000,0 |
| 25-29 | 964,7 | 901,1 | 0,9340 | 30 | 0,9251 | 0,535 | 0,139 | 1997,7 |
| 30-34 | 766,9 | 702,2 | 0,9156 | 35 | 0,9031 | 0,467 | 0,128 | 1995,6 |
| 35-39 | 626,4 | 554,4 | 0,8850 | 40 | 0,8706 | 0,473 | 0,129 | 1993,5 |
| 40-44 | 552,8 | 490,7 | 0,8877 | 45 | 0,8774 | 0,226 | 0,093 | 1991,8 |
| 45-49 | 495,9 | 401,5 | 0,8095 | 50 | 0,7940 | 0,436 | 0,124 | 1989,9 |

soit le 30/04/2003. La localisation dans le temps de chaque estimation est calculée à partir du groupe d'âge des répondants et des proportions encore en vie grâce aux équations du tableau 23.2. Par exemple, pour la première estimation de la survie des sœurs, le calcul est :

$$T = 3,23 - 1,12 \times \ln(0,9533) = 3,28.$$

DIAGNOSTICS, ANALYSE ET INTERPRÉTATION

Contrôles et validation

Le nombre de répondants qui ont déclaré ne pas savoir combien de leurs frères et sœurs ont atteint l'âge de 15 ans ou combien d'entre eux sont encore en vie ou qui n'ont pas répondu du tout à la question doit être contrôlé avant d'être exclu de l'analyse. Si ce nombre est important, il se peut que l'information donnée par ceux qui ont répondu à ces questions ne soit pas représentative de l'ensemble de la population. En outre, un taux élevé de non réponse peut indiquer que les enquêteurs ou les répondants ont eu des difficultés avec les questions. Il est alors possible que les réponses, même quand elles ont été fournies, ne soient pas fiables. S'il y a un taux élevé de non réponse à une question, il peut être intéressant de savoir si le défaut est concentré sur une minorité d'enquêteurs ou sur un certain type de répondants.

Si l'information sur la survie des frères et sœurs a été obtenue à la fois de répondants hommes et femmes dans un recensement ou une grande enquête, il est recommandé d'établir les proportions de frères et de sœurs encore en vie séparément pour les répondants de chaque sexe en vue de comparer la cohérence des déclarations. La cohérence des déclarations ne garantit pas leur exactitude, mais des différences statistiquement significatives entre les proportions calculées pour les répondants hommes et femmes impliquent qu'au moins un des deux sexes, et peut-être les deux, donnent des réponses inexactes. Il est assez courant que les répondants hommes déclarent moins de frères et sœurs que les répondantes et, en particulier, moins de frères et sœurs décédés. Dans d'autres enquêtes, les deux sexes peuvent déclarer les mêmes nombres de frères et sœurs en vie à 15 ans, mais des nombres différents d'entre eux encore en vie. Le premier type de désaccord peut tenir à des différences dans les erreurs de déclaration des âges, mais pas le second type.

Tout biais dû à l'existence de grappes de mortalité au sein des familles entraîne des sous-estimations. Il semble en outre peu vraisemblable que les répondants inventent des frères et

sœurs ou qu'ils les déclarent vivants alors qu'ils sont décédés. L'analyse doit donc probablement se concentrer sur les données fournies par le sexe qui déclare les proportions les plus faibles de frères et sœurs vivants.

Interprétation

Les résultats de l'application de la méthode indirecte des fratries adultes aux données de l'Enquête mondiale sur la santé au Bangladesh en 2003 sont représentés graphiquement sur la figure 23.1. L'indicateur de la table de mortalité choisi dans cet exemple est la probabilité de décéder entre 15 et 40 ans.

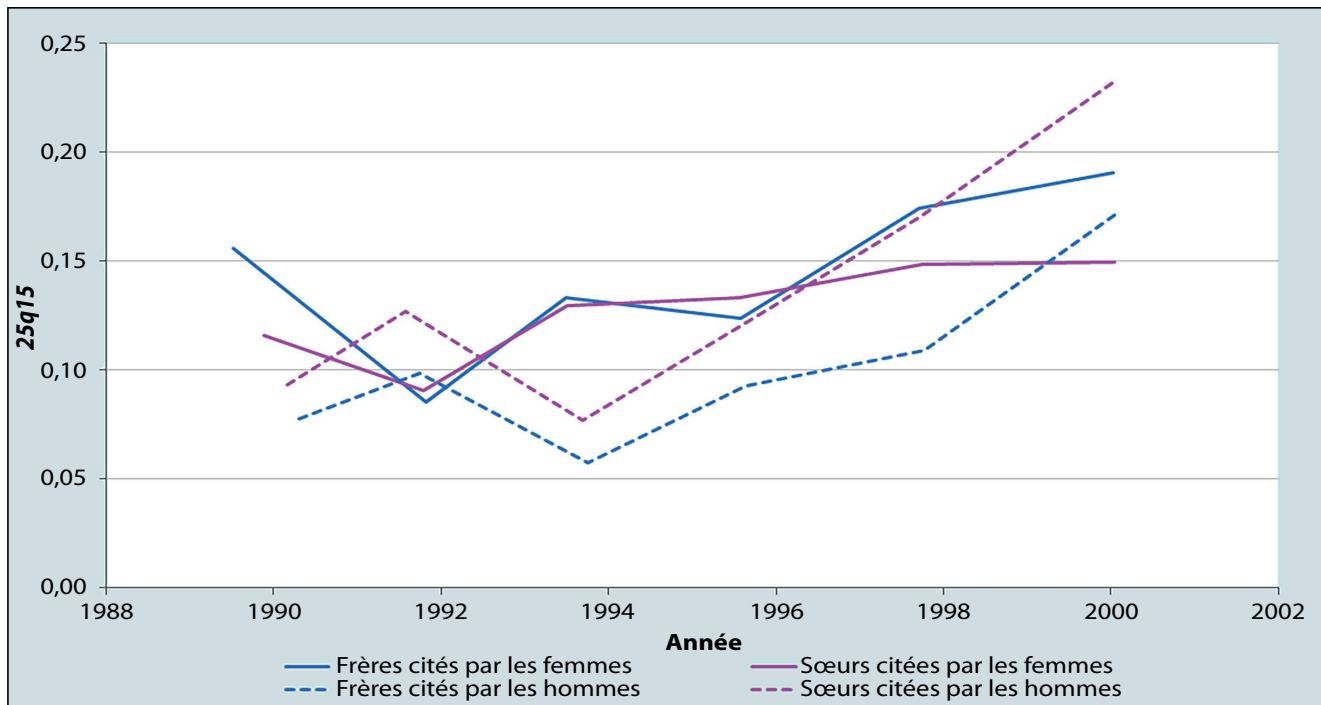
Les tableaux des proportions des sœurs encore en vie parmi celles qui étaient vivantes à 15 ans, selon les réponses des femmes, et des proportions des frères survivants parmi ceux qui étaient vivants à 15 ans, selon les réponses des hommes, ont été réalisés selon la démarche standard, en pondérant les réponses uniquement par les pondérations de l'enquête. En revanche, les proportions équivalentes concernant les sœurs des hommes et les frères des femmes ont fait l'objet d'une pondération supplémentaire par l'inverse du nombre de frères ou sœurs de même sexe que le répondant.

Les quatre séries tendent à fluctuer de façon erratique et suggèrent une hausse de la mortalité adulte au Bangladesh pendant les années 1990. Ceci semble peu vraisemblable et laisse penser que les estimations faites à partir des données des répondants âgés sont biaisées vers le bas par l'omission de frères et sœurs décédés ou par d'autres biais.

Les estimations de la mortalité féminine (c'est-à-dire la mortalité des sœurs) faites à partir des réponses des hommes et des femmes indiquent à peu près le même niveau de mortalité. Mais les estimations de la mortalité masculine (c'est-à-dire la mortalité des frères) faites à partir des réponses des hommes indiquent une mortalité nettement plus faible que les estimations faites à partir des réponses des femmes. Ces dernières estimations suggèrent que la mortalité est à peu près la même pour les hommes et les femmes en début de vie adulte, mais les estimations faites à partir des réponses des hommes suggèrent que les jeunes hommes ont une mortalité nettement plus faible que les jeunes femmes au Bangladesh.

Ces estimations apparaissent clairement de piètre qualité. Elles sous-estiment sans doute fortement la mortalité des jeunes adultes au Bangladesh. Une explication très plausible des discordances apparentes entre les différentes séries d'estimation est que les hommes risquent davantage que les femmes d'omettre des frères et sœurs décédés dans leurs

FIGURE 23.1 Évolution de la probabilité de décéder entre le 15^e et le 40^e anniversaire estimée à partir de la survie de la fratrie adulte, Bangladesh, Enquête mondiale sur la santé, 2003



réponses, mais que le biais à la baisse résultant des grappes de mortalité au sein des fratries est plus fort dans les estimations fondées sur les déclarations concernant les fratries de même sexe que les estimations fondées sur les déclarations concernant les fratries de sexe opposé. Les estimations fondées sur les réponses des hommes concernant leurs frères sont particulièrement basses car elles sont sévèrement affectées par les deux biais. À l'inverse, les deux séries féminines sont relativement cohérentes car chacune est sévèrement affectée par un biais mais moins par l'autre. La conséquence de ce schéma d'erreurs, si c'est bien l'explication des différences entre les séries, est que les deux ensembles d'estimations pour les femmes sous-estiment probablement la mortalité davantage que les estimations pour les frères fondées sur les réponses des femmes. Il semble donc possible que la mortalité des jeunes femmes au Bangladesh reste plus élevée que celle des hommes du même âge.

DESCRIPTION DÉTAILLÉE DE LA MÉTHODE

Introduction

Les méthodes développées initialement pour estimer la mortalité à partir des informations sur la survie des frères et sœurs s'appuyaient sur l'idée que, en moyenne, les âges des frères et sœurs sont proches de l'âge du répondant. La proportion de frères et sœurs du répondant encore en vie est donc un bon estimateur de la survie dans la table de mortalité jusqu'à l'âge du répondant (Hill and Trussell 1977 ; Division de la Population des NU 1984). Malheureusement, l'expérience de terrain a montré que la qualité des données collectées sur les frères et sœurs était souvent médiocre, car les frères et sœurs décédés avant ou peu après la naissance du répondant étaient souvent omis par celui-ci, qui peut ne pas les connaître du tout (Blacker and Brass 1983 ; Zaba 1986).

L'intérêt pour l'estimation de la mortalité à partir des données sur la fratrie a été relancé par le développement de la méthode des sœurs pour la mesure de la mortalité maternelle (Graham, Brass, et Snow 1989). Celle-ci nécessite des données sur le nombre de sœurs du répondant ayant survécu jusqu'à 15 ans, le nombre de celles décédées par la suite, et si les sœurs décédées l'ont été pendant une grossesse ou dans les 6 à 8 semaines suivant un accouchement. Le fait de limiter la prise en compte des sœurs à celles qui ont survécu jusqu'à 15 ans exclut les sœurs décédées jeunes et que le répondant risque de ne pas avoir connues ou d'avoir oubliées. Les réponses données aux deux premières de ces questions permettent de calculer les proportions de sœurs encore en vie parmi celles ayant survécu jusqu'à 15 ans par groupe quinquennal d'âge. Ces proportions peuvent être utilisées pour estimer la mortalité féminine adulte toutes causes. Les données comparables sur les frères des répondants peuvent être utilisées pour estimer la mortalité masculine.

La seule information nécessaire pour l'application de la méthode indirecte de la fratrie adulte est donc la donnée résumée sur la proportion de frères et sœurs adultes encore en vie parmi ceux qui ont survécu jusqu'à 15 ans, classée par groupe d'âge des répondants. Comme les frères et sœurs ont environ le même âge en moyenne que les répondants, ces proportions sont à peu près égales à

$$l_x / l_{15}$$

pour les répondants âgés de x . Comme la relation est étroite entre cette mesure et la proportion de frères et sœurs encore en vie parmi ceux ayant survécu jusqu'à l'âge adulte, elle peut être estimée de façon relativement précise, même dans les populations ayant un profil inhabituel de mortalité par âge, comme celles subissant une forte épidémie de VIH.

Exposé mathématique

En utilisant la démarche probabiliste développée par Goodman, Keyfitz et Pullum (1974), Timæus *et al.* (2001) montrent que, dans une population stable, le nombre de frères et sœurs déjà nés y années avant un répondant aujourd'hui âgé de a est égal à

$${}_aV_y = \int_s^{\omega} e^{-r(z-a)} f(z) l(z) f(z-y) dz, y \geq 0 \quad (1)$$

$${}_aV_y = \int_s^{\omega} e^{-r(z-a)} f(z) l(z-y) f(z-y) dz, y < 0 \quad (2)$$

où l'équation 1 donne le nombre de frères et sœurs plus âgés et l'équation 2 le nombre de ceux plus jeunes que le répondant, l'intégrale couvrant tous les âges à la naissance des enfants de s à ω , et

z = âge de la mère à la naissance du répondant

$f(z-y)$ = probabilité que les mères des répondants aient une naissance à l'âge $z-y$, sachant qu'elles ont eu le répondant à l'âge z

r = taux de croissance dans une population stable.

Noter que dans les équations 1 et 2, $f(x)$ est une *distribution des naissances*, c'est-à-dire la distribution des âges à la maternité d'une femme donnée, pas la distribution de la fécondité dans l'ensemble de la population.

La proportion de frères et sœurs encore en vie parmi ceux qui survivaient à 15 ans pour les répondants dans un groupe quinquennal d'âge, x à $x+5$, est donnée par

$${}_5S_x^{15+} = \frac{\int_x^{x+5} l(a) \int_a^{\omega-s} {}_aV_y l(a+y) dy da}{l(15) \int_x^{x+5} l(a) \int_a^{\omega-s} {}_aV_y dy da}, x \geq 15 \quad (3)$$

Le calcul de la proportion de frères et sœurs en vie pour les répondants d'un âge donné nécessite un modèle de distribution des naissances pour une femme donnée. Hill et Trussell (1977) ont supposé que toutes les mères ont eu les taux de fécondité par âge de la population générale. Ils ont pu ainsi en tirer une distribution par âge des frères et sœurs comme une convolution de la distribution de la fécondité. Toutefois, si les femmes débutent la constitution de leur descendance à un large éventail d'âges, mais compriment ensuite le temps de cette constitution dans une faible fraction de leur période féconde, comme on le voit dans les populations de faible fécondité, on peut s'attendre à ce que la variance de la distribution des naissances soit bien moindre que la variance du schéma de fécondité.

En revanche, en développant la méthode des sœurs, Graham, Brass, et Snow (1989) ont supposé que ${}_aV_y$, la distribution des différences d'âge entre frères et sœurs, pouvait être représentée par une distribution normale de moyenne zéro et de variance 80 années-carrées. Cette hypothèse simplifie considérablement la procédure d'estimation de la proportion de frères et sœurs encore en vie, mais elle est difficile à justifier théoriquement. En particulier, la distribution des différences d'âge entre frères et sœurs ne serait normale que si la distribution des naissances de la mère était elle-même

normale. L'utilisation d'une distribution normale pour la distribution des différences d'âge entre frères et sœurs est une approximation raisonnable si la distribution des naissances présente un pic (c'est-à-dire $\sigma^2 < 35$), mais elle est moins satisfaisante quand les distributions des naissances sont plus plates, comme c'est le cas dans les populations de fécondité élevée et moyenne.

Une autre question apparaît dans les populations en croissance ou en décroissance. Goldman (1978) a montré que, dans une population croissante, un individu choisi au hasard parmi ceux dont les mères ont terminé la constitution de leur descendance ont davantage de frères et sœurs plus jeunes qu'eux que de frères et sœurs plus âgés. Le contraire est vrai dans une population décroissante. On peut comprendre ceci intuitivement en considérant des répondants âgés aujourd'hui de 40 ans, dont les mères ont terminé la constitution de leur descendance. Dans une population croissante stable, les répondants seront relativement plus nombreux à avoir des mères jeunes (disons aujourd'hui âgées de moins de 65 ans, si elles ont survécu) que dans une population stationnaire car, au moment de leur naissance, il y aura eu davantage de femmes âgées de moins de 25 ans que dans la population stationnaire correspondante. Mais si les répondants sont des enfants de mères jeunes, ils ont plus de chances d'avoir des frères et sœurs plus jeunes plutôt que plus âgés car leurs mères ont davantage de naissances d'enfants devant elles que derrière elles.

La distribution des différences d'âge entre frères et sœurs n'est donc pas symétrique : sa moyenne est inférieure à zéro dans une population croissante et supérieure dans une population décroissante. Plus précisément, si la variance de la distribution des naissances sous-jacente est σ^2 , la moyenne de la distribution des âges des frères et sœurs est approximativement $-r\sigma^2$, où r est le taux de croissance de la population. Donc, même si toutes les femmes avaient la même fécondité par âge, la variance de la distribution par âge des frères et sœurs dans une population croissante serait légèrement inférieure au double de la variance de la distribution de fécondité et la distribution aurait une asymétrie positive. Cette distribution a des caractéristiques inverses dans une population décroissante.

En vue de traiter ces questions, Timæus, Zaba et Ali (2001) ont proposé un modèle des différences d'âge entre frères et sœurs qui synthétise les deux démarches précédentes. Sur la base d'une étude des distributions des âges de fratries âgées dans des historiques de naissances recueillis

dans 12 enquêtes représentatives nationalement, réalisées dans le cadre de l'Enquête mondiale de fécondité (EMF), ils ont conclu que les variances des distributions de naissances dans le monde en développement s'étagent entre environ 40 et 110 années-carrées. Ils ont ensuite adapté le modèle relationnel de Gompertz de fécondité (Brass 1974, 1981) pour représenter ces distributions de naissances, en fixant le paramètre b dans leur ensemble de modèles à des valeurs qui varient entre 1 et 1,8, afin de générer des distributions de la largeur appropriée (à mesure que b augmente la variance des distributions du modèle diminuent). Pour tenir compte de l'absence d'intervalles entre naissances très courts dans les populations humaines, ${}_aV_0$ a été mis à 0 et ${}_aV_1$ et ${}_aV_{-1}$ à 40 % des valeurs du modèle. La valeur de 40 % correspond à la moyenne des rapports ${}_aV_1 / {}_aV_2$ dans les 12 populations de l'EMF.

Mise en œuvre de la méthode

L'équation 3 doit être calculée numériquement, mais il n'y a pas de raison en principe qu'elle ne puisse pas être résolue directement pour une fonction de survie d'une table de mortalité, en utilisant le solveur d'Excel ou un outil équivalent et une distribution des naissances, ${}_aV_y$, qui soit adaptée à la population étudiée. Pour obtenir une solution unique, il faut encore faire une hypothèse sur le profil par âge de la mortalité adulte, par exemple en choisissant le standard dans un système de tables-types de mortalité à 1 paramètre. En pratique, les estimations sont généralement obtenues grâce à des modèles de régression qui ont été ajustés à des données simulées sur la survie des frères et sœurs, générées pour des populations représentant un large éventail de structures par âge, de distribution des naissances et de schémas de mortalité (Timæus, Zaba and Ali 2001). Les coefficients de ces modèles figurent au tableau 23.1.

Exposé mathématique – localisation dans le temps des estimations

Les méthodes de localisation dans le temps visent à estimer le moment T où les mesures de survie dans les cohortes, qui ont entraîné la proportion de frères et sœurs survivants ${}_aP_y^c$, sont égales aux mesures équivalentes du moment, ${}_aP_y(T)$, de sorte que

$$S(a) = \int_s^{\omega} {}_aV_y {}_aP_y(T) dy$$

où :

$${}_a v_y = \frac{{}_a V_y}{\int_s {}_a V_y dy}$$

Si nous désignons par ${}_a g_y$ le temps moyen écoulé depuis le décès de ceux qui sont décédés entre y et $y + a$, par définition

$${}_a g_y = \frac{\int_y^{y+a} (z - y - a) \mu(z) l(z) dz}{\int_y^{y+a} \mu(z) l(z) dz} \quad (4)$$

où $\mu(z)l(z)$ sont les décès de la table de mortalité à l'âge z . Brass et Bamgboye (1981) montrent que si les schémas de mortalité sont conformes à un système de tables-types de mortalité logit relationnelles à 1 paramètre et que l'évolution de la mortalité adulte est supposée linéaire en α , le paramètre de ce système relationnel de tables-types, le moment où la survie des adultes dans les cohortes est égale à la survie du moment est une moyenne pondérée des durées écoulées depuis le décès des frères et sœurs des répondants :

$$T = \frac{\int_x^{x+5} l(a) \int_s^\omega {}_a v_y l(y + a, T) (1 - {}_a p_y(T)) {}_a g_y dy da}{\int_x^{x+5} l(a) \int_s^\omega {}_a v_y l(y + a, T) (1 - {}_a p_y(T)) dy da} \quad (5)$$

Cette localisation dans le temps dépend du niveau de la mortalité et des âges des parents, mais elle est indépendante du taux de variation en α . Pour établir l'équation 5, Brass et Bamgboye tirent parti d'une relation entre les variations de la mortalité au fil des âges et au fil du temps qui est propre au système logit relationnel de tables-types, mais il est possible d'arriver à des formules similaires pour T sur la base d'autres hypothèses raisonnables sur l'évolution dans le temps de la mortalité par âge (Palloni, Massagli and Marcotte 1984).

L'équation 5 peut être évaluée numériquement, en utilisant des valeurs de ${}_a v_y$ et pour des mesures de la table de mortalité choisies sur la base de données observées. Pour développer une procédure aisée d'estimation de T à partir des caractéristiques observées d'une population, il faut supposer une relation beaucoup plus simple que celle postulée. Brass et Bamgboye (1981) font valoir que la variation de T avec a dans des intervalles d'âge limités est suffisamment proche de la linéarité pour que tous les

répondants dans un groupe quinquennal d'âge puissent être traités comme regroupés à l'âge central N . Ils font valoir ensuite qu'aux âges et aux niveaux de mortalité où les méthodes indirectes sont utilisées pour estimer la mortalité adulte, la force de mortalité s'accroît à peu près exponentiellement avec l'âge. En conséquence, pour de telles applications, la variation de ${}_a g_y$ avec y est insignifiante. C'est pourquoi les facteurs de pondération de ${}_a g_y$ dans l'équation 5 ont peu d'effet et tous les parents adultes peuvent être traités comme si leur exposition au risque débutait à leur âge moyen à l'entrée, M . D'où l'approximation satisfaisante,

$$T = {}_N g_M$$

Si la survie des adultes recule linéairement avec l'âge, de sorte que le même nombre de décès survienne à chaque âge, alors ${}_N g_M$ est égal à $N/2$, quelle que soit la valeur de M . Dans des tables de mortalité moins extrêmes, la mortalité s'accroît avec l'âge plus rapidement que cela et les décès des frères et sœurs sont concentrés aux âges avancés et, en conséquence, dans la portion récente de la période de N années. Ceci signifie que la localisation dans le temps des estimations est plus proche de l'enquête que $N/2$. En remplaçant $\mu(z)l(z)$ par $e^{kz} \mu(y)l(y)$ dans l'équation 4 et en développant la partie droite en puissances de N , Brass et Bamgboye (1981) démontrent que l'ajustement approprié pour ${}_N g_M$ est une fonction du niveau de mortalité tel que mesuré par k et de l'âge des répondants mesuré par N :

$${}_N g_M \approx \frac{N}{2} \left(1 - k \frac{N}{6} \right) \quad (6)$$

Brass et Bamgboye (1981) démontrent aussi que l'hypothèse d'une augmentation exponentielle de la mortalité avec l'âge implique que, dans un système de tables de mortalité logit relationnelles :

$$\frac{e^{kN}}{e^{k_s N}} = \frac{({}_N p_M)^2}{(l_s(M + N) / l_s(M))^2}$$

En résolvant pour kN et en remplaçant cette expression dans l'équation 6, on obtient une estimation de ${}_N g_M$ et donc de T :

$$T = \frac{N}{2} \left(1 - \frac{\ln({}_N p_M)}{3} + \frac{1}{3} \ln \left(\frac{l_s(M + N)}{l_s(M)} \right) - \frac{k_s N}{6} \right)$$

Dans cette formulation, les localisations dans le temps des mesures de survie conditionnelle tirées des données sur les frères et sœurs adultes sont estimées à la moitié de la durée d'exposition, N , réduite par un facteur qui dépend du niveau de la survie conditionnelle comparé au niveau de la survie dans une table de mortalité standard.

Ayant obtenu cette expression de T sur une base théorique, Brass (1985) approxime ${}_Np_M$ par ${}_5S_x$ et adopte comme standard une table de mortalité où $l_s(x)$ est linéaire aux âges adultes et est pris égal à $(1 - x/80)/2$. Comme $l_s(x)$ est linéaire, $T = \frac{1}{2}N$ et ks devient nul. T est donc estimé à partir des données observées grâce à

$$T = \frac{N}{2} \left(1 - \frac{\ln({}_5S_x)}{3} + \frac{1}{3} \ln \left(\frac{80 - M - N}{80 - M} \right) \right) \quad (7)$$

Dans la méthode de la fratrie adulte, M , l'âge où commence l'exposition au risque est au quinzième anniversaire pour chaque frère et sœur. L'asymétrie de la distribution des différences d'âge entre frères et sœurs signifie que, dans une population croissante, les frères et sœurs sont en moyenne un peu plus jeunes que les répondants. La différence d'âge varie entre environ zéro et 1,75 an dans les populations où on a des chances de vouloir appliquer la méthode. On peut utiliser une valeur centrale de 0,8 an dans toutes les applications. La durée d'exposition au risque, N , devient donc $(n - 2.5 - 0.8) - 15$, où n est la limite supérieure du groupe d'âge des répondants. Comme M est fixé à 15 ans, l'équation 7 peut être simplifiée pour chaque groupe d'âge en une équation linéaire de la forme Timæus, Zaba et Ali (2001):

$$T = a(n) - b(n) \ln({}_5S_{n-5}^{15+}).$$

FONCTIONNEMENT DANS LES POPULATIONS SUBISSANT UNE ÉPIDÉMIE GÉNÉRALISÉE DE VIH

L'épidémie de VIH pose deux problèmes pour les méthodes indirectes d'estimation de la mortalité fondées sur la survie de parents. D'une part, les voies sexuelle et verticale de transmission créent des biais significatifs de sélection dans les données recueillies par enquête sur la survie de parents. D'autre part, l'incidence de l'infection de VIH est concentrée chez les jeunes adultes. Les populations avec une importante mortalité par Sida ont des profils de mortalité par âge très différents à la fois des autres populations et des tables-types de mortalité utilisées pour l'obtention des coefficients

permettant de convertir les données sur la survie de parents en mesures de la survie des tables de mortalité.

Un avantage majeur de la méthode des fratries pour mesurer la mortalité adulte, comparée à la méthode des proportions d'orphelins, est qu'elle est exempte des biais de sélection résultant de la transmission directe du virus. Il subsiste un biais résiduel dû à l'effet de grappe de la mortalité par Sida au sein des fratries. En particulier, le risque d'infection par le VIH varie fortement entre localités et les frères et sœurs vivent souvent à proximité les uns des autres. L'impact est cependant limité comparé aux biais qui affectent les données que les parents fournissent sur leurs enfants ou vice versa.

Le biais dans les coefficients de régression utilisés pour estimer la survie dans les tables de mortalité est un problème plus important. Par rapport à l'équation 3, c'est la modification du profil de la mortalité par âge des frères et sœurs qui pose problème, et non l'impact de l'épidémie sur la distribution des différences d'âge entre frères et sœurs, puisque le principal facteur de cette distribution est le profil des âges à la naissance des enfants plutôt que la mortalité ou la structure par âge.

Timæus, Zaba et Ali (2001) évaluent la sensibilité des estimations par la méthode de la fratrie adulte à ces problèmes grâce à une combinaison de données empiriques et simulées. Ils trouvent que même en présence d'un profil par âge de la mortalité inhabituel, tel qu'on l'observe dans les populations avec une forte mortalité par Sida, la méthode des fratries adultes donne des estimations de la survie proches des valeurs effectives. Les estimations fondées sur les données de répondants âgés de 20-24 ans et de plus de 40 ans sont extrêmement précises, alors que celles fondées sur les données des répondants âgés de 25-39 ans surestiment légèrement la survie. C'est parce que les coefficients de régression ne réussissent pas à refléter la concentration des décès par Sida dans cet intervalle d'âge.

Pour que les estimations de la survie adulte par la méthode des fratries permettent de suivre l'évolution de la mortalité, il faut ajuster une table-type de mortalité aux estimations pour des intervalles d'âge spécifiques et en tirer un indice se référant à un intervalle d'âge commun. Étonnamment, si on convertit la série complète des estimations en mesures de la survie de 15 à 50 ans, ${}_{35}p_{15}$, celles-ci restent tout à fait précises. Celles obtenues à partir des données de répondants âgés de 25 à 34 ans sont plus précises que les estimations de sur lesquelles elles sont fondées. Les erreurs dues à l'incapa-

cité de tenir compte de l'impact du Sida sur le schéma de mortalité, d'abord en calculant les coefficients, puis en tirant une mesure commune de survie, se compensent en grande partie. Ce résultat vaut quels que soient les caractéristiques générales de la mortalité et le standard de mortalité choisi. Les estimations de ${}_{35}p_{15}$ obtenues par la méthode des fratries adultes représentent sans doute des indices relativement robustes permettant de suivre l'évolution de la mortalité à mesure que l'épidémie de Sida se développe. Comme avec les autres méthodes indirectes, si des ensembles successifs de données sont collectés pour la même population, des contrôles de la cohérence des estimations pour les périodes où elles se recouvrent donnent une indication puissante de la précision des résultats.

AUTRES LECTURES ET RÉFÉRENCES

La méthode des fratries adultes n'est pas présentée dans les manuels classiques d'estimation indirecte (Sloggett, Brass, Eldridge et al. 1994 ; Division de la Population des NU 1984), mais elle est décrite dans le manuel des Nations Unies sur l'estimation de la mortalité adulte (UN Population Division 2002). La principale référence expliquant la base théorique de la méthode des fratries adultes et le développement des coefficients de régression pour convertir les proportions de frères et sœurs survivants en indice de la table de mortalité est Timæus et al. (2001). L'article passe en revue les contributions antérieures.

- Blacker JGC and W Brass. 1983. "Experience of retrospective enquiries to determine vital rates," in Moss, L and H Goldstein (eds). *The Recall Method in Social Surveys*. London: University of London Institute of Education, pp. 48-61.
- Brass W. 1974. "Perspectives in population prediction: illustrated by the statistics of England and Wales", *Journal of the Royal Statistical Society* **A137**(4):532-583.
- Brass W. 1981. "The use of the Gompertz relational model to estimate fertility," Paper presented at International Population Conference, Manila, 1981. Liège. International Union for the Scientific Study of Population. Vol. 3:345-362.
- Brass W. 1985. *Advances in Methods for Estimating Fertility and Mortality from Limited and Defective Data*. London: London School of Hygiene & Tropical Medicine.
- Brass W and EA Bamgboye. 1981. *The Time Location of Reports of Survivorship: Estimates for Maternal and Paternal Orphanhood and the Ever-widowed*. London: London School of Hygiene & Tropical Medicine.

EXTENSIONS ET VARIANTES DE LA MÉTHODE

La plupart des enquêtes qui ont recueilli l'information nécessaire à l'estimation de la mortalité toutes causes des adultes à partir des données sur les fratries adultes ont aussi demandé si les sœurs décédées étaient mortes pendant une grossesse ou peu après un accouchement. Ensemble, ces données fournissent la base pour l'application de la méthode des sœurs permettant d'estimer la mortalité maternelle (Graham, Brass and Snow 1989).

Il est aussi possible de calculer des estimations directes de la mortalité adulte à partir d'historiques détaillés des fratries collectées dans de nombreuses Enquêtes démographiques et de santé et certaines autres enquêtes.

- Coale AJ, P Demeny and B Vaughan. 1983. *Regional Model Life Tables and Stable Populations*. London: Academic Press.
- Division de la Population des NU. 1984. *Manuel X. Techniques indirectes d'estimation démographique*. New York : Nations Unies, Département des affaires économiques et sociales internationales, ST/ESA/SER.A/81. http://unstats.un.org/unsd/demographic/standmeth/handbooks/Manuel_X-fr.pdf
- Gakidou E and G King. 2006. "Death by survey: estimating adult mortality without selection bias from sibling survival data", *Demography* **43**(3):569-585. doi: <http://dx.doi.org/10.1353/dem.2006.0024>
- Goldman N. 1978. "Estimating the intrinsic rate of increase of a population from the average numbers of younger and older sisters", *Demography* **15**(4):499-521. doi: <http://dx.doi.org/10.2307/2061202>
- Goodman LA, N Keyfitz and TW Pullum. 1974. "Family formation and the frequency of various kinship relationships", *Theoretical Population Biology* **5**(1):1-27. doi: [http://dx.doi.org/10.1016/0040-5809\(74\)90049-5](http://dx.doi.org/10.1016/0040-5809(74)90049-5)
- Graham W, W Brass and RW Snow. 1989. "Estimating maternal mortality: The sisterhood method", *Studies in Family Planning* **20**(3):125-135. doi: <http://dx.doi.org/10.2307/1966567>
- Hill K and TJ Trussell. 1977. "Further developments in indirect mortality estimation", *Population Studies* **31**(2):313-334. doi: <http://dx.doi.org/10.2307/2173920>
- Masquelier B. 2013. "Adult mortality from sibling survival data: A reappraisal of selection biases", *Demography* **50**(1):207-228. doi: <http://dx.doi.org/10.1007/s13524-012-0149-1>
- Obermeyer Z, JK Rajaratnam, CH Park, E Gakidou et al. 2010. "Measuring adult mortality using sibling survival: a new analytical method and new results for 44 countries, 1974-2006", *PLoS Medicine* **7**(4):e1000260. doi: <http://dx.doi.org/10.1371/journal.pmed.1000260>

- Palloni A, M Massagli and J Marcotte. 1984. "Estimating adult mortality with maternal orphanhood data: analysis of sensitivity of the techniques", *Population Studies* **38**(2):255-279. doi: <http://dx.doi.org/10.1080/00324728.1984.10410289>
- Sloggett A, W Brass, SM Eldridge, IM Timæus, P Ward and B Zaba. 1994. *Estimation of Demographic Parameters from Census Data*. Tokyo, Japan: United Nations Statistical Institute for Asia and the Pacific.
- Timæus IM, B Zaba and M Ali. 2001. "Estimation of adult mortality from data on adult siblings," in Zaba, B and J Blacker (eds). *Brass Tacks: Essays in Medical Demography*. London: Athlone, pp. 43-66.
- Trussell J and G Rodriguez. 1990. "A note on the sisterhood estimator of maternal mortality", *Studies in Family Planning* **21**(6):344-346. doi: <http://dx.doi.org/10.2307/1966923>
- UN Population Division. 1982. *Model Life Tables for Developing Countries*. New York: United Nations, Department of Economic and Social Affairs. http://www.un.org/esa/population/techcoop/DemMod/model_lifetabs/model_lifetabs.html
- UN Population Division. 2002. *Methods for Estimating Adult Mortality*. New York: United Nations, Department of Economic and Social Affairs, ESA/P/WP.175. http://www.un.org/esa/population/techcoop/DemEst/methods_adultmort/methods_adultmort.html
- Zaba B. 1986. *Measurement of Emigration using Indirect Techniques: Manual for the Collection and Analysis of Data on Residence of Relatives*. Liège: Ordina.
- Zaba B and PH David. 1996. "Fertility and the distribution of child mortality risk among women", *Population Studies* **50**(2):263-278. doi: <http://dx.doi.org/10.1080/0032472031000149346>

Méthodes fondées
sur plusieurs recensements

Chapitre 24 LA MÉTHODE GÉNÉRALISÉE DE LA BALANCE DE L'ACCROISSEMENT DÉMOGRAPHIQUE

Rob Dorrington

DESCRIPTION DE LA MÉTHODE

Kenneth Hill (Hill 1987) a proposé une généralisation de la méthode de Brass de la balance de l'accroissement démographique (*Brass growth balance method*) faite pour estimer la complétude de l'enregistrement des décès par rapport à une estimation de la population, et qui repose sur l'hypothèse que la population est démographiquement stable. Cette généralisation s'applique au cas où la population n'est plus stable, mais reste fermée aux migrations. Elle peut s'utiliser dans le cas où l'on dispose des données sur les effectifs de population par groupe d'âge à deux recensements, et du nombre de décès par groupe d'âge entre les dates des deux recensements. Avec l'information supplémentaire des deux recensements, il est possible d'estimer les taux de croissance par âge, au lieu de prendre un taux constant comme on le fait dans l'hypothèse d'une population stable. Mais la méthode repose toujours sur les mêmes autres hypothèses : les proportions de décès déclarés et la complétude du recensement sont constants aux âges adultes, et les données déclarées sont précises. De plus, dans sa formulation la plus simple, la méthode repose sur l'hypothèse que la population est fermée aux migrations ; on peut cependant adapter la méthode pour prendre en compte les migrations si on dispose des données correspondantes.

Dans toutes les populations fermées, la relation suivante est vérifiée : $r(x+) = b(x+) - d(x+)$, où le taux partiel de natalité, $b(x+)$, est défini comme le taux auquel la population atteint l'âge x au sein de la population âgée de x années et plus ; et le taux partiel de mortalité, $d(x+)$, est défini comme le taux de mortalité dans la population âgée de x années et plus. Si, dans cette population, les décès sont sous-déclarés dans une même proportion à tous les âges, alors la relation devient : $b(x+) - r(x+) = d^r(x+) / c$, où $d^r(x+)$ est le taux de mortalité déclaré au-delà de l'âge x , et

c est la proportion de décès qui sont effectivement déclarés. En pratique, les recensements de population qui servent au calcul de $r(x+)$ peuvent ne pas être complets ; on fera alors l'hypothèse que la sous-déclaration est la même à tous les âges, ce qui permet de trouver c d'après la relation linéaire entre les points donnés par : $b(x+) - r(x+)$ et $d^r(x+)$. Les taux de mortalité corrigés s'obtiennent alors en divisant les nombres de décès déclarés dans chaque groupe d'âge par c , et en divisant ces nombres par l'estimation de la population au risque. Ceci permet de calculer les taux partiel de natalité, de mortalité et de croissance. De plus, un sous-produit de cette procédure est de permettre de corriger un des deux recensements, dans le cas où l'un des deux est moins complet que l'autre, de manière à ce que les deux soient cohérents, même s'ils ne sont ni l'un ni l'autre parfaitement complets.

DONNÉES REQUISES ET HYPOTHÈSES

Tabulation des données nécessaires

- Effectifs de femmes (ou d'hommes), par groupe d'âge de 5 ans, et pour l'intervalle d'âge ouvert $A+$ (A étant aussi élevé que possible), à deux instants donnés, le plus souvent provenant de deux recensements exhaustifs (voir les mises-en-garde ci-dessous pour le cas où l'on utilise des enquêtes par sondage au lieu des recensements).
- Nombre de décès de femmes (ou d'hommes), par groupe d'âge de 5 ans, et pour l'intervalle d'âge ouvert $A+$, au cours de la période située entre les deux recensements ou les deux enquêtes.

Hypothèses importantes

- La couverture de chaque recensement est la même à tous les âges.

- La complétude de la déclaration des décès est la même à tous les âges au-delà d'un âge minimal (en général 5 ans ou 15 ans).
- La population est fermée aux migrations. On peut cependant adapter la méthode pour tenir compte des migrations, mais les données précises des effectifs de migrations nettes n'existent que rarement. Pour les populations nationales, les migrations nettes sont souvent suffisamment faibles pour être négligées. Mais dans les situations où les migrations sont importantes, il convient de les prendre en compte pour interpréter les résultats et décider de l'estimation finale de la complétude.

TRAVAUX PRÉPARATOIRES ET RECHERCHES PRÉLIMINAIRES

Avant d'appliquer la méthode, il convient d'analyser la qualité des données, au moins en ce qui concerne les dimensions suivantes :

- structure par âge de la population;
- répartition par sexe de la population;
- structure par âge des décès; et
- répartition par sexe des décès.

Lorsque les décès déclarés se rapportent à une période autre que celle comprise entre les deux recensements, il convient d'estimer le nombre de décès qui auraient dû être déclarés pendant la période intercensitaire. Ainsi, lorsqu'on dispose des décès déclarés à l'état civil par année de calendrier, l'ajustement consistera à répartir proportionnellement les décès de l'année du premier recensement et ceux de l'année du second recensement. Si l'on dispose des décès déclarés par les ménages au cours des 12 derniers mois avant chaque recensement, il faudra estimer le nombre de décès au cours de la période intercensitaire par interpolation entre les deux estimations précédentes (voir la feuille de calcul : *Estimating deaths.xlsx*).

PRÉCAUTIONS ET MISES EN GARDE

Pour appliquer cette méthode, les analystes devront faire attention tout particulièrement aux points suivants.

Les processus d'interprétation et d'estimation finale doivent tenir compte de la source des données sur les décès (décès déclarés à l'état civil, décès déclarés par les ménages lors d'un recensement, décès enregistrés dans les hôpitaux), comme cela est expliqué ci-dessous. Mais les biais associés aux sources de données tendent à avoir moins d'impact sur l'estimation de la complétude obtenue par la méthode de la balance de l'accroissement

démographique, que par la méthode de l'extinction des cohortes synthétiques.

La question des migrations devient plus importante lorsqu'on applique la méthode à des zones géographiques régionales.

Le choix de l'intervalle d'âge utilisé pour ajuster la ligne droite aux taux partiels de natalité et de mortalité, qui détermine l'estimation de la complétude, est important. Les problèmes en la matière sont les suivants : le meilleur choix pour l'âge de début de l'intervalle ouvert lorsqu'on sait que l'âge est exagéré ; comment traiter les points qui sont situés au-dessus de la droite de régression aux âges élevés, du fait de la baisse de la complétude ; celle-ci peut être due aux migrations associées à la retraite de personnes qui migrent de l'urbain vers le rural où l'enregistrement des décès est moins complet. Enfin, il faut décider s'il convient d'exclure les âges en-dessous de 30 ou 35 ans lorsqu'on ne peut pas corriger l'effet des migrations.

Si la complétude semble être inférieure à 60 %, alors le degré d'incertitude devient grand, et ce point doit être pris en compte pour interpréter les résultats.

Dans la situation où les données sur la population et les décès des ménages sont collectées lors d'un recensement unique, il est tentant de chercher à utiliser des données d'une enquête par sondage pour avoir une répartition par âge de la population à une autre date, antérieure ou postérieure, de manière à pouvoir utiliser la méthode. Mais, pour des raisons qui ne sont pas claires et qui n'ont pas fait l'objet de recherches approfondies, une telle combinaison de sources de données très différentes donne rarement des résultats satisfaisants.

APPLICATION DE LA MÉTHODE

D'un point de vue purement technique, on pourrait appliquer cette méthode à des données présentées par année d'âge, mais ce type de données est sujet à de fortes erreurs sur l'âge, si bien que, en pratique, on travaille le plus souvent avec des données présentées par groupe d'âge de 5 ans. Comme la plupart des données disponibles sont publiées selon ce format, la feuille de calcul correspondante est organisée pour travailler avec des groupes d'âge quinquennaux classiques. Il faut remarquer que John Blacker (Blacker 1988) a montré que ces groupes d'âge classiques sont sensibles aux préférences pour les chiffres ronds ; dans ce cas on pourra adapter la méthode pour travailler avec des groupes d'âge quinquennaux différents, centrés sur les chiffres qui font l'objet de l'attraction (12-17, 18-21 etc.), plutôt qu'avec des groupes d'âge commençant par ces chiffres (15-19, 20-24, etc.).

Étape 1 : Lorsque ce nombre n'est pas immédiatement disponible, estimer le nombre de décès déclarés au cours de la période entre les deux estimations de population

Dans le cas où l'on dispose de données de l'état civil par année de calendrier, l'ajustement consiste à répartir proportionnellement les décès de la première et de la dernière année. On les répartit proportionnellement à la fraction de l'année couverte après le premier recensement et avant le second recensement. Sauf si la distribution des décès par âge change très rapidement au cours de l'année, cette approximation n'aura pas d'effet notable sur les résultats.

Dans le cas où l'on ne dispose pas des décès entre les deux enquêtes, mais si l'intervalle tombe entre deux périodes pour lesquelles on a de telles données (par exemple dans le cas où chaque recensement inclut la question sur les décès des douze derniers mois dans le ménage), alors on peut utiliser la feuille de calcul : *Estimating deaths.xlsx*. Cette feuille de calcul permet d'estimer le nombre de décès entre deux instants donnés d'après les décès qui se sont produits au cours de deux périodes les encadrant. Pour utiliser cette feuille de calcul, on a besoin du nombre de décès répartis par groupes d'âge quinquennaux pour les deux périodes, les dates de début et de fin de chaque période, ainsi que les dates de début et de fin de la période au cours de laquelle on veut estimer le nombre de décès.

Étape 2 : Cumuler la population, les décès et les migrants vers les plus jeunes âges

Pour calculer les taux partiels de natalité, mortalité (et migration), on doit cumuler les effectifs de population, le nombre de décès (et le nombre de migrations nettes), pour les âges au-delà de x . Ainsi, dans le cas de la population, on utilise la formule suivante :

$$N(x+) = \sum_{y=x}^{A-5} {}_5N_y + {}_{\infty}N_A$$

où A désigne l'âge au début du groupe d'âge ouvert. On utilise des formules équivalentes pour calculer les décès cumulés au-delà de l'âge x , soit $D(x+)$. On peut faire de même pour les migrations nettes, quand des données sont disponibles (même si cela est peu probable), et cumuler leur nombre au-delà de l'âge x , soit $NM(x+)$. Lorsque le nombre de migrants est inconnu, on laisse cette colonne en blanc (ou bien on fixe les effectifs à zéro), et la méthode s'utilise en prenant en compte cette omission, comme cela est expliqué plus loin.

Étape 3 : Calcul des personnes-années vécues, $PYL(x+)$

Pour calculer les taux partiels de natalité et de mortalité (ainsi que les taux partiels de migration nette s'ils sont disponibles), on doit calculer les personnes-années vécues au risque. On utilise pour cela la formule suivante :

$$PYL(x+) = (t_2 - t_1) \left({}_{\infty}N_x(t_1) \times {}_{\infty}N_x(t_2) \right)^{1/2}$$

où t_1 est la date du premier recensement, et t_2 la date du second recensement.

Étape 4 : Calcul du nombre de personnes qui atteignent leur x -ème anniversaire, $N(x)$

Le nombre de personnes qui atteignent leur x -ème anniversaire (c'est à dire ceux arrivent, ou sont « nés », dans le groupe d'âge $x+$) dans la population se calcule comme la moyenne géométrique des effectifs de la cohorte au temps t_1 et t_2 divisé par 5, multiplié par la durée de la période entre les deux recensements, exprimée en années, selon la formule suivante :

$$N(x) = \frac{t}{5} \left({}_5N_{x-5}(t_1) \times {}_5N_x(t_2) \right)^{1/2}$$

Étape 5 : Calcul des taux partiels de natalité et de mortalité, $b(x+)$ et $d(x+)$, et de croissance $r(x+)$, corrigés des migrations, $i(x+)$

Les taux partiels de natalité et de mortalité se calculent selon les formules suivantes :

$$b(x+) = \frac{N(x)}{PYL(x+)}$$

$$d(x+) = \frac{D(x+)}{PYL(x+)}$$

Alors, le taux partiel de croissance, moins le taux partiel de migration, se calcule selon la formule suivante:

$$r(x+) - nm(x+) = \frac{{}_{\infty}N_x(t_2) - {}_{\infty}N_x(t_1) - NM(x+)}{PYL(x+)}$$

Étape 6 : Représentation graphique de $b(x+) - r(x+) + nm(x+)$ en fonction de $d(x+)$, et examen visuel pour décider de l'intervalle d'âge sur lequel on ajustera une droite

On commence en prenant pour âge minimal l'âge de 50 ans, et pour âge maximal $A-1$, où A est l'âge du début

de l'intervalle ouvert. On examine ensuite le graphique, et on décide de l'intervalle d'âge sur lequel on ajustera une droite. Lorsque l'exagération des âges au décès est plus forte que celle des âges des survivants dans la population, les points situés à droite du graphique (qui représentent les âges élevés) vont tendre à baisser progressivement avec l'âge en-dessous de la droite. Ceci indique qu'il faut rechercher un âge maximal plus faible, de proche en proche tous les 5 ans, jusqu'à ce que l'effet soit éliminé. De plus, si la valeur des résidus est trop forte aux âges extrêmes (par exemple s'ils excèdent 1 %), alors l'âge maximal devra être abaissé, pour éviter que ces points erronés influencent la pente de la droite ajustée. Mais si l'exagération de l'âge est la même pour la population et pour les décès, alors ceci n'aura pas d'influence sur la pente ni sur l'estimation de la complétude de la déclaration des décès ; par contre, les taux de mortalité par âge seront sous-estimés à ces âges élevés.

Lorsque les points placés sur le graphique correspondants aux âges jeunes (c'est-à-dire la partie gauche du graphique), surtout aux âges entre 15 et 30 ans, dévient notablement de la ligne droite, et si l'on n'a pas introduit de données sur les migrations, ceci indique probablement qu'il y a un niveau important de migrations (sauf s'il y a une couverture du recensement différentielle selon l'âge). Dans ce cas, il convient d'augmenter l'âge minimal du début de l'intervalle utilisé pour ajuster la droite à 30 ou 35 ans, selon celui qui produit le meilleur ajustement aux données.

Étape 7 : Ajustement d'une droite, et estimation de la complétude (c)

Pour estimer la complétude de l'enregistrement des décès par rapport à la population, on commence par tracer un graphique représentant les points d'abscisse : $d(x+)$ et d'ordonnées : $b(x+) - r(x+) + i(x+)$. On calcule les coefficients de la droite liant les deux séries de points par régression orthogonale, selon la formule suivante :

$$b = \frac{\sigma_y}{\sigma_x} = \frac{\frac{1}{n-1} \sum_{i=1}^n (y_i - \mu_y)^2}{\frac{1}{n-1} \sum_{i=1}^n (x_i - \mu_x)^2}$$

et $a = \mu_y - b\mu_x$
où b représente la pente de la droite, et a l'ordonnée à l'origine. La série des ordonnées y_i représente les points : $b(x+) - r(x+) + i(x+)$, la série des abscisses x_i représente les

points $d(x+)$, et les variables μ_y et μ_x représentent, respectivement, les moyennes des deux séries.

Lorsque l'on a tracé la droite de régression, on observe la position des points par rapport à la droite, ainsi que les résidus, de manière à décider quel est le meilleur intervalle d'âge à garder pour calculer la complétude de l'enregistrement des décès. La manière de le faire est discutée plus en détail plus loin, mais on considère qu'il faut exclure les points dont les résidus sont supérieurs à 1 % en valeur absolue. Après exclusion des points aberrants, une nouvelle droite est ajustée, et on calcule de nouvelles valeurs de a et b . En règle générale, on recommande de ne pas terminer à un âge qui se termine par zéro dans les populations qui font preuve d'une forte attraction des chiffres ronds.

On calcule alors la complétude de l'enregistrement des décès, c , d'après les valeurs de a et de b , comme suit. Comme :

$$a = \frac{\ln(k_1 / k_2)}{t_2 - t_1}$$

et

$$b = \frac{(k_1 k_2)^{1/2}}{c}$$

$$\frac{k_1}{k_2} = e^{a(t_2 - t_1)}$$

on fait alors l'hypothèse que le maximum de k_1 et k_2 = 1. Si

$\frac{k_1}{k_2} < 1$, on suppose que $k_2 = 1$, et donc : $k_1 = e^{a(t_2 - t_1)}$ et

si $\frac{k_1}{k_2} > 1$, on suppose que $k_1 = 1$, et donc : $k_2 = e^{-a(t_2 - t_1)}$

et $c = \frac{(k_1 k_2)^{1/2}}{b}$.

Étape 8 : Estimer les taux de mortalité après correction de la complétude des décès déclarés

Afin de calculer les taux de mortalité corrigés, il convient au préalable de corriger les recensements de population en tenant compte de leurs couvertures relatives. Ceci se fait en divisant les effectifs du premier recensement par k_1 et les effectifs du second recensement par k_2 .

Ensuite, on doit corriger les nombres de décès du sous-enregistrement, en divisant les nombres de décès déclarés par l'estimation de la complétude c .

Les personnes-années vécues corrigées, $PYL^a(x,5)$, se calculent de la même manière qu'avant, mais en utilisant les effectifs corrigés du sous-enregistrement relatif des recensements, soit :

$$PYL^a(x,5) = (t_2 - t_1) \left({}_5N_x(t_1) / k_1 \times {}_5N_x(t_2) / k_2 \right)^{1/2}$$

Ensuite, on calcule les taux de mortalité corrigés de tous les biais (complétude de l'enregistrement des décès et couverture des recensements) en divisant les effectifs corrigés des décès et des personnes-années vécues, soit :

$${}_5m_x = \frac{{}_5D_x / c}{PYL^a(x,5)}$$

Remarquons que, d'un point de vue purement technique, on pourrait omettre le coefficient d'ajustement k_1/k_2 et obtenir les mêmes valeurs pour les taux de mortalité corrigés (puisque l'on fait la même correction au numérateur et au dénominateur). Mais dans ce cas, l'estimation de la complétude correspondrait à la moyenne des populations recensées et ignorerait le fait que les recensements ont des couvertures différentes.

Étape 9 : Lissage à l'aide d'une table-type et d'un système relationnel logit

Lorsque les taux de mortalité par âge sont erratiques, on peut les lisser à l'aide d'un modèle. Ceci peut se faire à l'aide du *système relationnel logit* de Brass et d'une table-type du même sexe que l'on considère avoir le même schéma par âge que celui qui s'applique à la population étudiée.

Le classeur Excel correspondant à cette méthode (voir le site web) comporte une feuille de calcul qui permet de produire des taux de mortalité lissés à l'aide d'un *système relationnel logit* et d'une table type de mortalité. L'utilisateur a le choix entre un standard du modèle *Général* de la famille des tables-type des Nations Unies, ou l'un des quatre modèles régionaux des tables-type de Princeton. La transformation logit de ces tables-type de mortalité, ainsi qu'une table-type d'une population affectée par l'épidémie de sida (Timæus 2004) sont donnés dans la feuille de calcul intitulée *Modèles*. Cette feuille de calcul permet aussi à l'utilisateur de choisir comme standard une autre table de mortalité sous forme de logit, s'il y a de bonnes raisons de penser que cette table a un schéma par âge similaire à celui de la population étudiée.

Pour ajuster le modèle logit, on calcule d'abord les quotients quinquennaux de mortalité, ${}_5q_x$, d'après les taux corrigés de mortalité, ${}_5m_x$, selon la formule :

$${}_5q_x = \frac{{}_5m_x}{1 + 2,5{}_5m_x}.$$

Puis on calcule la table de mortalité de proche en proche, en partant d'une racine $l_5 = 1$, selon la formule de récurrence suivante :

$$l_{x+5} = l_x (1 - {}_5q_x)$$

Les coefficients, α et β , du *système relationnel logit* se calculent par régression linéaire, comme :

$$\gamma_x = \alpha + \beta \gamma_x^s$$

où la fonction logit est définie comme :

$$\gamma_x = 0,5 \ln \left(\frac{1 - l_x}{l_x} \right)$$

et où l'exposant s désigne les valeurs de la table de mortalité de référence (le standard).

La table de mortalité lissée se calcule ensuite en appliquant les coefficients α and β comme suit :

$$\gamma_x^{fitted} = \alpha + \beta \gamma_x^s$$

et

$$l_x^{fitted} = \frac{1}{\exp(1 + 2\gamma_x^{fitted})}$$

Les taux de mortalité lissés se calculent dans la table de mortalité comme suit :

$${}_5m_x^{fitted} = \frac{l_x^{fitted} - l_{x+5}^{fitted}}{T_x - T_{x+5}}$$

et

$${}_{\infty}m_x^{fitted} = \frac{l_x^{fitted}}{T_x}$$

où

$$T_x = \sum_{x=x,5}^{\omega} \frac{5}{2} (l_x^{fitted} + l_{x+5}^{fitted})$$

soit :

$$T_x = T_{x+5} + \frac{5}{2} (l_x^{fitted} + l_{x+5}^{fitted})$$

où « ω » désigne l'âge au-delà duquel il ne reste plus de survivants.

EXEMPLE DÉTAILLÉ

Cet exemple utilise les données de la population masculine d'Afrique du Sud, recueillies au recensement de 2001 et à l'enquête de communauté de 2007, et les données des décès enregistrés à l'état civil au cours des années de 2001 à 2007. Le solde migratoire est estimé d'après le changement du nombre de personnes nées à l'étranger repérées aux deux enquêtes, moins une estimation du nombre de sud-africains qui ont émigré à l'étranger entre les deux enquêtes. Cet exemple est détaillé dans le classeur Excel intitulé : *GGB_South Africa_males.xlsx*.

Étape 1 : Calculer le nombre de décès déclarés au cours de la période située entre les deux estimations de population

Les décès déclarés au cours des années 2001 à 2007 pour les hommes sud-africains apparaissent dans le tableau 24.1.

TABLEAU 24.1 Calcul des décès déclarés entre les deux dates des recensements, sud-africains de sexe masculin, 2001-2007

| Âge | 2001 | 2002-2006 | 2007 | Total entre les deux recensements |
|-------|--------|-----------|--------|-----------------------------------|
| 0-4 | 29 005 | 186 346 | 40 314 | 197 912 |
| 5-9 | 2 118 | 14 733 | 2 854 | 15 566 |
| 10-14 | 1 745 | 10 535 | 2 233 | 11 207 |
| 15-19 | 4 470 | 23 857 | 4 860 | 25 473 |
| 20-24 | 8 931 | 51 588 | 10 875 | 54 960 |
| 25-29 | 16 834 | 96 705 | 18 405 | 102 802 |
| 30-34 | 20 892 | 137 355 | 28 245 | 145 588 |
| 35-39 | 21 068 | 137 502 | 29 258 | 145 900 |
| 40-44 | 19 322 | 128 217 | 26 973 | 135 936 |
| 45-49 | 17 881 | 113 891 | 24 761 | 121 010 |
| 50-54 | 16 883 | 104 508 | 22 790 | 111 157 |
| 55-59 | 14 544 | 90 919 | 21 317 | 96 854 |
| 60-64 | 15 097 | 84 351 | 17 410 | 89 930 |
| 65-69 | 13 011 | 77 680 | 17 878 | 82 843 |
| 70-74 | 14 035 | 68 147 | 13 771 | 73 036 |
| 75-79 | 10 846 | 59 859 | 12 534 | 63 871 |
| 80-84 | 9 161 | 44 986 | 8 872 | 48 163 |
| 85+ | 7 602 | 43 233 | 10 009 | 46 196 |

La date de référence du recensement de 2001 était la nuit du 9 au 10 octobre 2001. L'enquête de communauté se déroula sur plusieurs semaines en février, et on prendra

comme date de référence la nuit du 14 au 15 février 2007. Donc, en supposant que les décès sont répartis uniformément au cours de l'année, on peut répartir proportionnellement les décès de 2001 et de 2007, et les ajouter au total des années 2002 à 2006. Ceci donne le total des décès qui se sont produits entre les deux dates de recensement. Par exemple, pour le groupe d'âge 20-24 ans, le nombre se calcule comme suit :

$$\frac{(22 + 30 + 31)}{365} 8931 + 51588 + \frac{(31 + 14)}{365} 10875 = 54960 .$$

Étape 2 : Cumul de la population, des décès et des migrants vers le bas

On cumule alors les effectifs de la population, des décès et des migrants, vers le bas, de l'âge x jusqu'au dernier groupe d'âge (tableau 24.2).

Étape 3 : Calcul des personnes-années vécues, $PYL(x+)$

Le calcul des personnes-années vécues nécessite de connaître d'intervalle de temps entre les deux recensements. Ceci peut se calculer à l'aide de la fonction YEARFRAC dans Excel d'après des dates de référence des deux recensements (soit 5,3541 années). Le décompte des jours entre les deux dates divisé par 365 (soit 5,3507 années) n'est pas identique, mais cette différence a un impact négligeable sur l'estimation de la complétude.

Les personnes-années vécues apparaissent en colonne 2 du tableau 24.3, et se calculent d'après les effectifs cumulés de population qui figurent en colonne 2 et 3 du tableau 24.2. Par exemple, pour l'âge de 20 ans, le calcul est le suivant :

$$PYL(20+) = 5,3541(2099417 \times 2362519)^{\frac{1}{2}} = 67053861 .$$

Étape 4 : Calcul du nombre de personnes qui atteignent l'âge x dans la population, $N(x)$

Les nombres de personnes qui atteignent l'âge x figurent en colonne 3 du tableau 24.3. Par exemple, le nombre de personnes qui atteignent l'âge de 20 ans se calcule d'après les effectifs de population qui figurent en colonne 2 et 3 du tableau 24.1, comme suit :

$$N(20) = \frac{5,3541}{5} (2518985 \times 2553293)^{\frac{1}{2}} = 25778889 .$$

TABLEAU 24.2 Calcul des effectifs cumulés de population, décès et migrants, Afrique du sud, sexe masculin, 2001-2007

| Âge | ${}_5N_x(t_1)$ | ${}_5N_x(t_2)$ | ${}_5D_x$ | ${}_5NM_x$ | $PI(x+)$ | $P2(x+)$ | $D(x+)$ | $NM(x+)$ |
|-----|----------------|----------------|-----------|------------|------------|------------|-----------|----------|
| 0 | 2 223 006 | 2 505 744 | 197 912 | 10 605 | 21 434 045 | 23 348 679 | 1 568 404 | 128 946 |
| 5 | 2 425 066 | 2 560 642 | 15 566 | 2 848 | 19 211 039 | 20 842 935 | 1 370 492 | 118 341 |
| 10 | 2 518 985 | 2 452 339 | 11 207 | 5 153 | 16 785 973 | 18 282 293 | 1 354 926 | 115 492 |
| 15 | 2 453 156 | 2 553 293 | 25 473 | 16 574 | 14 266 988 | 15 829 955 | 1 343 719 | 110 339 |
| 20 | 2 099 417 | 2 362 519 | 54 960 | 14 803 | 11 813 832 | 13 276 662 | 1 318 246 | 93 766 |
| 25 | 1 899 275 | 2 033 165 | 102 802 | 4 714 | 9 714 415 | 10 914 143 | 1 263 286 | 78 963 |
| 30 | 1 594 624 | 1 875 483 | 145 588 | 13 331 | 7 815 140 | 8 880 977 | 1 160 484 | 74 249 |
| 35 | 1 441 657 | 1 548 185 | 145 900 | 9 693 | 6 220 516 | 7 005 495 | 1 014 896 | 60 918 |
| 40 | 1 233 813 | 1 306 900 | 135 936 | 7 464 | 4 778 859 | 5 457 310 | 868 996 | 51 225 |
| 45 | 967 744 | 1 104 294 | 121 010 | 8 719 | 3 545 046 | 4 150 410 | 733 060 | 43 761 |
| 50 | 769 627 | 888 042 | 111 157 | 9 413 | 2 577 302 | 3 046 116 | 612 050 | 35 042 |
| 55 | 552 402 | 708 812 | 96 854 | 4 640 | 1 807 675 | 2 158 074 | 500 893 | 25 629 |
| 60 | 444 592 | 491 871 | 89 930 | 5 081 | 1 255 273 | 1 449 261 | 404 039 | 20 989 |
| 65 | 304 835 | 394 305 | 82 843 | 4 922 | 810 681 | 957 391 | 314 108 | 15 908 |
| 70 | 232 604 | 241 976 | 73 036 | 4 334 | 505 846 | 563 086 | 231 266 | 10 986 |
| 75 | 136 466 | 163 112 | 63 871 | 2 980 | 273 242 | 321 110 | 158 229 | 6 652 |
| 80 | 90 856 | 87 698 | 48 163 | 1 662 | 136 776 | 157 998 | 94 359 | 3 672 |
| 85 | 45 920 | 70 299 | 46 196 | 2 009 | 45 920 | 70 299 | 46 196 | 2 009 |

TABLEAU 24.3 Calcul des effectifs cumulés de la population, des décès et des migrants, Afrique du sud, sexe masculin, 2001-2007

| Âge | $PYL(x+)$ | $N(x)$ | $b(x+)$ | $r(x+) - i(x+)$ | $d(x+) = X$ | $b(x+) - r(x+) + i(x+) = Y$ | $a + bx$ | Résidus $y - (a + bx)$ |
|-----|-------------|-----------|---------|-----------------|-------------|-----------------------------|----------|---------------------------|
| 0 | 119 775 275 | | | #N/A | 0,00000 | | -0,0047 | |
| 5 | 107 136 837 | 2 554 810 | 0,02385 | 0,01413 | 0,01279 | 0,00972 | 0,0093 | 0,0004 |
| 10 | 93 793 458 | 2 611 355 | 0,02784 | 0,01472 | 0,01445 | 0,01312 | 0,0111 | 0,0020 |
| 15 | 80 461 835 | 2 715 670 | 0,03375 | 0,01805 | 0,01670 | 0,01570 | 0,0135 | 0,0022 |
| 20 | 67 053 861 | 2 577 889 | 0,03845 | 0,02042 | 0,01966 | 0,01803 | 0,0168 | 0,0013 |
| 25 | 55 129 886 | 2 212 329 | 0,04013 | 0,02033 | 0,02291 | 0,01980 | 0,0203 | -0,0005 |
| 30 | 44 604 915 | 2 020 991 | 0,04531 | 0,02223 | 0,02602 | 0,02308 | 0,0237 | -0,0006 |
| 35 | 35 344 071 | 1 682 498 | 0,04760 | 0,02049 | 0,02871 | 0,02712 | 0,0266 | 0,0005 |
| 40 | 27 342 320 | 1 469 826 | 0,05376 | 0,02294 | 0,03178 | 0,03082 | 0,0300 | 0,0008 |
| 45 | 20 537 160 | 1 249 916 | 0,06086 | 0,02735 | 0,03569 | 0,03352 | 0,0343 | -0,0007 |
| 50 | 15 001 678 | 992 684 | 0,06617 | 0,02891 | 0,04080 | 0,03726 | 0,0398 | -0,0026 |
| 55 | 10 574 924 | 790 897 | 0,07479 | 0,03071 | 0,04737 | 0,04408 | 0,0470 | -0,0029 |
| 60 | 7 221 483 | 558 171 | 0,07729 | 0,02396 | 0,05595 | 0,05334 | 0,0564 | -0,0030 |
| 65 | 4 716 866 | 448 343 | 0,09505 | 0,02773 | 0,06659 | 0,06732 | 0,0680 | -0,0006 |
| 70 | 2 857 463 | 290 826 | 0,10178 | 0,01619 | 0,08093 | 0,08559 | 0,0836 | 0,0020 |
| 75 | 1 585 932 | 208 577 | 0,13152 | 0,02599 | 0,09977 | 0,10553 | 0,1041 | 0,0014 |
| 80 | 787 071 | 117 144 | 0,14884 | 0,02230 | 0,11989 | 0,12654 | 0,1261 | 0,0005 |
| 85 | 304 201 | | | | | | | |

Étape 5 : Calcul des taux partiels de natalité et de mortalité, $b(x+)$ et $d(x+)$, et du taux partiel de croissance $r(x+)$ après correction des migrations, $i(x+)$

Les taux partiels de natalité et de mortalité figurant en colonne 4 et 6 du tableau 24.3. Par exemple, pour l'âge de 20 ans, les taux partiels de natalité et de mortalité se calculent à partir des naissances partielles (colonne 3 du tableau 24.3) et des décès partiels (colonne 8 du tableau 24.2) comme suit :

$$b(20+) = \frac{2577889}{67053861} = 0,03845$$

$$d(20+) = \frac{1318246}{67053861} = 0,01966$$

Le taux partiel de croissance, diminué du taux partiel de migration nette, figure en colonne 5 du tableau 24.3. Par exemple, pour l'âge de 20 ans, il se calcule à partir des populations cumulées qui figurent en colonne 2 et 3 du

tableau 24.3, et des migrations nettes cumulées qui figurent dans la dernière colonne du tableau 24.2, comme suit :

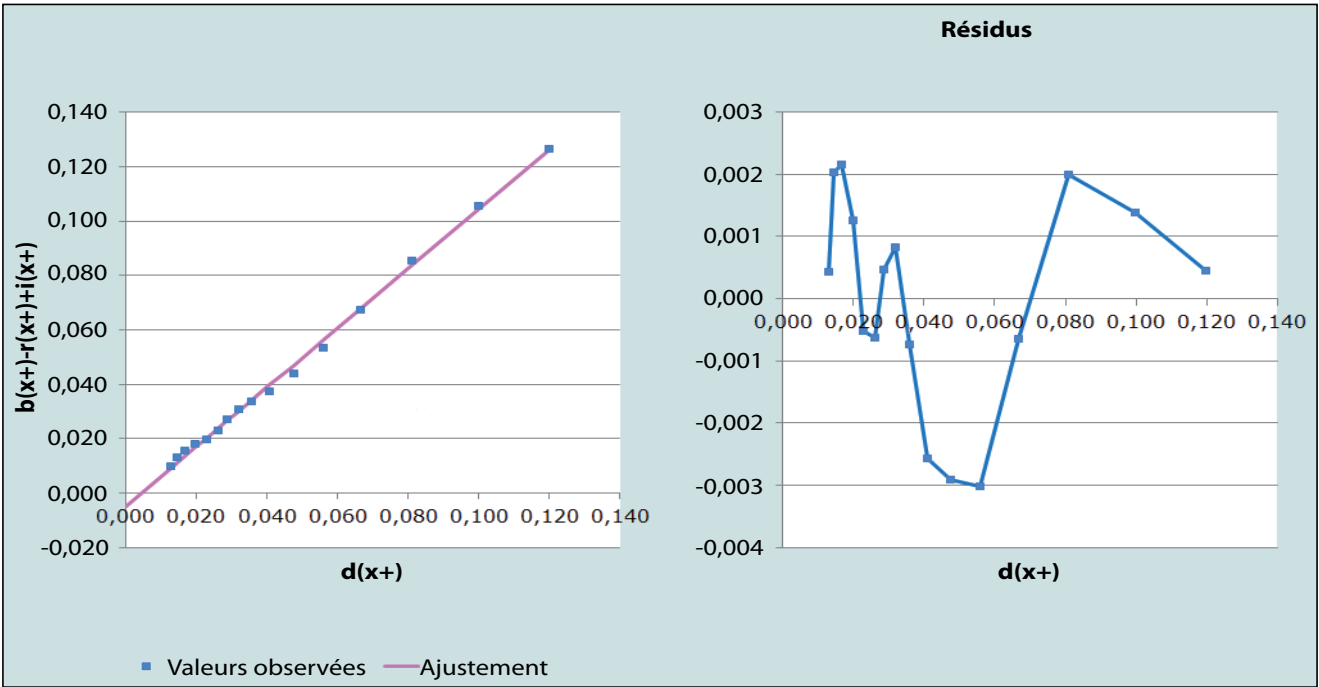
$$r(20+) - i(20+) = \frac{13276662 - 11813832 - 93766}{67053861} = 0,02042 .$$

Étape 6 : Représentation graphique de $b(x+) - r(x+) + i(x+)$ en fonction de $d(x+)$, et analyse de la figure pour décider quel intervalle d'âges choisir pour ajuster la droite

Pour tracer le graphique et ajuster la droite liant tous les points, on commence par fixer l'âge minimal à 5 ans, et l'âge maximal à 84 ans (car, dans cet exemple, l'intervalle d'âge ouvert est de 85 ans et plus). La figure 24.1 donne le graphique des points de coordonnées : $b(x+) - r(x+) + i(x+)$ et $d(x+)$.

L'analyse visuelle des points de ce graphique (figure 24.1) montre que ceux-ci sont assez bien alignés,

FIGURE 24.1 Graphique de diagnostic, Afrique du sud, sexe masculin, 2001-2007



et assez proches de la ligne de régression, ce qui indique que l'essentiel des migrations a bien été pris en compte. Dans ce cas, il n'y a pas de raison de changer l'intervalle d'âge sur lequel la droite est ajustée. Et donc, comme on peut s'y attendre, augmenter l'âge minimal n'aurait que peu d'effet sur l'estimation de la complétude, qui est ici de 91 %. De même, l'estimation de la complétude ne serait guère affectée en diminuant l'âge maximal, c'est-à-dire en excluant le dernier ou les deux derniers points, même si les résultats sont un peu affectés par une baisse de la complétude aux âges élevés (voir l'application de la méthode d'extinction des cohortes synthétiques). Par contre, éliminer d'autres points conduirait à augmenter la complétude et atteindre des niveaux irréalistes, ce qui suggère un problème de qualité des données (probablement les données de population).

Étape 7 : Ajustement de la droite, et estimation de la complétude, c

Les coefficients de la droite ajustant au mieux les points de la Figure 24.1 se calculent comme suit :

$$b = \frac{\sigma_y}{\sigma_x} = \frac{0,03483}{0,03193} = 1,0907$$

$$a = 0,0444 - 1,0907 \times 0,045 = -0,00467$$

et la couverture relative des deux recensements se calcule comme suit :

$$\frac{k_1}{k_2} = \exp(-0,00467 \times 5,3541) = 0,9753.$$

Comme $k_2 > k_1$ on considère que $k_2 = 1$ et donc que $k_1 = 0,9753$ (c'est-à-dire que le décompte du premier recensement est sous-estimé de 2,5 % par rapport au second).

La complétude de la déclaration des décès, c , est estimée à 91 % (relativement au décompte de l'enquête de 2007), et se calcule comme suit :

$$c = \frac{(0,9753)^{1/2}}{1,0907} = 0,9054.$$

Étape 8 : Calcul des taux de mortalité corrigés de la sous-déclaration des décès

On corrige d'abord la population au premier recensement, en divisant la population recensée, qui figure en colonne 2 du tableau 24.2, par k_1 . Par exemple, pour l'âge de 20 ans, la population corrigée se calcule comme :

$$\frac{2099417}{0,9753} = 2152629.$$

La population corrigée au second recensement se calcule de même, comme la population recensée, qui figure en colonne 3 du tableau 24.2, divisée par k_2 . Comme, par hypothèse, on a pris $k_2 = 1$, ces nombres sont donc les mêmes que ceux qui figurent en colonne 3 du tableau 24.2.

Ensuite, les décès sont corrigés pour tenir compte de la sous-déclaration, en divisant les nombres de décès déclarés dans chaque groupe d'âge, qui figurent en colonne 4 du tableau 24.2, par l'estimation de la complétude. Ces nombres figurent en colonne 4 du tableau 24.4. Par exemple, pour l'âge de 20 ans, le nombre se calcule à partir du nombre de décès déclarés : 54 960 décès, comme suit :

$$\frac{54960}{0,9054} = 60701.$$

Les personnes-années vécues corrigées (colonne 5 du tableau 24.4) se calculent comme la moyenne géométrique des populations qui figurent en colonne 2 et 3 du tableau 24.4, multipliées par la durée (en années) de la période intercensitaire, qui est de 5,3541 années. Par exemple, pour l'âge de 20 ans, cela donne :

$$PYL(20,5) = 5,3541(2152629 \times 2362519)^{1/2} = 12074140.$$

Les taux de mortalité corrigés du sous-enregistrement des décès (colonne 6 du tableau 24.4) se calculent en divisant les décès corrigés par les personnes-années vécues corrigées. Par exemple, pour le groupe d'âge de 20-24 ans, le taux de mortalité corrigé se calcule comme suit :

$$\frac{60701}{12074140} = 0,0050.$$

Étape 9 : Lissage en utilisant une table-type et un système relationnel logit

Les quotients quinquennaux de mortalité, corrigés du sous-enregistrement, qui figurent en colonne 2 du tableau 24.5, se calculent à partir des taux de mortalité corrigés, qui figuraient en colonne 6 du tableau 24.4. Par exemple, la probabilité pour un homme de décéder entre 20 et 25 ans, se calcule comme suit :

$${}_5q_{20} = \frac{5 \times 0,005}{1 + 2,5 \times 0,005} = 0,0248.$$

| Âge | Valeurs corrigées du sous-enregistrement | | | | |
|-----|--|----------------|-----------|------------|-----------|
| | ${}_5N_x(t_1)$ | ${}_5N_x(t_2)$ | ${}_5D_x$ | $PYL(x,5)$ | ${}_5m_x$ |
| 0 | | | | | |
| 5 | 2 486 532 | 2 560 642 | 17 193 | 13 510 001 | 0,0013 |
| 10 | 2 582 831 | 2 452 339 | 12 378 | 13 474 797 | 0,0009 |
| 15 | 2 515 334 | 2 553 293 | 28 134 | 13 568 508 | 0,0021 |
| 20 | 2 152 629 | 2 362 519 | 60 701 | 12 074 140 | 0,0050 |
| 25 | 1 947 414 | 2 033 165 | 113 541 | 10 653 675 | 0,0107 |
| 30 | 1 635 041 | 1 875 483 | 160 796 | 9 375 725 | 0,0172 |
| 35 | 1 478 197 | 1 548 185 | 161 141 | 8 099 564 | 0,0199 |
| 40 | 1 265 085 | 1 306 900 | 150 136 | 6 884 383 | 0,0218 |
| 45 | 992 273 | 1 104 294 | 133 651 | 5 604 563 | 0,0238 |
| 50 | 789 134 | 888 042 | 122 768 | 4 482 045 | 0,0274 |
| 55 | 566 403 | 708 812 | 106 972 | 3 392 442 | 0,0315 |
| 60 | 455 861 | 491 871 | 99 325 | 2 535 277 | 0,0392 |
| 65 | 312 561 | 394 305 | 91 497 | 1 879 609 | 0,0487 |
| 70 | 238 500 | 241 976 | 80 666 | 1 286 217 | 0,0627 |
| 75 | 139 925 | 163 112 | 70 543 | 808 863 | 0,0872 |
| 80 | 93 159 | 87 698 | 53 194 | 483 940 | 0,1099 |
| 85 | 47 084 | 70 299 | 51 021 | 308 032 | 0,1656 |

TABLEAU 24.4 Calcul des taux de mortalité corrigés, Afrique du Sud, sexe masculin, 2001-2007

TABLEAU 24.5 Calcul des taux de mortalité lissés à l'aide d'une table-type et d'un système relationnel logit, Afrique du sud, sexe masculin, 2001-2007

| Âge | ${}_5q_x$ | l_x/l_5 | $Y(x)$ obs. | Sida Cdn. $l^s(x)$ | Logit standard $Y^s(x)$ | Logit lissé $Y(x)$ | Surviv. lissé $l(x)$ | $T(x)$ | Taux lissé ${}_5m_x$ |
|-----|-----------|-----------|-------------|-----------------------|----------------------------|--------------------|----------------------|--------|----------------------|
| 0 | | | | | | | | | |
| 5 | 0,0063 | 1 | | 1,0000 | | | 1 | 50,699 | 0,0032 |
| 10 | 0,0046 | 0,9937 | - 2,5270 | 0,9785 | - 1,9081 | - 2,0551 | 0,9839 | 45,739 | 0,0030 |
| 15 | 0,0103 | 0,9891 | - 2,2542 | 0,9632 | - 1,6326 | - 1,7292 | 0,9695 | 40,856 | 0,0025 |
| 20 | 0,0248 | 0,9789 | - 1,9186 | 0,9512 | - 1,4853 | - 1,5550 | 0,9573 | 36,039 | 0,0043 |
| 25 | 0,0519 | 0,9546 | - 1,5229 | 0,9324 | - 1,3120 | - 1,3500 | 0,9370 | 31,303 | 0,0090 |
| 30 | 0,0822 | 0,9051 | - 1,1273 | 0,8969 | - 1,0818 | - 1,0777 | 0,8962 | 26,722 | 0,0159 |
| 35 | 0,0948 | 0,8306 | - 0,7951 | 0,8420 | - 0,8365 | - 0,7875 | 0,8285 | 22,415 | 0,0206 |
| 40 | 0,1034 | 0,7519 | - 0,5544 | 0,7794 | - 0,6311 | - 0,5446 | 0,7482 | 18,482 | 0,0241 |
| 45 | 0,1125 | 0,6742 | - 0,3636 | 0,7148 | - 0,4593 | - 0,3414 | 0,6644 | 14,964 | 0,0244 |
| 50 | 0,1282 | 0,5983 | - 0,1992 | 0,6560 | - 0,3228 | - 0,1799 | 0,5890 | 11,848 | 0,0234 |
| 55 | 0,1461 | 0,5216 | - 0,0433 | 0,6048 | - 0,2127 | - 0,0497 | 0,5248 | 9,084 | 0,0258 |
| 60 | 0,1784 | 0,4454 | 0,1097 | 0,5530 | - 0,1064 | 0,0760 | 0,4621 | 6,640 | 0,0337 |
| 65 | 0,2170 | 0,3659 | 0,2749 | 0,4918 | 0,0163 | 0,2212 | 0,3912 | 4,530 | 0,0503 |
| 70 | 0,2711 | 0,2865 | 0,4562 | 0,4119 | 0,1781 | 0,4125 | 0,3047 | 2,814 | 0,0717 |
| 75 | 0,3580 | 0,2089 | 0,6659 | 0,3178 | 0,3819 | 0,6536 | 0,2130 | 1,542 | 0,1008 |
| 80 | 0,4311 | 0,1341 | 0,9327 | 0,2173 | 0,6408 | 0,9598 | 0,1279 | 0,708 | 0,1470 |
| 85 | #N/A | 0,0763 | 1,2470 | 0,1201 | 0,9959 | 1,3799 | 0,0595 | 0,252 | 0,2081 |

Les survivants de la table de mortalité à l'âge $x + 5$ se calculent à partir des survivants à l'âge x à partir des quotients quinquennaux, et figurent en colonne 3 du tableau 24.5. Par exemple, la proportion de survivants à l'âge de 25 ans se calcule comme suit :

$$l_{25} = 0,9789(1 - 0,0248) = 0,9546.$$

La transformation logit appliquée aux proportions de survivants figure en colonne 4 du tableau 24.5. Par exemple, à l'âge de 20 ans, le logit de l_{20} se calcule comme suit :

$$\gamma_{20} = 0,5 \ln \left(\frac{1 - 0,9789}{0,9789} \right) = -1,919.$$

La table-type utilisée est une table-type qui tient compte du sida, avec une espérance de vie de $e_0 = 50$ ans. Les logits des survivants de la table-type conditionnelle (qui figurent en colonne 5 du tableau 24.5) apparaissent en colonne 6 du tableau 24.5. Comme on peut le voir sur la figure 24.2, la table-type qui tient compte du sida n'ajuste pas bien les données, mais elle marche mieux que toute autre table-type qui ne tient pas compte de l'impact du sida sur la mortalité.

Les coefficients, α et β , sont déterminés comme la pente et l'ordonnée à l'origine de la droite qui ajuste au mieux les transformations logit, qui figurent en colonnes 4 et 6 du tableau 24.5, dans l'intervalle d'âge choisi par l'utilisateur (entre 45 et 80 ans dans cet exemple), soit ici 0,2119 et 1,1893 respectivement.

On applique ensuite ces coefficients aux logits de la table de mortalité conditionnelle, afin de produire les logits lissés (colonne 7 du tableau 24.5). Par exemple, pour l'âge de 20 ans, le logit lissé se calcule comme suit :

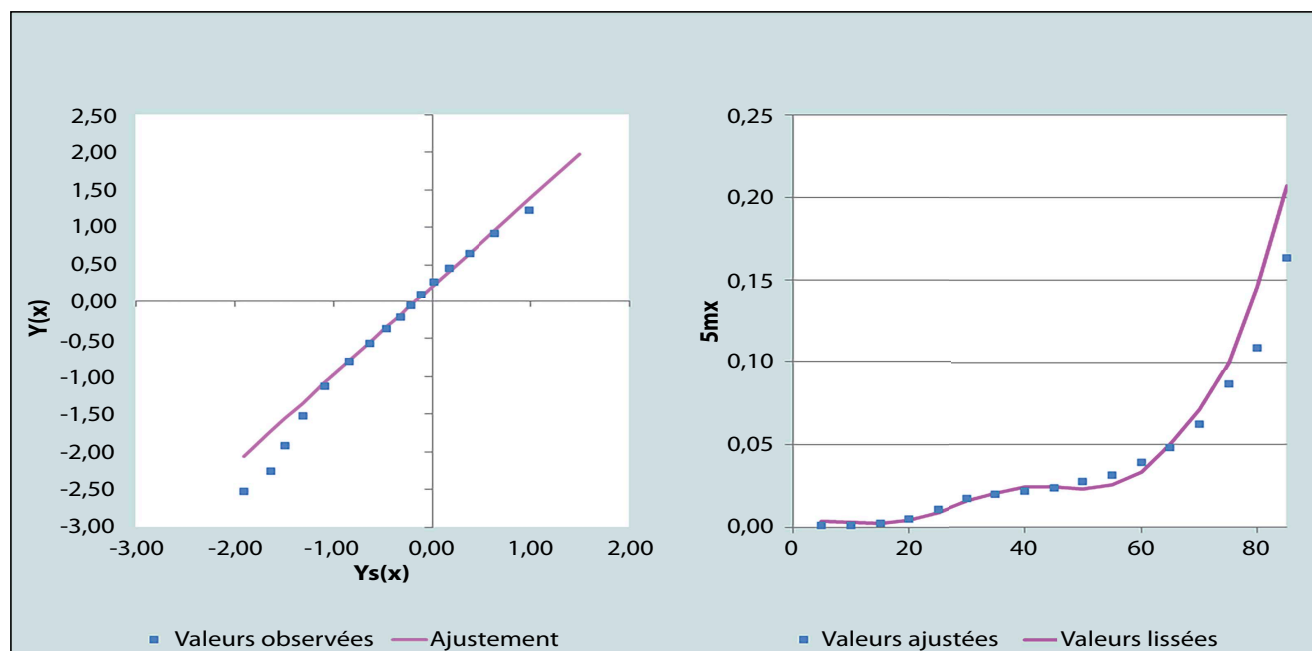
$$\gamma_{20}^{fitted} = 0,2119 + 1,1893 \times (-1,4853) = -1,555.$$

Ces valeurs sont ensuite prises pour calculer la table de mortalité lissée (colonne 8 du tableau 24.5). Par exemple, la probabilité de survie à 20 ans se calcule comme suit :

$$l_{20}^{fitted} = \frac{1}{1 + \exp(2 \times (-1,555))} = 0,9573.$$

Les personnes-années vécues cumulées au-delà de l'âge x , soit T_x , figurent en colonne 9 du tableau 24.5, et se calculent à partir de la table de mortalité lissée. Ces valeurs

FIGURE 24.2 Logits des proportions de survivants et taux de mortalité obtenus après correction, comparés aux valeurs correspondantes de la table-type de mortalité obtenue en tenant compte du sida, Afrique du sud, sexe masculin, 2001-2007



sont alors utilisées pour calculer les taux de mortalité lissés, qui figurent en colonne 10 du tableau 24.5. Par exemple, à l'âge de 80 ans :

$$T_{80} = 0,2523 + \frac{5}{2}(0,1248 + 0,0577) = 0,7085$$

$${}_5m_{80}^{fitted} = \frac{0,1248 - 0,0577}{0,7085 - 0,2523} = 0,1470 .$$

DIAGNOSTICS, ANALYSE ET INTERPRÉTATION

Contrôles et validation

L'estimation de la complétude est de 91 %. La première vérification à faire sur ce résultat est la comparaison avec le résultat correspondant pour l'autre sexe. Ainsi, en appliquant la même méthode décrite pour les hommes aux données concernant les femmes au cours de la même période (voir la feuille de calcul : *GGB_South Africa_females*) donne une estimation de la complétude de 89 %. Des recherches antérieures (Dorrington, Moultrie and Timæus 2004) ont montré que l'on peut s'attendre à des résultats voisins pour les deux sexes, et le fait que ces résultats le sont en effet valide par conséquent ces estimations.

Une seconde vérification des résultats consiste à les comparer avec ceux obtenus par la Méthode de l'extinction des cohortes synthétiques (voir la feuille de calcul : *SEG_South Africa_males*). Cette méthode donne une estimation de la complétude de 94 % sur l'intervalle d'âge de 5 à 84 ans, ce qui est aussi assez proche pour valider les résultats précédents.

Une troisième vérification consiste à comparer diverses estimations de la mortalité avec d'autres sources, telles que des estimations antérieures pour le pays, ou les *Perspectives de la population mondiale* (UN Population Division 2011). L'estimation du quotient de mortalité de 15 à 60 ans, ${}_{45}q_{15}$, est de 523 pour 1 000 après correction pour tenir compte de la sous-déclaration des décès, une valeur très proche des 528 pour 1 000 retenue par les *Perspectives de la population mondiale* pour la période 2000-2005, ce qui une fois encore montre qu'il y a peu de raisons de mettre en question les résultats obtenus par la méthode.

Il est intéressant de remarquer que l'application de la Méthode de Brass de la balance de l'accroissement démographique à ces mêmes données, en prenant pour population moyenne la population au milieu de l'intervalle entre les deux recensements, donne une estimation de 85 % lors-

qu'on utilise le même intervalle d'âge. Si on augmente l'âge minimal à 35 ans pour ajuster la droite, on augmente l'estimation à 88 %, toujours un peu plus faible que l'estimation de 91 % produite par la méthode utilisée ci-dessus.

Interprétation

Comme il a déjà été mentionné plus haut, tous les derniers intervalles d'âge entre 75 et plus et 85 ans et plus retenus pour ajuster la droite fournissent des estimations équivalentes de la complétude de l'enregistrement des décès. Par contre, si on s'arrête avant 75 ans et plus, l'estimation augmente, et atteint 100 % pour 70 ans et plus, 105 % pour 65 ans et plus, et 108 % pour 60 ans et plus. Bien qu'il soit probable que le recensement de 2001 et l'enquête de 2007 sous-estiment le nombre d'hommes dans la population, il est vraisemblable que la sous-estimation soit concentrée chez les jeunes adultes, mais improbable qu'elle soit suffisamment forte pour produire une estimation supérieure à 100 % pour la complétude de l'enregistrement des décès par rapport au recensement. De plus, toutes choses égales par ailleurs, plus le seuil du groupe d'âge ouvert est bas, moins les estimations de la complétude sont robustes. En conséquence, on préférera les estimations plus modestes obtenues avec un groupe d'âge ouvert commençant à un âge plus élevé.

PROBLÈMES D'INTERPRÉTATION SPÉCIFIQUES DE LA MÉTHODE

Source des données sur les décès déclarés

On trouve en général deux types de problèmes avec les données concernant les décès : ceux qui conduisent à un biais de déclaration (sous- ou sur-déclaration) constant par âge, ce qui est précisément ce que la méthode vise à corriger, et ceux qui conduisent à des biais de déclaration qui varient selon l'âge, qui peuvent entraîner une distorsion dans les estimations finales. Bien que l'approche générale reste essentiellement la même quelle que soit la source des données, des sources différentes peuvent être sujettes à des biais spécifiques, qui peuvent avoir un effet sur l'interprétation des résultats. On va illustrer ces cas par des exemples concrets, et d'une manière générale, l'analyse devra prêter attention aux biais suivants dans les données concernant les décès.

(i) Décès déclarés à l'état civil

Si la répartition de la population entre les zones urbaines et rurales (ou une division équivalente) diffère significati-

vement par âge, et que la complétude de la déclaration des décès en milieu urbain est significativement supérieure à celle du milieu rural, alors l'hypothèse que la complétude est indépendante de l'âge sera vraisemblablement violée, et la complétude va tendre à baisser avec l'âge au-delà de 50 ans si la proportion de personnes qui migrent de l'urbain vers le rural au moment de leur retraite. Si on l'ignore, cette violation de l'hypothèse produira probablement une sous-estimation du niveau moyen de la complétude.

(ii) Décès déclarés par les ménages

Ces données sont sujettes à quatre problèmes potentiels :

- Si une proportion significative des ménages sont dissous après le décès d'une personne clé (comme la seule personne à subvenir aux besoins du ménage), alors les décès de ces personnes risquent de ne pas être déclarés, ce qui conduit à violer l'hypothèse que la complétude est constante selon l'âge. Si une proportion significative des décès dans certains groupes d'âge sont relatifs à des individus qui ne vivent pas dans des ménages ordinaires (par exemple, ceux qui vivent dans des maisons pour personnes âgées), la violation de l'hypothèse peut même être encore plus sévère. Mais ce n'est pas encore un problème dans la plupart des pays en développement.
- Dans les situations où les jeunes adultes quittent leur famille lorsqu'ils grandissent, pour aller vivre en ville, il est possible qu'ils soient considérés comme membres de plusieurs ménages (ou d'aucun ménage), et leurs décès peuvent être déclarés plus d'une fois (ou jamais), ce qui conduit encore à une violation de l'hypothèse de déclaration des décès constante selon l'âge. Dans ce cas, on peut en limiter l'impact en ignorant les données concernant les décès d'âge inférieur à un âge donné pour estimer la complétude.
- Erreurs sur la période de référence. Comme il arrive fréquemment qu'il y ait confusion sur la période précise au cours de laquelle on doit déclarer les décès, sans même parler d'erreurs sur la date précise du décès, il est possible que cela se traduise par soit une sous-déclaration, soit une sur-déclaration des décès. En supposant que l'on puisse faire l'hypothèse que ces biais sont indépendants de l'âge des décédés, cette distorsion sera prise en compte dans l'estimation de la complétude, et ne constituera pas un problème pour le calcul des taux de mortalité corrigés.
- La période de référence ne couvre qu'une petite partie de la période intercensitaire. Par exemple, on a fréquemment

le cas suivant, où l'on demande aux ménages de déclarer les décès au cours des 12 derniers mois précédant le recensement. Non seulement une période très courte comme celle-ci produit de fortes fluctuations aléatoires, mais de plus le problème qui se pose est d'estimer la population au début de cette période de référence. On illustrera ce point dans les exemples ci-dessous, et on expliquera comment résoudre le problème. Par contre, si en plus on dispose des décès déclarés par les ménages au premier recensement, on peut utiliser les deux ensembles de données sur les décès pour estimer le nombre de décès qui sont produit entre les deux recensements, comme cela a été expliqué ci-dessus. Mais, comme la question sur les décès des 12 derniers mois au sein des ménages n'a été posée que relativement peu fréquemment avant le cycle des recensements des années 2010, il se peut que l'on ait un seul ensemble de données sur les décès. Dans ce cas, en supposant qu'il n'y ait pas de raison de présumer que le schéma par âge de mortalité ait changé rapidement au cours de la période, on recommande de calculer les taux de mortalité par âge pour l'année concernée, et d'appliquer ces taux aux personnes-années vécues au cours de l'intervalle intercensitaire, de manière à obtenir une estimation des décès au cours de la période. Mais si l'on a de bonnes raisons de penser que la mortalité a changé rapidement, par exemple du fait de l'épidémie de VIH/sida, alors cette adaptation va vraisemblablement biaiser les résultats, soit en sous-estimant, soit en surestimant la mortalité, et dans ce cas on ne recommande pas d'utiliser cette méthode de répartition des décès.

(iii) Décès enregistrés dans les centres de santé

On sait peu de choses à propos de cette source de données, ni comment elle fonctionne. On peut cependant supposer que la complétude dépendra de la répartition géographique des centres de santé auprès desquels les données seront collectées. Dans de nombreux pays en développement, ce type de services a tendance à être concentré dans les zones urbaines. Alors, une nouvelle fois, si la proportion de la population vivant en milieu urbain plutôt qu'en milieu rural varie selon l'âge, on ne pourra pas supposer que la complétude est indépendante de l'âge. Il est aussi possible que certaines causes de décès soient plus fréquentes dans les centres de santé, et si ces causes sont importantes en nombre et qu'elles sont liées à l'âge, tout ceci pourrait entraîner une nouvelle violation de l'hypothèse de la complétude constante selon l'âge.

EXEMPLES UTILISANT LES DÉCÈS DÉCLARÉS PAR LES MÉNAGES DANS LES RECENSEMENTS ET ENQUÊTES

Les exemples présentés ci-dessous utilisent les mêmes données de population que celles qui sont présentées dans les classeurs Excel : *GGB_South Africa_males* et *GGB_South Africa_females*, mais on a utilisé les données des décès des 12 derniers mois déclarés par les ménages au recensement de 2001 et à l'enquête de communauté de 2007, au lieu d'utiliser les données de l'état civil pour la période 2001-2007. Ces données figurent dans le tableau 24.6.

TABLEAU 24.6 Décès déclarés par les ménages au-cours des 12 mois précédant le recensement ou l'enquête, Afrique du sud

| Âge | Recensement de 2001 | | Enquête de communauté de 2007 | |
|-------|---------------------|--------------|-------------------------------|--------------|
| | Sexe masculin | Sexe féminin | Sexe masculin | Sexe féminin |
| 0-4 | 35 873 | 32 096 | 48 322 | 44 418 |
| 5-9 | 3 868 | 3 155 | 4 505 | 5 216 |
| 10-14 | 2 590 | 2 284 | 3 442 | 3 259 |
| 15-19 | 5 628 | 5 122 | 8 246 | 7 878 |
| 20-24 | 10 976 | 13 246 | 16 360 | 21 702 |
| 25-29 | 17 787 | 19 727 | 27 551 | 35 840 |
| 30-34 | 20 038 | 18 292 | 34 832 | 42 576 |
| 35-39 | 19 816 | 15 521 | 38 061 | 34 809 |
| 40-44 | 17 417 | 12 124 | 33 604 | 28 823 |
| 45-49 | 15 840 | 10 105 | 27 829 | 20 973 |
| 50-54 | 15 077 | 9 144 | 28 223 | 18 891 |
| 55-59 | 12 781 | 7 755 | 22 868 | 13 118 |
| 60-64 | 13 428 | 10 367 | 18 775 | 14 912 |
| 65-69 | 11 820 | 10 195 | 17 532 | 14 298 |
| 70-74 | 11 885 | 10 809 | 14 879 | 14 645 |
| 75-79 | 8 794 | 8 393 | 12 966 | 14 151 |
| 80-84 | 7 484 | 9 371 | 9 204 | 12 063 |
| 85+ | 7 115 | 12 389 | 11 735 | 18 178 |

Les effectifs de décès qui se sont produits entre la date du recensement (nuit du 9 au 10 octobre 2001) et la date de l'enquête de communauté (supposée être la nuit du 14 au 15 février 2007) sont estimés dans les feuilles de calcul suivantes : *Estimating deaths_South Africa_males_hhd* et *Estimating deaths_South Africa_females_hhd*.
L'application de la méthode généralisée de la balance de l'accroissement démographique à ces données pour les

hommes est fournie dans la feuille de calcul : *GGB_South Africa_males_hhd*. Les résultats suggèrent que les décès sont approximativement aussi bien déclarés que ceux de l'état civil. Cependant, l'estimation du quotient de mortalité de 15 à 60 ans, $_{45}q_{15}$, est de 548 pour 1 000, ce qui est un peu plus élevé que l'estimation produite par les données de l'état civil (523 pour 1 000). Mais lorsqu'on applique la méthode généralisée de la balance de l'accroissement démographique aux données du sexe féminin, (voir le classeur Excel *GGB_South Africa_females_hhd*), on trouve une complétude des décès de femmes déclarés par les ménages inférieure à celle obtenue par les décès de l'état civil. On trouve donc une estimation du quotient de mortalité de 15 à 60 ans, $_{45}q_{15}$, de 509 pour 1 000, bien supérieure aux 420 pour 1 000 obtenus par les décès de l'état civil, et bien moins plausible si on la compare au quotient obtenu pour les hommes.
On peut rechercher les raisons pour lesquelles la méthode semble moins bien marcher lorsqu'elle est appliquée aux décès de sexe féminin déclarés par les ménages en comparant les effectifs de décès calculés d'après les décès déclarés par les ménages aux effectifs de décès attendus par ceux de l'état civil après correction du sous-enregistrement, comme cela apparaît dans le tableau 24.7. Cette comparaison montre que le nombre de décès de femmes de 55 ans et plus est nettement plus faible. Ceci pourrait être dû au fait que les ménages sont dissous après le décès de ces femmes, car il est fréquent que les femmes qui décèdent à cet âge aient été le chef du ménage.
La comparaison suggère aussi un sur-enregistrement des décès d'hommes de moins de 30 ans et de femmes de moins de 25 ans. Ceci pourrait être dû au fait que leurs décès sont déclarés dans plusieurs ménages, ou au fait que l'état civil est moins complet chez les moins de 25 ans.
Pour simuler une situation dans laquelle seul le dernier recensement possède la question sur les décès des 12 derniers mois, on calcule ici une estimation du nombre de décès dans chaque groupe d'âge entre l'instant du recensement de 2001 et l'enquête de communauté de 2007 en utilisant seulement les décès déclarés à l'enquête de communauté de 2007 selon la formule suivante :

$$\frac{{}_5D_x(t_2)}{{}_5N_x(t_2)}(t_2 - t_1)({}_5N_x(t_1) \times {}_5N_x(t_2))^{1/2}$$

En appliquant la méthode à ces nouveaux effectifs de décès, on trouve des estimations des quotients de mortalité à 15-59 ans, $_{45}q_{15}$, de 586 pour 1 000 chez les

TABEAU 24.7 Rapport des estimations des décès déclarés par les ménages aux décès attendus, déclarés à l'état civil après correction du sous-enregistrement, Afrique du sud

| Âge | Sexe masculin | | | Sexe féminin | | |
|-------|---------------|----------|---------|--------------|----------|---------|
| | Déclarés | Attendus | Rapport | Déclarés | Attendus | Rapport |
| 0-4 | | | | | | |
| 5-9 | 22 683 | 17 193 | 132 % | 22 995 | 14 670 | 157 % |
| 10-14 | 16 462 | 12 378 | 133 % | 15 173 | 10 417 | 146 % |
| 15-19 | 38 013 | 28 134 | 135 % | 35 666 | 27 050 | 132 % |
| 20-24 | 74 934 | 60 701 | 123 % | 95 993 | 85 167 | 113 % |
| 25-29 | 124 403 | 113 541 | 110 % | 152 718 | 155 452 | 98 % |
| 30-34 | 150 792 | 160 796 | 94 % | 166 488 | 171 801 | 97 % |
| 35-39 | 159 016 | 161 141 | 99 % | 137 837 | 142 328 | 97 % |
| 40-44 | 140 172 | 150 136 | 93 % | 111 910 | 116 506 | 96 % |
| 45-49 | 120 016 | 133 651 | 90 % | 85 284 | 94 022 | 91 % |
| 50-54 | 118 989 | 122 768 | 97 % | 76 941 | 82 330 | 93 % |
| 55-59 | 97 977 | 106 972 | 92 % | 57 353 | 72 605 | 79 % |
| 60-64 | 88 088 | 99 325 | 89 % | 69 220 | 79 395 | 87 % |
| 65-69 | 80 451 | 91 497 | 88 % | 67 007 | 86 665 | 77 % |
| 70-74 | 72 827 | 80 666 | 90 % | 69 536 | 94 017 | 74 % |
| 75-79 | 59 632 | 70 543 | 85 % | 61 942 | 88 894 | 70 % |
| 80-84 | 45 365 | 53 194 | 85 % | 58 410 | 77 590 | 75 % |
| 85+ | 51 779 | 51 021 | 101 % | 83 753 | 108 712 | 77 % |

hommes et de 578 pour 1 000 chez les femmes. Mais, à la différence des estimations précédentes, ce sont des estimations de mortalité qui se rapportent essentiellement à l'année qui précède l'enquête de 2007. On pouvait donc s'attendre à une mortalité plus forte que la moyenne de la période 2001-2007 car la mortalité était en augmentation au cours de la période du fait de l'épidémie de VIH/sida. Il est bien évident que lorsqu'on fait une estimation à partir des décès d'une seule année (qui plus est provenant d'une enquête par sondage, sur un échantillon relativement faible) on peut s'attendre à des résultats moins fiables, ce qui semble être le cas ici, surtout pour les femmes. D'autres estimations faites pour l'année 2006 (Bradshaw, Dorrington and Laubscher 2012) donnent des valeurs proches de 550 pour 1000 chez les hommes et de 450 pour 1 000 chez les femmes.

DESCRIPTION DÉTAILLÉE DE LA MÉTHODE

Exposé mathématique

La méthode généralisée de la balance de l'accroissement démographique suit la même logique que la méthode de Brass de la balance de l'accroissement démographique (Brass 1975). Cette dernière a pour origine les travaux de Carrier (1958), qui le premier a proposé une méthode d'estimation de la mortalité à partir de la distribution par âge de décès. La méthode est fondée sur une relation simple, l'équation de l'équilibre des populations, que l'on suppose par convention être fermée aux migrations. Dans une telle population fermée, le nombre de personnes vivantes au temps t_2 est égal au nombre de personnes vivantes au temps t_1 plus les naissances qui se produisent entre les instants t_1 et t_2 , moins les décès qui se produisent entre les instants t_1 et t_2 , qui peut s'écrire :

${}_{\infty}N_0(t_2) = {}_{\infty}N_0(t_1) + B - D$ où B sont les naissances et D sont les décès qui se produisent entre les instants t_1 et t_2 . On peut généraliser cette équation à la population d'âge x et plus. Dans ce cas, les naissances correspondent aux personnes qui atteignent leur x -ème anniversaire au cours de la période entre les instants t_1 et t_2 , noté N_x , et les décès sont les décès d'âge x et plus qui se produisent entre les instants t_1 et t_2 , noté μD_x . L'équation s'écrit alors :

$${}_{\infty}N_x(t_2) = {}_{\infty}N_x(t_1) + N_x - {}_{\infty}D_x \quad (1)$$

L'équation (1) peut se réécrire comme suit :

$${}_{\infty}N_x(t_2) - {}_{\infty}N_x(t_1) = N_x - {}_{\infty}D_x.$$

On peut diviser les deux termes par les personnes-années vécues au cours de la période entre les instants t_1 et t_2 , soit :

$$\int_{t_1}^{t_2} {}_{\infty}N_x(t) dt$$

et l'équation de la balance de l'accroissement démographique s'écrit :

$$r(x+) = b(x+) - d(x+) \quad (2)$$

où

$$r(x+) = \frac{{}_{\infty}N_x(t_2) - {}_{\infty}N_x(t_1)}{\int_{t_1}^{t_2} {}_{\infty}N_x(t) dt}$$

$$b(x+) = \frac{N_x}{\int_{t_1}^{t_2} {}_{\infty}N_x(t) dt}$$

et

$$d(x+) = \frac{{}_{\infty}D_x}{\int_{t_1}^{t_2} {}_{\infty}N_x(t) dt}.$$

On appelle souvent les taux $b(x+)$ et $d(x+)$ respectivement les « taux partiels » de natalité et de mortalité, ou parfois les « taux segmentiels » (*segmental birth and death rates*).

Ces relations ne sont vérifiées que dans le cas où l'enregistrement des décès est complet pour tous les âges entre les instants t_1 et t_2 , que les âges sont exacts, et qu'il en va de même pour le décompte de la population aux instants t_1 et t_2 .

Supposons maintenant qu'au lieu d'un enregistrement complet, seule une certaine proportion de décès sont déclara-

és, la même à tous les âges, et qu'il en va de même pour les recensements de population, c'est-à-dire qu'une certaine proportion est recensée, qui peut être différente d'un recensement à l'autre, et différente de celle des décès. On a alors, au lieu des vraies valeurs : ${}_{\infty}N_x(t_1)$, ${}_{\infty}N_x(t_2)$ et ${}_{\infty}D_x$ des valeurs déclarées, notées : ${}_{\infty}N_x^r(t_1)$, ${}_{\infty}N_x^r(t_2)$ et ${}_{\infty}D_x^r$ telles que : ${}_{\infty}N_x^r(t_1) = k_1 \cdot {}_{\infty}N_x(t_1)$, ${}_{\infty}N_x^r(t_2) = k_2 \cdot {}_{\infty}N_x(t_2)$ et ${}_{\infty}D_x^r = c \cdot {}_{\infty}D_x$.

Alors, si on utilise les approximations suivantes :

$$r(x+) \approx \frac{\ln\left(\frac{{}_{\infty}N_x(t_2)}{{}_{\infty}N_x(t_1)}\right)}{t_2 - t_1}$$

$$N_x \approx (t_2 - t_1) \frac{\sqrt{{}_nN_{x-n}(t_1) \times {}_nN_x(t_2)}}{n}$$

et

$$\int_{t_1}^{t_2} {}_{\infty}N_x(t) dt \approx (t_2 - t_1) \left({}_{\infty}N_x(t_1) \times {}_{\infty}N_x(t_2) \right)^{1/2}$$

alors on a :

$$r(x+) = r^r(x+) + \frac{\ln\left(\frac{k_1}{k_2}\right)}{t_2 - t_1}$$

où

$$r^r(x+) = \frac{\ln\left(\frac{{}_{\infty}N_x^r(t_2)}{{}_{\infty}N_x^r(t_1)}\right)}{t_2 - t_1}$$

$$b(x+) = b^r(x+) = \frac{(t_2 - t_1) \left({}_nN_{x-5}^r(t_1) \times {}_nN_x^r(t_2) \right)^{1/2} / n}{(t_2 - t_1) \left({}_{\infty}N_x^r(t_1) \times {}_{\infty}N_x^r(t_2) \right)^{1/2}}$$

et

$$d(x+) = \frac{(k_1 k_2)^{1/2}}{c} d^r(x+)$$

et où

$$d^r(x+) = \frac{{}_{\infty}D_x^r}{(t_2 - t_1) \left({}_{\infty}N_x^r(t_1) \times {}_{\infty}N_x^r(t_2) \right)^{1/2}}$$

L'équation 2 devient alors :

$$r^r(x+) + \frac{\ln\left(\frac{k_1}{k_2}\right)}{t_2 - t_1} = b^r(x+) - d^r(x+) \frac{(k_1 k_2)^{1/2}}{c},$$

C'est-à-dire : $b^r(x+) - r^r(x+) = a + bd^r(x+)$ avec :

$$a = \frac{\ln(k_1 / k_2)}{t_2 - t_1}$$

et

$$b = \frac{(k_1 k_2)^{1/2}}{c}.$$

On peut donc résoudre cette équation pour k_1 , k_2 et c , en faisant l'hypothèse que la complétude du meilleur des recensements est de 100 %, c'est-à-dire en prenant la valeur du maximum de k_1 et $k_2 = 1$.

Ajustement de la droite de régression

Deux aspects doivent être considérés pour ajuster au mieux la droite qui décrit la relation entre le taux partiel de natalité et le taux partiel de mortalité : le choix de la méthode de régression, et le choix des points qui sont retenus pour calculer la pente et l'ordonnée à l'origine.

L'ajustement d'une droite en utilisant une régression non-pondérée et la méthode des moindres carrés n'est pas recommandé ici, car il donne trop de poids aux points aberrants qui ne sont pas fiables, surtout ceux qui représentent les âges élevés. On recommande plutôt une méthode plus robuste, telle que la ligne des moyennes (*mean line*), c'est-à-dire la droite qui joint deux points fixes, qui eux-mêmes sont les moyennes en abscisse et en ordonnée de la première moitié et de la seconde moitié des points correspondant à l'intervalle d'âge retenu. On peut aussi utiliser la ligne des moyennes tronquées (*trimmed mean line*) : c'est une ligne analogue, qui joint deux points fixes qui eux-mêmes sont les moyennes pondérées des points, les pondérations étant choisies pour donner moins de poids aux points qui sont moins fiables, en général les points situés aux extrémités. Ces méthodes sont expliquées en détail dans le *Manuel X* (UN Population Division 1984 : 147–148). Une autre alternative est décrite dans le *Manuel des Nations Unies sur la Mortalité Adulte* (UN Population Division 2002 : 105–110). Cette méthode est là encore analogue à la méthode de la ligne des moyennes, mais dans ce cas on divise les points en trois groupes de même taille,¹ et on choisit la droite qui joint les

médianes des points en abscisse et en ordonnée correspondant au premier tiers et au dernier tiers des points classés par âge.

Mari Bhat signale que chaque méthode a ses points faibles (Bhat 2002). Puisque les taux partiels de natalité et de mortalité sont traités tous deux comme des variables dépendantes, il propose d'utiliser la régression orthogonale comme la meilleure méthode pour tenir compte d'éventuelles erreurs sur les âges. Cette stratégie permet de tenir compte à la fois des distances horizontales et verticales à la ligne de régression, puisqu'elle minimise la somme des carrés des résidus orthogonaux (ORSS), définis comme suit :

$$\sum_i \frac{x_i^2 y_i^2}{x_i^2 + y_i^2}.$$

L'utilisation de cette méthode conduit à calculer la complétude de la déclaration des décès, c , comme le rapport des écarts-types des taux partiels de mortalité et de natalité. La valeur de l'ordonnée à l'origine est la différence entre la moyenne des taux partiels de natalité et la moyenne des taux partiels mortalité divisés par c . C'est cette approche qui est utilisée dans l'application de la méthode de Brass de la balance généralisée de l'accroissement démographique présentée dans le classeur Excel associé à ce manuel.

Limites de la méthode

Cette méthode est moins vulnérable aux erreurs sur l'âge que la Méthode de l'extinction des cohortes synthétiques. Cependant, la tendance fréquente à exagérer l'âge au décès tel qu'il est déclaré (par rapport à l'âge déclaré au recensement) se manifesterait dans le graphique : les points correspondants à l'intervalle d'âge où l'âge est exagéré auraient tendance à se placer en-dessous de la droite de régression. On peut alors traiter ce problème en diminuant l'âge de début du dernier intervalle, en le choisissant en-dessous de l'âge qui correspond aux points qui font problème.

Les migrations ne sont pas prises en compte par ce modèle, mais elles peuvent affecter la population des jeunes adultes (surtout entre 20 et 35 ans). Elles ont en général beaucoup moins d'effet sur les décès, car ceux-ci sont concentrés aux âges élevés. Ignorer les immigrations va tendre à diminuer la pente de la droite, et donc à surestimer le sous-enregistrement des décès, et par conséquent à sous-estimer les taux de mortalité corrigés. Ignorer les émigrations aura l'effet contraire.

1 Lorsqu'il n'est pas possible de diviser le nombre total de points, n , en trois groupes de même taille, on peut alors prendre comme groupes ceux formés par les k premiers et k derniers points, k étant calculé comme : $\text{ent}(n/3) + 1$.

Les estimations du nombre de migrations nettes par âge au cours de la période intercensitaire font souvent défaut. Dans une telle situation, on peut procéder en suivant les recommandations suivantes. Si les migrations sont importantes mais restent inconnues, et que les points situés au-delà de l'âge de 30 ans sont bien alignés, on peut estimer la complétude en ajustant une droite restreinte aux points correspondants aux âges de 35 ans et plus. Si les migrations sont peu importantes, certains démographes préconisent d'ajuster une droite en partant de l'âge de 5 ans, pour réduire cette distorsion, en faisant l'hypothèse que toute différence entre le sous-enregistrement des décès à ces âges jeunes et celui aux âges élevés n'entraînera pas de distorsion sensible, puisque la mortalité reste très faible entre 5 et 14 ans. Dans le cas où les migrations sont peu importantes, d'autres auteurs (Hill, You and Choi 2009) proposent une autre stratégie : on pourrait obtenir une meilleure estimation en prenant la moyenne entre l'estimation de la complétude obtenue avec cette méthode et celle obtenue en appliquant la Méthode de l'extinction des cohortes synthétiques aux mêmes données. Il reste possible que ces adaptations produisent de meilleurs estimations que celles obtenues en ignorant les migrations, mais on ne dispose que de peu de recherches sur ce point, ainsi que sur la précision des estimations de la complétude obtenues en procédant à de telles adaptations.

Les variations de la complétude de la déclaration des décès en fonction de l'âge peuvent introduire une courbure dans l'alignement des points du graphique. En conséquence, une des forces de cette méthode est que si les points qui représentent les groupes d'âge successifs sont approximativement situés sur une ligne droite, on peut alors raisonnablement supposer que la complétude est approximativement constante selon l'âge. Mais si certains points sont sur la ligne, et que d'autres en sont éloignés, une méthode pour décider quels points rejeter consiste à calculer le taux partiel de croissance pour chacun des intervalles ouverts successifs, et de garder seulement les points pour lesquels les valeurs de r_{a+} sont assez cohérentes.

Une des principales limites de cette méthode est peut-être que la valeur du diagnostic fait à partir de la représentation graphique des taux partiels de natalité et de mortalité reste faible, hormis les exceptions mentionnées ci-dessus.

Extensions de la méthode

Si les âges sont précis, et si l'hypothèse de la couverture constante par âge des recensements est vérifiée, alors la

méthode peut être adaptée aux cas où la complétude de la déclaration des décès n'est constante que sur un intervalle d'âges limité, $(x \text{ to } x + n)$.² Dans ce cas on restreint le groupe d'âge choisi pour l'équation de la balance de la croissance démographique. L'équation 2 devient alors :

$${}_n r_x = b(x+) - {}_n d_x - b((x + n) +),$$

où

$${}_n r_x = \frac{{}_n N_x(t_2) - {}_n N_x(t_1)}{\int_{t_1}^{t_2} {}_n N_x(t) dt}$$

$$b(x+) = \frac{N_x}{\int_{t_1}^{t_2} N_x(t) dt}$$

et

$${}_n d_x = \frac{{}_n D_x}{\int_{t_1}^{t_2} {}_n N_x(t) dt}.$$

La partie gauche de l'équation de régression qui correspond aux observations sélectionnées devient alors : $b^r(x+) - b^r(x + n+) - {}_n r_x^r$. Mais on a très peu d'expérience sur la manière dont cette approche alternative fonctionne en pratique, probablement parce que les données des pays en développement sont rarement assez précises.

AUTRES LECTURES ET RÉFÉRENCES

L'analyse de la sensibilité de cette méthode aux erreurs de données les plus fréquentes et aux violations des hypothèses reste assez limitée. Toutefois, on renvoie le lecteur à Hill, You et Choi (2009) pour ce qui concerne les hypothèses de la méthode de la répartition des décès dans les populations sans sida, et à Dorrington et Timæus (2008) pour le cas des populations souffrant d'une forte épidémie de VIH/sida. Murray, Rajaratnam, Marcus *et al.* (2010) ont utilisé des simulations stochastiques et conclu que ces méthodes ne sont pas vraiment fiables. Mais leur travail n'a eu jusqu'ici qu'un impact très limité sur l'utilisation de ces méthodes,

2 Un tel cas où la complétude de l'état civil peut baisser au-delà de l'âge à la retraite peut se produire lorsque les retraités migrant des zones urbaines vers les zones rurales, ou lorsque les décès déclarés par les ménages se produisent dans des ménages qui sont dissous à la suite du dernier décès d'adulte.

peut-être parce que ces auteurs ne fournissent que peu de détails sur la description de leurs simulations, ou peut-être que leur étude est basée sur des hypothèses irréalistes concernant une fréquence élevée des migrations.

- Bhat M. 2002. "General Growth Balance method: A reformulation for populations open to migration", *Population Studies* **56**(1):23-34. doi: <http://dx.doi.org/10.1080/00324720213798>
- Blacker J. 1988. *An Evaluation of the Pakistan Demographic Survey*. Karachi: Pakistan Federal Bureau of Statistics.
- Bradshaw D, RE Dorrington and R Laubscher. 2012. *Rapid Mortality Surveillance Report 2012*. Cape Town: South African Medical Research Council. <http://www.mrc.ac.za/bod/Rapid-Mortality2011.pdf>
- Brass W. 1975. *Methods for Estimating Fertility and Mortality from Limited and Defective Data*. Chapel Hill NC: Carolina Population Centre.
- Carrier NH. 1958. "A note on the estimation of mortality and other population characteristics, given death by age", *Population Studies* **12**(2):149-163. doi: <http://dx.doi.org/10.2307/2172187>
- Division de la Population des Nations Unies. 1984. *Manuel X. Techniques indirectes d'estimation démographique*. New York : Nations Unies, Département des affaires économiques et sociales internationales, ST/ESA/SER.A/81. https://unstats.un.org/unsd/demographic/standmeth/handbooks/Manuel_X-fr.pdf
- Dorrington RE, TA Moultrie and IM Timæus. 2004. *Estimation of mortality using the South African 2001 census data*. Monograph 11. Centre for Actuarial Research, University of Cape Town. <http://www.commerce.uct.ac.za/care/Monographs/Monographs/Mono11.pdf>
- Dorrington RE and IM Timæus. 2008. "Death Distribution Methods for Estimating Adult Mortality: Sensitivity Analysis with Simulated Data Errors, Revisited," Paper presented at Population Association of America 2008 Annual Meeting. New Orleans, Louisiana, 17-19 April.
- Hill K. 1987. "Estimating census and death registration completeness", *Asian and Pacific Census Forum* **1**(3):8-13, 23-24. <http://hdl.handle.net/10125/3602>.
- Hill K, D You and Y Choi. 2009. "Death distribution methods for estimating adult mortality: Sensitivity analysis with simulated data error", *Demographic Research* **21**(Article 9):235-254. doi: <http://dx.doi.org/10.4054/DemRes.2009.21.9>
- Murray CJL, JK Rajaratnam, J Marcus, T Laakso and AD Lopez. 2010. "What can we conclude from death registration? Improved methods for evaluating completeness", *PLoS Med* **7**(4):e1000262. doi: <http://dx.doi.org/10.1371/journal.pmed.1000262>
- Timæus IM. 2004. "Impact of HIV on mortality in Southern Africa: Evidence from demographic surveillance," Paper presented at Seminar of the IUSSP Committee "Emerging Health Threats" HIV, Resurgent Infections and Population Change in Africa. Ougadougou, 12-14 February.
- UN Population Division. 2002. *Methods for Estimating Adult Mortality*. New York: United Nations, Department of Economic and Social Affairs, ESA/P/WP.175. http://www.un.org/esa/population/techcoop/DemEst/methods_adultmort/methods_adultmort.html
- UN Population Division. 2011. *World Population Prospects: The 2010 Revision, Volume I: Comprehensive Tables*. New York: United Nations, Department of Economic and Social Affairs, ST/ESA/SER.A/313. http://esa.un.org/unpd/wpp/Documentation/pdf/WPP2010_Volume-I_Comprehensive-Tables.pdf

Chapitre 25 MÉTHODE DE L'EXTINCTION DES COHORTES SYNTHÉTIQUES

Rob Dorrington

DESCRIPTION DE LA MÉTHODE

Les démographes Neil Bennett et Shiro Horiuchi (Bennett et Horiuchi 1981, 1984) ont proposé une généralisation de la méthode de Preston et Coale pour estimer la complétude de la déclaration des décès par rapport à une estimation de la population, et qui est maintenant connue sous le nom de la méthode de l'extinction des cohortes synthétiques (*Synthetic Extinct Generations (SEG) method*). Ces deux méthodes partent du même constat que, dans une population fermée, le nombre de personnes d'âge a survivantes à un instant t doit être égal au nombre de décès cumulés au-delà du temps t dans la même cohorte. Le principe de base sur lequel se fondent les méthodes de l'extinction des cohortes synthétiques est que le flux de décès à venir dans une cohorte peut être estimé par les décès d'âge a se produisant au temps t corrigés en tenant compte de la croissance démographique. Le cas le plus simple, mais très particulier, est celui d'une population stationnaire, associée à une table de mortalité, dans laquelle le nombre de décès d'âge a au temps t sera égal aux décès d'âge a dans la même cohorte. Un cas plus général, mais un peu plus complexe, est celui d'une population stable, c'est-à-dire une population fermée aux migrations, croissant à un taux constant r , et ayant une mortalité par âge constante et une structure par âge constante. Dans ce cas, si les données déclarées sont complètes et précises, le nombre de décès à l'âge x dans t années sera égal au nombre de décès mesuré à un instant précis multiplié par le coefficient de la croissance démographique, soit e^{rt} . C'est cette relation qui fonde la méthode de Preston et Coale (chapitre 21).

Dans le cas encore plus général dans lequel la population n'est pas stable, on peut trouver une relation équivalente en remplaçant : e^{rt} par $e^{\int_t^T r(y,t)dy}$, où $r(y,t)$ représente le taux de croissance instantané de la population d'âge y au temps t .

Si les décès déclarés au temps t sont sous-déclarés dans la même proportion, c , à tous les âges, au moins pour les âges adultes, alors le nombre de décès attendus dans la cohorte sera sous-estimé dans la même proportion. On pourra donc estimer la complétude de la déclaration des décès en divisant la somme des estimations du nombre de décès attendus dans le futur dans la cohorte calculées d'après les décès mesurés à l'instant t par la population mesurée à la même date. On pourra ainsi recalculer les taux de mortalité du moment en divisant le nombre de décès déclarés dans chaque groupe d'âge par le coefficient c , puis en divisant ces nombres corrigés par une estimation de la population au risque, ce qui permettra de recalculer les taux partiels de natalité et de mortalité.

DONNÉES REQUISES ET HYPOTHÈSES

Tabulation des données requises

- Les effectifs de femmes (ou d'hommes), répartis par groupe d'âge de cinq ans, et pour l'intervalle d'âge ouvert $A+$ (avec A aussi élevé que possible), à deux instants donnés, le plus souvent provenant de deux recensements (voir ci-dessous les inconvénients qui résultent de l'utilisation d'enquêtes par sondage au lieu de recensements).
- Les nombres de décès de femmes (ou d'hommes), répartis par groupe d'âge de cinq ans, et pour l'intervalle d'âge ouvert $A+$, pour toute la période située entre les deux recensements ou les deux enquêtes.

Hypothèses importantes

- La couverture de chaque recensement est la même à tous les âges
- La complétude de la déclaration des décès est la même à tous les âges au-delà d'un âge minimal (en général 15 ans)

- La population est fermée aux migrations. La méthode pourra être adaptée aux cas des migrations, mais on ne dispose que très rarement des effectifs de migrations nettes par âge. Pour les populations nationales, les migrations nettes sont souvent suffisamment faibles pour être négligées, mais lorsque ce n'est pas le cas, il conviendra de prendre en compte les migrations lorsqu'on interprétera les résultats et que l'on devra retenir une estimation finale de la complétude.

TRAVAUX PRÉPARATOIRES ET RECHERCHES PRÉLIMINAIRES

Avant de mettre en œuvre cette méthode, on doit au préalable analyser la qualité des données, et au moins les points suivants :

- structure par âge de la population ;
- répartition par sexe de la population ;
- structure par âge des décès ; et
- répartition par sexe des décès.

Si la période au cours de laquelle les décès sont déclarés est différente de celle située entre les deux recensements, il faudra corriger ces nombres pour estimer ceux qui se sont produits dans la période intercensitaire. Lorsqu'on dispose de données de l'état civil, cet ajustement consistera à répartir proportionnellement les décès des années correspondant aux années des deux recensements, c'est-à-dire la première et la dernière année de la période. Lorsqu'on dispose des décès déclarés par les ménages au cours des 12 mois précédents chacun des recensements, on doit estimer le nombre de décès de la période intercensitaire par interpolation (voir la feuille de calcul suivante : *Estimating deaths.xlsx*).

La méthode de l'extinction des cohortes synthétiques utilise des taux de croissance par âge pour effectuer ces calculs. Si la couverture du décompte de population varie d'un recensement à l'autre, dans un rapport constant par âge, alors les taux de croissance par âge vont être systématiquement biaisés, d'une quantité fixe, notée *delta*. La méthode généralisée de la balance de l'accroissement démographique permet d'ailleurs de calculer directement ce biais. Par contre la méthode de l'extinction des cohortes synthétiques ne permet pas d'estimer directement *delta*, mais des valeurs non nulles de *delta* se traduisent par une tendance linéaire dans les estimations de la complétude par âge. On pourra alors estimer *delta* par itération, en recherchant la valeur qui produit des estimations de la complétude constantes par âge sur un intervalle situé entre deux âges, minimal et maximal.

PRÉCAUTIONS ET MISES EN GARDE

Lorsqu'ils appliquent cette méthode, les analystes doivent prendre en considération les points suivants :

- Le processus d'interprétation et d'estimation doit prendre en compte la source des données de mortalité (décès déclarés à l'état civil, décès déclarés par les ménages dans les recensements, ou décès déclarés dans les structures sanitaires), comme cela est expliqué plus loin. Les biais associés à la source des données tendent à avoir plus d'effet sur l'estimation de la complétude avec la méthode de l'extinction des cohortes synthétiques qu'avec la méthode généralisée de la balance de l'accroissement démographique.
- Le problème des migrations devient plus sérieux quand on applique la méthode à des zones géographiques régionales, et non plus nationales.
- Il est nécessaire de choisir correctement l'intervalle d'âges pour calculer *delta* (l'estimation relative de la couverture d'un recensement par rapport à l'autre). Les problèmes spécifiques sont les suivants : est-ce que la meilleure estimation de *delta* est fournie par l'ordonnée à l'origine obtenue par application de la méthode généralisée de la balance de l'accroissement démographique aux mêmes données (ce qui peut être le cas, par exemple, si on pense que le sous-enregistrement des décès augmente après l'âge à la retraite, pour des raisons expliquées ci-dessous) ? Est-ce qu'il convient d'exclure les âges en dessous de 30 ou 35 ans, car l'impact des migrations n'y est pas pris en compte spécifiquement ?
- Il est nécessaire de choisir correctement l'intervalle d'âges pour calculer la complétude. En général, cet intervalle d'âge pourrait exclure les jeunes adultes lorsqu'on suspecte de fortes migrations qui ne sont pas prises en compte, ou exclure les personnes âgées si les résultats indiquent que leurs décès sont moins bien déclarés que ceux des jeunes adultes, ou si les erreurs sur l'âge des personnes âgées semble affecter différemment les âges des survivants et ceux des décédés.
- Il convient de s'assurer que le module *Solveur* de *Excel* a fonctionné correctement, c'est à dire qu'il a fourni des résultats réalistes. Il peut arriver que le module *Solveur* fournisse une solution qui soit manifestement trop basse. Dans de telles situations, il est préférable d'ajuster *delta* à la main dans la bonne direction, et de faire tourner le module *Solveur* en démarrant par cette nouvelle valeur.

- Il convient de s'assurer que l'estimation de l'espérance de vie dans le groupe d'âge ouvert est réaliste. Il arrive souvent que les données sur les décès des personnes âgées sont rares et sujettes à des erreurs. Les estimations de l'espérance de vie qui en résultent peuvent être irréalistes, surestimant le plus souvent l'espérance de vie. Plus l'âge au début de l'intervalle ouvert est élevé, plus l'impact de ce type d'erreur sera faible.
- Lorsque la complétude des décès déclarés semble être inférieure à 60 %, on conseille de prendre garde aux résultats, car l'incertitude sur les estimations devient importante.
- Dans le cas où les données sur la population et les décès des ménages sont collectées lors d'un recensement unique, il est tentant de chercher à utiliser des données d'une enquête par sondage pour avoir une répartition par âge de la population à une autre date, antérieure ou postérieure, de manière à pouvoir utiliser la méthode. Mais, pour des raisons qui ne sont pas claires et qui n'ont pas fait l'objet de recherches approfondies, une telle combinaison de sources de données très différentes donne rarement des résultats satisfaisants.

APPLICATION DE LA MÉTHODE

D'un point de vue purement technique, on pourrait appliquer cette méthode à des données présentées par année d'âge, mais ce type de données est sujet à de fortes erreurs sur l'âge, si bien que, en pratique, on travaille le plus souvent avec des données présentées par groupe d'âge de 5 ans. Comme la plupart des données disponibles sont publiées selon ce format, la feuille de calcul correspondante est organisée pour travailler avec des groupes d'âge quinquennaux classiques. Il faut remarquer que John Blacker (Blacker 1988) a montré que ces groupes d'âge classiques sont sensibles aux préférences pour les chiffres ronds ; dans ce cas on pourra adapter la méthode pour travailler avec des groupes d'âge quinquennaux différents, centrés sur les chiffres qui font l'objet de l'attraction (12-17, 18-21 etc.), plutôt qu'avec des groupes d'âge commençant par ces chiffres (15-19, 20-24, etc.).

Étape 1 : Lorsque ce nombre n'est pas immédiatement disponible, estimation du nombre de décès déclarés au cours de la période entre les deux recensements de population

Dans le cas où l'on dispose de données de l'état civil par année de calendrier, l'ajustement consiste à répartir pro-

portionnellement les décès de la première et de la dernière année. On les répartit proportionnellement à la fraction de l'année couverte après le premier recensement et avant le second recensement. Sauf si la distribution des décès par âge change très rapidement au cours de l'année, cette approximation n'aura pas d'effet notable sur les résultats.

Dans le cas où l'on ne dispose pas des décès entre les deux enquêtes, mais si l'intervalle tombe entre deux périodes pour lesquelles on a de telles données (par exemple dans le cas où chaque recensement inclut la question sur les décès des douze derniers mois dans le ménage), on peut alors utiliser la feuille de calcul suivante : *Estimating deaths.xlsx*. Cette feuille de calcul permet d'estimer le nombre de décès entre deux instants donnés d'après les décès qui se sont produits au cours de deux périodes les encadrant. Pour utiliser cette feuille de calcul, on a besoin du nombre de décès répartis par groupes d'âge quinquennaux pour les deux périodes, les dates de début et de fin de chaque période, ainsi que les dates de début et de fin de la période au cours de laquelle on veut estimer le nombre de décès.

Étape 2 : Calcul des taux de croissance par âge, après correction pour tenir compte des migrations et de la différence de couverture entre les deux recensements

Les taux de croissance par âge, après correction pour tenir compte des migrations et de la différence de couverture des deux recensements, se calculent à partir des deux recensements, et du nombre de migrants au cours de la période intercensitaire, répartis par groupe d'âge, comme suit :

$${}_5r_x = \frac{\ln({}_5N_x(t_2) / {}_5N_x(t_1))}{t_2 - t_1} - \frac{{}_5NM_x}{(t_2 - t_1)({}_5N_x(t_1) \times {}_5N_x(t_2))^{\frac{1}{2}}} + \delta$$

où ${}_5N_x(t)$ est la population du groupe d'âge x à $x + 5$ à l'instant t , ${}_5NM_x$ est le nombre de migrations nettes (immigrants moins émigrants) dans le groupe d'âge x à $x + 5$, et t_1 et t_2 sont les dates moyennes des deux recensements. *Delta* (δ) représente la correction à effectuer pour prendre en compte la différence de couverture des deux recensements. On le calcule soit en utilisant la méthode de la balance de l'accroissement démographique, soit par itération comme cela est expliqué plus loin.

Étape 3 : Calcul de l'espérance de vie à l'âge A et au début de chaque groupe d'âge quinquennal, jusqu'à 65 ans

On peut faire ce calcul de plusieurs manières :

- 1) Utiliser une estimation provenant d'autres sources fiables, s'il en existe. Ce peut être une estimation provenant de recherches antérieures, ou une estimation tirée des projections de population comme celles fournies par les *Perspectives de la population mondiale* (UN Population Division 2011).
- 2) Utiliser une estimation calculée à partir des mêmes données auxquelles on a appliqué la méthode de la balance de l'accroissement démographique. Le classeur *Excel* qui correspond à cette méthode fournit, parmi ses résultats, une telle estimation.
- 3) Utiliser une table-type de mortalité. Pour cela, on pourra calculer le rapport des décès dans les groupes d'âge 10-39 ans et 40-59 ans (âge au dernier anniversaire), soit $(_{30}D_{10}/_{20}D_{40})$, comme entrée dans le système des tables-type. Ce rapport pourra être comparé à celui du modèle Ouest des tables-type de Princeton, qui servira à calculer l'espérance de vie. On trouvera dans le classeur *Excel* correspondant une feuille de calcul qui effectue ces calculs. Mais, comme le modèle Ouest des tables-type de Princeton ne permet pas d'ajuster un schéma par âge résultant des décès causés par le VIH/sida, cette approche ne convient pas aux pays qui souffrent d'une forte mortalité par sida.
- 4) Utiliser une méthode itérative : on part d'une valeur raisonnable, que l'on peut prendre dans une table-type du modèle Ouest de Princeton (sauf si le pays souffre d'une forte mortalité par sida), ou provenant d'une source indépendante. Ensuite, on estime la complétude, comme cela est expliqué plus loin, on recopie les nouvelles valeurs de l'espérance de vie après correction pour tenir compte du sous-enregistrement dans la feuille de calcul *Espérances de vie* du classeur *Excel*, on les colle dans la feuille de calcul *Méthode* du classeur *Excel* et on recommence l'estimation de la complétude. On peut répéter l'opération plusieurs fois si nécessaire, jusqu'à ce que l'estimation devienne stable, c'est-à-dire que les variations de l'espérance de vie deviennent négligeables. Malheureusement, s'il y a de bonnes raisons de penser que la mortalité reste sous-estimée aux âges élevés, même après correction du sous-enregistrement (par exemple si les âges sont exagérés, ou si le niveau de complétude baisse avec l'âge), alors cette approche va tendre à surestimer les espérances de vie, et

par conséquent à surestimer le niveau de complétude de la déclaration des décès.

Étape 4 : Calcul du nombre de personnes qui atteignent l'âge x, et ceux qui atteignent le groupe d'âge x à x + 4, d'après le nombre de décès déclarés

Le nombre de personnes qui atteignent l'âge x au cours de la période de référence, c'est-à-dire celle au cours de laquelle les décès sont déclarés, se calcule à partir des décès déclarés comme suit :

$$\widehat{N}_x = \widehat{N}_{x+5} \exp(5 \, {}_5r_x) + {}_5D_x \exp(2,5 \, {}_5r_x)$$

et

$$\widehat{N}_A = {}_{\infty}D_A \left(\exp({}_{\infty}r_A \times e_A) - ({}_{\infty}r_A \times e_A)^2 / 6 \right)$$

où A désigne le début de l'intervalle ouvert, ${}_nr_x$ est le taux de croissance annuel dans le groupe d'âge de x à x + n (âge au dernier anniversaire), et e_A est l'espérance de vie à l'âge A.

Le nombre de personnes qui atteignent le groupe d'âge de x à x + 4 (âge au dernier anniversaire), pendant la période de référence, se calcule d'après les nombres de personnes qui atteignent l'âge x, de 5 ans en 5 ans, comme suit :

$${}_5\widehat{N}_x = 2,5(\widehat{N}_x + \widehat{N}_{x+5}).$$

Étape 5 : Calcul du nombre de personnes qui atteignent le groupe d'âge x à x + 4 au cours de la période située entre les deux recensements d'après les effectifs dénombrés

Le nombre de personnes qui atteignent le groupe d'âge de x à x + 4 (âge au dernier anniversaire), pendant la période de référence, se calcule en multipliant la moyenne géométrique des effectifs dénombrés dans le même groupe d'âge aux deux recensements par la durée de la période intercensitaire (mesurée en années), comme suit :

$${}_5N_x = (t_2 - t_1)({}_5N_x(t_1) \times {}_5N_x(t_2))^{1/2}$$

Étape 6 : Calcul des rapports des estimations faites à partir des décès et des estimations faites à partir des recensements de population

On peut calculer deux séries de rapports entre les estimations qui proviennent des décès et celles qui proviennent des populations recensées. Les premiers rapports sont ceux des groupes d'âge quinquennaux, que l'on peut calculer directement. Les seconds rapports sont ceux des effectifs cumulés depuis l'âge x jusqu'à l'âge de début de l'intervalle ouvert, A,

aux effectifs des personnes qui atteignent le même groupe d'âge, c'est-à-dire de l'âge x à l'âge $A - 1$ au cours de la période de référence, qui se calculent comme les sommes des effectifs dans les groupes d'âge quinquennaux, soit en formule :

$${}_{A-x}\widehat{N}_x = \sum_{a=x,5}^{A-5} {}_5\widehat{N}_a$$

Étape 7 : Estimation de la complétude de la déclaration des décès

Afin de déterminer le niveau de sous-enregistrement des décès, il faut d'abord décider si les taux de croissance choisis doivent être corrigés pour prendre en compte la différence de couverture entre les deux recensements. L'interprétation de la représentation graphique des rapports est discutée plus en détail ci-dessous. Mais, en résumé, la correction à faire aux taux de croissance (*delta*) pour prendre en compte la différence de couverture entre les recensements peut se calculer comme la quantité qui produit un ensemble de rapports de même niveau, c'est-à-dire horizontaux sur le graphique. La feuille de calcul intitulée *Méthode* est préparée de telle manière que le module Solveur (Données / Solveur / Résoudre) trouve la valeur de *delta* qui minimise l'écart absolu entre la moyenne des rapports calculés sur un intervalle d'âge spécifié par l'utilisateur.

On suggère à l'analyste de commencer l'optimisation par la valeur de *delta* qui résulte de la méthode généralisée de la balance de l'accroissement démographique appliquée aux mêmes données. Si cette valeur initiale de *delta* produit un ensemble de rapports de même niveau pour les âges adultes, mais avec une courbure marquée aux âges élevés, ceci pourrait indiquer une baisse de la complétude aux âges élevés, qui pourrait être due à la désintégration de certains ménages suite au décès de l'un de ses membres. Dans ce cas, il conviendra de ne pas chercher un *delta* qui produise un ensemble de rapports de même niveau, mais plutôt de conserver la valeur initiale.

Lorsqu'on cherche à optimiser à la fois *delta* et les espérances de vie de manière itérative, les valeurs des espérances de vie devront être copiées à partir de la feuille de calcul *Espérances de vie* et collées dans la feuille de calcul *Méthode* pour rechercher une nouvelle valeur de *delta*. Ce processus peut être répété deux ou trois fois, jusqu'à ce que les valeurs des espérances de vie soient stables.

Finalement, on doit choisir l'intervalle d'âge sur lequel on calcule la complétude à partir des rapports. Si l'on observe une courbure marquée vers le haut aux âges élevés, on est probablement en présence d'une exagération des âges, surtout pour les décès, et il conviendra d'essayer d'identifier un âge de début de l'intervalle ouvert en dessous duquel cette exagération est négligeable. Le fait que la complétude baisse pour les âges en-dessous de 35 ans, pourrait indiquer des émigrations qui n'ont pas été prises en compte. Si l'on suspecte ce cas, il conviendra d'exclure les âges correspondants pour le calcul de *delta* et de la complétude.

La complétude se calcule à partir des rapports des effectifs des groupes d'âge quinquennaux. Pour obtenir un résultat assez robuste, on propose de prendre la somme pondérée comme suit : 50 % de la médiane et 25 % de chacun des deux quartiles (les percentiles correspondant à 25 % et 75 % de la distribution des rapports).

Étape 8 : Calcul des taux de mortalité corrigés pour tenir compte du sous-enregistrement des décès déclarés

Pour calculer les taux de mortalité corrigés, il faut d'abord corriger le dénominateur, c'est-à-dire les effectifs de population, de la couverture relative du recensement. Si *delta* est négatif, ce calcul se fait en multipliant les effectifs du premier recensement par le coefficient : $\exp(-\delta(t_2 - t_1))$, et ceux du second recensement par 1, (c'est à dire qu'ils restent inchangés) ; si *delta* est positif, on fait l'inverse, c'est à dire que le premier recensement reste inchangé et le second est multiplié par le coefficient correspondant.

Les personnes-années vécues, soit $PYL^a(x,5)$, se calculent en multipliant la moyenne géométrique des populations corrigées par la durée de la période intercensitaire (en années) comme suit :

$$PYL^a(x,5) = (t_2 - t_1) \left({}_5N_x(t_1) e^{-\delta(t_2-t_1)} \times {}_5N_x(t_2) \right)^{1/2} \text{ if } \delta < 0$$

$$PYL^a(x,5) = (t_2 - t_1) \left({}_5N_x(t_1) \times {}_5N_x(t_2) e^{-\delta(t_2-t_1)} \right)^{1/2} \text{ if } \delta > 0$$

Pour tenir compte du sous-enregistrement, on doit ensuite corriger les nombres de décès déclarés, en divisant ces nombres par l'estimation de la complétude, c . Puis on calcule les taux de mortalité corrigés en divisant les décès corrigés par les personnes-années vécues corrigées, $PYL^a(x,5)$, comme suit :

$${}_5m_x = \frac{{}_5D_x / c}{PYL^a(x,5)}$$

Si on pense que la complétude baisse après l'âge à la retraite, on pourra essayer d'améliorer les estimations des taux de mortalité aux âges élevés en remplaçant la valeur trouvée c par une valeur arbitraire correspondant à la complétude dans ces groupes d'âge. La feuille de calcul contient une option pour réaliser cette correction.

D'un point de vue purement technique, on pourrait ignorer la correction due à la différence de couverture des deux recensements, et trouver les mêmes taux de mortalité, car la correction est identique au numérateur et au dénominateur. Mais, dans ce cas, l'estimation de la complétude de l'enregistrement des décès par rapport à la moyenne des deux recensements ne serait plus correcte, puisqu'elle supposerait que les deux recensements ont la même couverture.

Étape 9 : Lissage à l'aide d'un système relationnel logit de tables-type de mortalité

Lorsque les taux de mortalité par âge sont erratiques, on peut les lisser à l'aide d'un modèle. Ceci peut se faire à l'aide du *système relationnel logit* de Brass et d'une table-type du même sexe que l'on considère avoir le même schéma par âge que celui qui s'applique à la population étudiée.

Le classeur *Excel* correspondant à cette méthode (*Estimating deaths.xls*) comporte une feuille de calcul qui permet de produire des taux de mortalité lissés à l'aide d'un *système relationnel logit* et d'une table-type de mortalité. L'utilisateur a le choix entre un standard du modèle *Général* de la famille des tables-type des Nations Unies, ou l'un des quatre modèles régionaux des tables-type de Princeton. La transformation logit de ces tables-type de mortalité, ainsi qu'une table-type d'une population affectée par l'épidémie de sida (Timæus 2004) sont donnés dans la feuille de calcul *Données des modèles*. Cette feuille de calcul permet aussi à l'utilisateur de choisir comme standard une autre table de mortalité sous forme de logit, s'il y a de bonnes raisons de penser de cette table présente un schéma par âge similaire à celui de la population étudiée.

Pour ajuster le modèle logit, on calcule d'abord les quotients quinquennaux de mortalité, ${}_5q_x$, d'après les taux corrigés de mortalité, ${}_5m_x$, selon la formule :

$${}_5q_x = \frac{5 {}_5m_x}{1 + 2,5 {}_5m_x}.$$

Puis on calcule la table de mortalité de proche en proche, en partant d'une racine $l_5 = 1$, selon la formule de récurrence suivante :

$$l_{x+5} = l_x(1 - {}_5q_x).$$

Les coefficients, α et β , du *système relationnel logit* se calculent par régression linéaire, comme suit :

$$\gamma_x = \alpha + \beta \gamma_x^s$$

où la fonction logit est définie comme :

$$\gamma_x = 0,5 \ln \left(\frac{1 - l_x}{l_x} \right)$$

et où l'exposant s désigne les valeurs de la table de mortalité de référence (le standard).

La table de mortalité lissée se calcule ensuite en appliquant les coefficients α and β comme suit :

$$\gamma_x^{fitted} = \alpha + \beta \gamma_x^s$$

et :

$$l_x^{fitted} = \frac{1}{\exp(1 + 2\gamma_x^{fitted})}$$

Les taux de mortalité lissés se calculent dans la table de mortalité comme suit :

$${}_5m_x^{fitted} = \frac{l_x^{fitted} - l_{x+5}^{fitted}}{T_x - T_{x+5}}$$

et

$${}_5m_x^{fitted} = \frac{l_x^{fitted}}{T_x}$$

où

$$T_x = \sum_{x=x,5}^{\omega} \frac{5}{2} (l_x^{fitted} + l_{x+5}^{fitted})$$

soit :

$$T_x = T_{x+5} + \frac{5}{2} (l_x^{fitted} + l_{x+5}^{fitted})$$

où ω désigne l'âge au-delà duquel il ne reste plus de survivant.

Les espérances de vie, qui sont particulièrement importantes si l'on veut estimer les espérances de vie aux âges élevés par itération, se calculent comme suit :

$$e_x = \frac{T_x}{l_x}.$$

EXEMPLE DÉTAILLÉ

Cet exemple utilise les données de la population masculine d'Afrique du Sud, recueillies au recensement de 2001 et à l'enquête de communauté de 2007, et les données des décès enregistrés à l'état civil au cours des années de 2001 à 2007. Le solde migratoire est estimé d'après le changement du nombre de personnes nées à l'étranger repérées aux deux enquêtes, moins une estimation du nombre de sud-africains qui ont émigré à l'étranger entre les deux enquêtes. Cet exemple est détaillé dans le classeur *Excel* intitulé : *SEG_South Africa males*.

Étape 1 : Puisqu'il n'est pas disponible directement, on calcule le nombre de décès déclarés au cours de la période située entre les deux estimations de population

Les décès déclarés au cours des années 2001 à 2007 pour les hommes sud-africains apparaissent dans le tableau 25.1.

TABLEAU 25.1 Calcul des décès de la période intercensitaire, Afrique du sud, sexe masculin, 2001-2007

| Âge | 2001 | 2002-2006 | 2007 | Total entre les deux recensements |
|-------|--------|-----------|--------|-----------------------------------|
| 0-4 | 29 005 | 186 346 | 40 314 | 197 912 |
| 5-9 | 2 118 | 14 733 | 2 854 | 15 566 |
| 10-14 | 1 745 | 10 535 | 2 233 | 11 207 |
| 15-19 | 4 470 | 23 857 | 4 860 | 25 473 |
| 20-24 | 8 931 | 51 588 | 10 875 | 54 960 |
| 25-29 | 16 834 | 96 705 | 18 405 | 102 802 |
| 30-34 | 20 892 | 137 355 | 28 245 | 145 588 |
| 35-39 | 21 068 | 137 502 | 29 258 | 145 900 |
| 40-44 | 19 322 | 128 217 | 26 973 | 135 936 |
| 45-49 | 17 881 | 113 891 | 24 761 | 121 010 |
| 50-54 | 16 883 | 104 508 | 22 790 | 111 157 |
| 55-59 | 14 544 | 90 919 | 21 317 | 96 854 |
| 60-64 | 15 097 | 84 351 | 17 410 | 89 930 |
| 65-69 | 13 011 | 77 680 | 17 878 | 82 843 |
| 70-74 | 14 035 | 68 147 | 13 771 | 73 036 |
| 75-79 | 10 846 | 59 859 | 12 534 | 63 871 |
| 80-84 | 9 161 | 44 986 | 8 872 | 48 163 |
| 85+ | 7 602 | 43 233 | 10 009 | 46 196 |

La date de référence du recensement de 2001 était la nuit du 9 au 10 octobre 2001. L'enquête de communauté se déroula sur plusieurs semaines en février, et on prendra comme date de référence la nuit du 14 au 15 février 2007. Donc, en sup-

posant que les décès sont répartis uniformément au cours de l'année, on peut répartir proportionnellement les décès de 2001 et de 2007, et les ajouter au total des années 2002 à 2006. Ceci donne le total des décès qui se sont produits entre les deux dates de recensement. Par exemple, pour le groupe d'âge 20-24 ans, le nombre se calcule comme suit :

$$\frac{(22 + 30 + 31)}{365}8931 + 51588 + \frac{(31 + 14)}{365}10875 = 54960 .$$

Étape 2 : Calcul des taux de croissance, après correction pour les migrations et pour la différence de couverture entre les deux recensements

Les taux de croissance par âge se calculent en soustrayant les taux de migrations nettes, et après correction pour tenir compte de la différence de couverture des deux recensements : ils sont présentés en colonne 6 du tableau 2. Par exemple, pour le groupe d'âge de 20-24 ans, les effectifs de population apparaissent en colonnes 2 et 3, les migrations nettes en colonne 5 du tableau 25.2, et avec un *delta* estimé plus loin, on trouve :

$${}_5r_{20} = \frac{\ln(2362519 / 2099417)}{5,3541} - \frac{14803}{5,3541(2099417 \times 2362519)^{\frac{1}{2}}} + (-0,00467) = 0,0161$$

où 5,3541 correspond à la durée de la période intercensitaires, calculée par la fonction YEARFRAC du logiciel *Excel*.

Étape 3 : Estimation de l'espérance de vie à l'âge A et aux anniversaires correspondants aux débuts des groupes d'âge quinquennaux, jusqu'à 65 ans

Les estimations provenant de l'application de la méthode généralisée de la balance de l'accroissement démographique en utilisant les mêmes données apparaissent en colonne 2 du tableau 25.3.

Le rapport des décès déclarés dans les groupes d'âge 10-39 ans et 40-59 ans (âge au dernier anniversaire), dont les données figurent en colonne 4 du tableau 25.2, se calcule comme suit :

$$\frac{485930}{464957} = 1,0451 .$$

TABLEAU 25.2 Taux de croissance par âge, et effectifs des personnes qui atteignent l'âge x , et qui atteignent le groupe d'âge x à $x + 5$, calculés d'après le nombre de décès déclarés, Afrique du sud, sexe masculin, 2001-2007

| Age | ${}_5N_x(t_1)$ | ${}_5N_x(t_2)$ | ${}_5D_x$ | ${}_5NM_x$ | ${}_5r_x$ | Est N_x | Est ${}_5N_x$ |
|-----|----------------|----------------|-----------|------------|-----------|-----------|---------------|
| 0 | 2 223 006 | 2 505 744 | 197 912 | 10 605 | 0,0168 | | |
| 5 | 2 425 066 | 2 560 642 | 15 566 | 2 848 | 0,0053 | 2 304 653 | 11 334 968 |
| 10 | 2 518 985 | 2 452 339 | 11 207 | 5 153 | -0,0101 | 2 229 335 | 11 405 753 |
| 15 | 2 453 156 | 2 553 293 | 25 473 | 16 574 | 0,0016 | 2 332 967 | 11 556 063 |
| 20 | 2 099 417 | 2 362 519 | 54 960 | 14 803 | 0,0161 | 2 289 459 | 10 871 687 |
| 25 | 1 899 275 | 2 033 165 | 102 802 | 4 714 | 0,0076 | 2 059 216 | 9 851 950 |
| 30 | 1 594 624 | 1 875 483 | 145 588 | 13 331 | 0,0242 | 1 881 564 | 8 529 425 |
| 35 | 1 441 657 | 1 548 185 | 145 900 | 9 693 | 0,0074 | 1 530 206 | 7 153 512 |
| 40 | 1 233 813 | 1 306 900 | 135 936 | 7 464 | 0,0050 | 1 331 199 | 6 238 580 |
| 45 | 967 744 | 1 104 294 | 121 010 | 8 719 | 0,0184 | 1 164 233 | 5 276 384 |
| 50 | 769 627 | 888 042 | 111 157 | 9 413 | 0,0199 | 946 320 | 4 242 847 |
| 55 | 552 402 | 708 812 | 96 854 | 4 640 | 0,0405 | 750 818 | 3 191 145 |
| 60 | 444 592 | 491 871 | 89 930 | 5 081 | 0,0122 | 525 640 | 2 332 526 |
| 65 | 304 835 | 394 305 | 82 843 | 4 922 | 0,0407 | 407 371 | 1 662 114 |
| 70 | 232 604 | 241 976 | 73 036 | 4 334 | -0,0007 | 257 475 | 1 106 744 |
| 75 | 136 466 | 163 112 | 63 871 | 2 980 | 0,0249 | 185 223 | 721 856 |
| 80 | 90 856 | 87 698 | 48 163 | 1 662 | -0,0148 | 103 519 | 412 486 |
| 85 | 45 920 | 70 299 | 46 196 | 2 009 | 0,0683 | 61 475 | |

TABLEAU 25.3 Espérances de vie calculées selon diverses méthodes, Afrique du sud, sexe masculin, 2001-2007

| x | Méthode généralisée de la balance de l'accroissement démographique | Table-type du modèle Ouest de Princeton | Itérations successives | Delta déterminé par la méthode généralisée de la balance de l'accroissement démographique |
|-----|--|---|------------------------|---|
| 65 | 11,7 | 9,45 | 11,6 | 11,7 |
| 70 | 9,4 | 7,37 | 9,3 | 9,4 |
| 75 | 7,4 | 5,55 | 7,3 | 7,4 |
| 80 | 5,7 | 4,06 | 5,6 | 5,7 |
| 85 | 4,4 | 2,90 | 4,3 | 4,4 |

Ce rapport sert pour entrer dans le système de tables-type de mortalité du modèle Ouest de Princeton (on trouve le tableau dans la feuille de calcul *Espérances de vie* du classeur *Excel*). La table correspondante est identifiée par interpolation, et les espérances de vie correspondantes apparaissent en colonne 3 du tableau 25.3. Par exemple, pour l'espérance de vie à 65 ans :

$$e_{65} = 9,65 + \frac{1,045 - 1,012}{1,062 - 1,012} (9,35 - 9,65) = 9,452 .$$

En partant de ces valeurs de la table-type du modèle Ouest, on peut procéder par itération, et calculer l'espérance de vie et *delta* (comme cela est expliqué plus en détail ci-dessous). On trouve alors une valeur de *delta* = -0,0066, et les estimations de l'espérance de vie apparaissent en colonne 4 du tableau 25.3.

Comme la prévalence du VIH est élevée en Afrique du sud, on ne peut pas utiliser les estimations provenant de la table-type du modèle Ouest, qui sont fournies par la feuille de calcul *Espérances de vie* du classeur *Excel* correspondant. De plus, comme cela est discuté plus loin, il semble que la complétude tende à

baïsser avec l'âge au-delà de 55 ans, et donc la procédure par itération s'applique probablement mal. Donc, dans cet exemple, on choisira une valeur de δ égale à l'ordonnée à l'origine obtenue par la Méthode généralisée de la balance de l'accroissement démographique, appliquée aux mêmes données. Les résultats apparaissent en colonne 5 du tableau 25.3.

Étape 4 : Estimation du nombre de personnes qui atteignent l'âge x et le groupe d'âge x à $x + 4$, (âge au dernier anniversaire) d'après les décès déclarés

Le nombre de personnes qui atteignent l'âge x (leur x -ème anniversaire) entre les deux recensements se calcule à partir du nombre de décès déclarés, en prenant comme intervalle d'âge ouvert 85 ans et plus, les taux de croissance qui figurent en colonne 6 du tableau 25.2, et l'estimation de l'espérance de vie à 85 ans (4,347 ans) qui figure en colonne 5 du tableau 25.3. Les résultats apparaissent en colonne 7 du tableau 25.2. Par exemple, le nombre de personnes qui atteignent 80 ans au cours de la période intercensitaire se calcule comme suit, en prenant le taux de croissance de + 0,0638 dans le groupe d'âge 85 ans et plus, et de – 0,0148 dans le groupe d'âge 80-85 ans :

$$\begin{aligned} \widehat{N}_{85} &= 46196 \left(\exp(0,0638 \times 4,347) \right. \\ &\quad \left. - (0,0638 \times 4,347)^2 / 6 \right) = 61475 \\ \widehat{N}_{80} &= 61475 \exp(5 \times (-0,0148)) \\ &\quad + 48163 \exp(2,5 \times (-0,0148)) = 103519 . \end{aligned}$$

Le nombre de personnes qui atteignent le groupe d'âge de x à $x + 4$ (âge au dernier anniversaire) pendant la période intercensitaire, se calcule d'après les décès déclarés, et apparaît en colonne 8 du tableau 25.2. Par exemple, le nombre de personnes qui atteignent le groupe d'âge 20-24 ans se calcule comme suit :

$${}_5\widehat{N}_{20} = 2,5(2289458,6 + 2059216,1) = 10871687 .$$

Étape 5 : Estimation du nombre de personnes qui atteignent le groupe d'âge x à $x + 4$, (âge au dernier anniversaire) d'après les données du recensement

Le nombre de personnes qui atteignent le groupe d'âge x à $x + 4$ au cours de la période intercensitaire apparaît en colonne 2 du tableau 25.4. Pour le groupe d'âge de 20-24 ans, par exemple, il se calcule à partir des populations

données en colonnes 2 et 3 du tableau 25.2, et de la durée entre les deux recensements, comme suit :

$${}_5N_{20} = 5,3540689(2099417 \times 2362519)^{\frac{1}{2}} = 11923972.$$

TABLEAU 25.4 Nombre de personnes qui atteignent le groupe d'âge x à $x + 4$, calculé d'après les recensements de population ; et ratios des estimations faites à partir des décès et à partir des populations, Afrique du sud, sexe masculin, 2001-2007

| Âge | ${}_5N_x$ obs | $c : {}_5N_x$ | $c : {}_{A-x}N_x$ |
|-----|---------------|---------------|-------------------|
| 0 | 12 636 377 | | |
| 5 | 13 341 976 | 0,8496 | 0,8981 |
| 10 | 13 307 209 | 0,8571 | 0,9050 |
| 20 | 11 923 972 | 0,9118 | 0,9231 |
| 25 | 10 521 174 | 0,9364 | 0,9256 |
| 30 | 9 259 118 | 0,9212 | 0,9230 |
| 35 | 7 998 828 | 0,8943 | 0,9235 |
| 40 | 6 798 761 | 0,9176 | 0,9322 |
| 45 | 5 534 858 | 0,9533 | 0,9371 |
| 50 | 4 426 301 | 0,9586 | 0,9310 |
| 55 | 3 350 250 | 0,9525 | 0,9191 |
| 60 | 2 503 746 | 0,9316 | 0,9028 |
| 65 | 1 856 232 | 0,8954 | 0,8865 |
| 70 | 1 270 220 | 0,8713 | 0,8799 |
| 75 | 798 803 | 0,9037 | 0,8885 |
| 80 | 477 921 | 0,8631 | 0,8631 |

Étape 6 : Calcul des rapports des estimations faites à partir des décès aux estimations faites à partir des recensements

Les rapports des estimations des personnes qui atteignent le groupe d'âge x à $x + 4$ au-cours de la période intercensitaire faites à partir des décès (colonne 8 du tableau 25.2) et à partir des populations (colonne 2 du tableau 25.4) apparaissent en colonnes 3 et 4 du tableau 25.4. Par exemple pour les groupes d'âge de 60-64 ans, et de 65-84 ans, les calculs se font comme suit :

$$\begin{aligned} \frac{{}_5\widehat{N}_{65}}{{}_5N_{65}} &= \frac{1662114}{1856232} = 0,8954 \\ \frac{{}_{20}\widehat{N}_{65}}{{}_{20}N_{65}} &= \frac{1662114 + 1106744 + 721856 + 412486}{1856232 + 1270220 + 798803 + 477921} \\ &= 0,8865. \end{aligned}$$

Étape 7 : Estimation de la complétude de l'enregistrement des décès

En prenant comme *delta* la valeur de l'ordonnée à l'origine obtenue en appliquant la méthode généralisée de la balance de l'accroissement démographique, on obtient une série de rapports qui sont d'un niveau assez stable, mais qui semblent baisser un peu au-delà de l'âge de 50 ans (voir la figure 25.1). Donc, dans cet exemple, on n'a pas utilisé l'optimisation avec le module *Solveur* pour calculer *delta*.

On propose ici d'estimer la complétude à partir des rapports calculés entre les âges de 25 à 64 ans. On fait ce choix pour éviter, dans une certaine mesure, de biaiser l'estimation vers le bas que l'on aurait si on utilisait les rapports aux âges élevés, même si la méthode utilisée est assez robuste aux variations des rapports dans les groupes d'âges quinquennaux. Ce choix donne une estimation de la complétude de 94 %, qui se calcule comme suit :

$$c = 0,5 \times 0,9340 + 0,25(0,9203 + 0,9527) = 0,9353 \\ = 94 \% .$$

où 0,9340 est la médiane, 0,9203 le premier quartile, et 0,9527 le troisième quartile des rapports calculés en colonne 3 du tableau 25.4, pour les âges de 25 à 65 ans.

Étape 8 : Estimation des taux de mortalité corrigés pour tenir compte du sous-enregistrement des décès

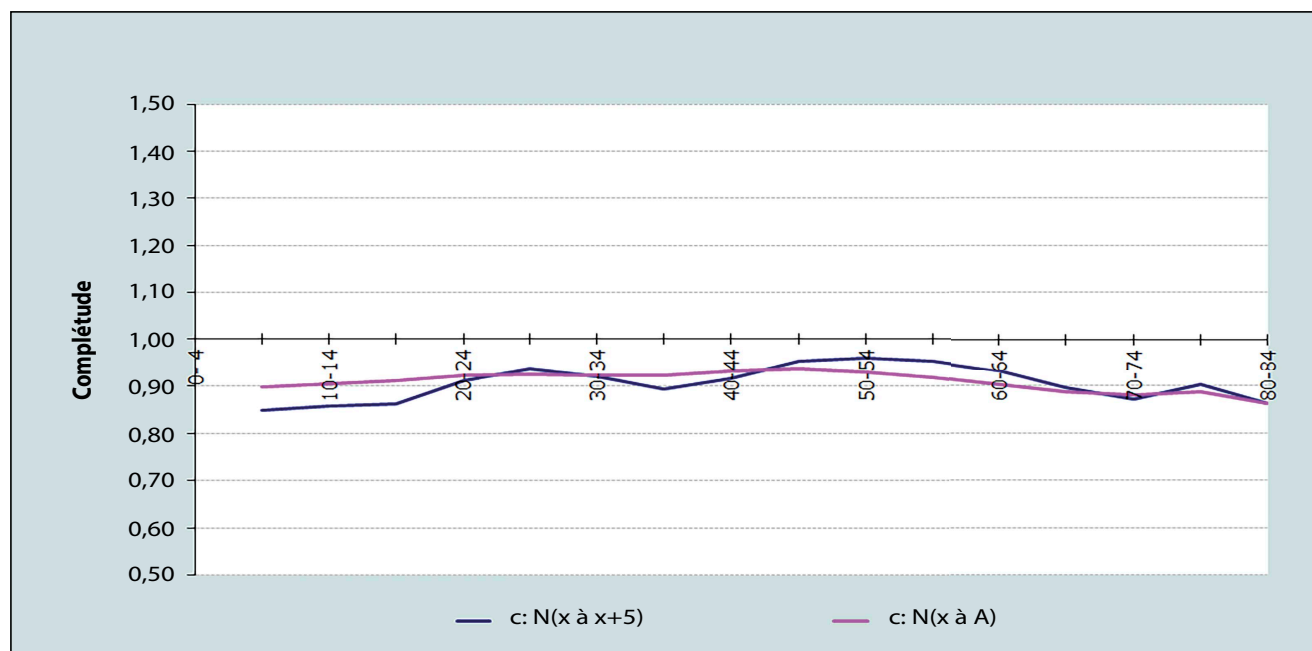
La population corrigée à la date du premier recensement qui apparaît en colonne 2 du tableau 25.5, est égale à la population effectivement recensée (colonne 2 du tableau 25.2), multipliée par le coefficient :

$\exp(-(-0,00467) \times 5,3541)$, puisque *delta* est inférieur à 0. Par exemple, pour le groupe d'âge 20-24 ans le calcul de la population corrigée se fait comme suit :

$$2099\,417 \times \exp(-(-0,00467 \times 5,3541)) = 2152\,629.$$

La population corrigée à la date du second recensement qui apparaît en colonne 3 du tableau 25.5, est égale à la population effectivement recensée (colonne 3 du tableau 25.2), puisque *delta* est inférieur à 0.

FIGURE 25.1 Complétude par âge de la déclaration des décès, Afrique du sud, sexe masculin, 2001-2007



Ensuite, on calcule les décès corrigés pour tenir compte du sous-enregistrement en divisant le nombre de décès déclarés dans chaque groupe d'âge (colonne 4 du tableau 25.2) par l'estimation de la complétude. Les résultats apparaissent en colonne 4 du tableau 25.5. Par exemple, pour le groupe d'âge 20-24 ans, le nombre corrigé s'obtient à partir du nombre observé (54 960 décès), comme suit :

$$\frac{54960}{0,9353} = 58764.$$

Comme il semble que la complétude ait tendance à baisser aux âges élevés, on choisit d'utiliser une complétude qui diffère par âge au-delà de 65 ans. Ainsi, par exemple pour le groupe d'âge 70-74 ans, on prend une nouvelle complétude (0,8713), et on obtient :

$$\frac{73036}{0,8713} = 83824.$$

Les personnes-années vécues corrigées (colonne 5 du tableau 25.5) se calculent comme la moyenne géométrique des populations figurant en colonnes 2 et 3 du tableau 25.5, multiplié par la durée de la période intercensitaire (exprimée

en années), qui dans ce cas est égale à 5,3541 années. Par exemple, pour le groupe d'âge 20-24 ans, le calcul se fait comme suit :

$$PYL(20,5) = 5,354069(2152629 \times 2362519)^{1/2} \\ = 12074140.$$

Les taux de mortalité corrigés, pour tenir compte à la fois du sous-enregistrement des décès et de la différence de couverture entre les recensements (colonne 6 du tableau 25.5) se calculent selon la formule classique, en divisant les décès corrigés par les personnes-années vécues corrigées. Par exemple, pour le groupe d'âge 20-24 ans, le taux de mortalité corrigé se calcule comme suit :

$$\frac{58764}{12074140} = 0,0049.$$

Étape 9 : Lissage en utilisant un système relationnel logit associé à une table-type

On peut dès lors calculer les quotients de mortalité quinquennaux corrigés, ${}_5q_x$, à partir des taux de mortalité corrigés

| Âge | Données corrigées | | | | |
|-----|-------------------|----------------|-----------|-------------|-----------|
| | ${}_5N_x(t_1)$ | ${}_5N_x(t_2)$ | ${}_5D_x$ | $PYL(x\ 5)$ | ${}_5m_x$ |
| 0 | | | | | |
| 5 | 2 486 532 | 2 560 642 | 16 644 | 13 510 001 | 0,0012 |
| 10 | 2 582 831 | 2 452 339 | 11 983 | 13 474 797 | 0,0009 |
| 15 | 2 515 334 | 2 553 293 | 27 236 | 13 568 508 | 0,0020 |
| 20 | 2 152 629 | 2 362 519 | 58 764 | 12 074 140 | 0,0049 |
| 25 | 1 947 414 | 2 033 165 | 109 919 | 10 653 675 | 0,0103 |
| 30 | 1 635 041 | 1 875 483 | 155 667 | 9 375 725 | 0,0166 |
| 35 | 1 478 197 | 1 548 185 | 156 001 | 8 099 564 | 0,0193 |
| 40 | 1 265 085 | 1 306 900 | 145 347 | 6 884 383 | 0,0211 |
| 45 | 992 273 | 1 104 294 | 129 387 | 5 604 563 | 0,0231 |
| 50 | 789 134 | 888 042 | 118 852 | 4 482 045 | 0,0265 |
| 55 | 566 403 | 708 812 | 103 560 | 3 392 442 | 0,0305 |
| 60 | 455 861 | 491 871 | 96 156 | 2 535 277 | 0,0379 |
| 65 | 312 561 | 394 305 | 92 518 | 1 879 609 | 0,0492 |
| 70 | 238 500 | 241 976 | 83 824 | 1 286 217 | 0,0652 |
| 75 | 139 925 | 163 112 | 70 679 | 808 863 | 0,0874 |
| 80 | 93 159 | 87 698 | 55 803 | 483 940 | 0,1153 |
| 85 | 47 084 | 70 299 | 53 524 | 308 032 | 0,1738 |

TABLEAU 25.5 Calcul des taux corrigés de mortalité, Afrique du sud, sexe masculin, 2001-2007

(colonne 6 du tableau 25.5) et qui apparaissent en colonne 2 du tableau 25.6. Par exemple, la probabilité pour une femme de décéder entre 20 et 25 ans se calcule comme suit :

$${}_5q_{20} = \frac{5 \times 0,0049}{1 + 2,5 \times 0,0049} = 0,0240 .$$

On dérive de proche en proche les survivants de la table de mortalité, de 5 ans en 5 ans, selon la formule classique qui lie les survivants à l'âge $x + 5$ aux survivants à l'âge x (colonne 3 du tableau 25.6). Par exemple, les survivants à 25 ans se calculent comme suit :

$$l_{25} = 0,9796(1 - 0,0240) = 0,9560 .$$

La transformation logit appliquée aux proportions de survivants figure en colonne 4 du tableau 25.6. Par exemple, à l'âge de 20 ans, le logit de l_{20} se calcule comme suit :

$$\gamma_{20} = 0,5 \ln \left(\frac{1 - 0,9796}{0,9796} \right) = -1,9350 .$$

Dans la table de mortalité conditionnelle pour le sexe masculin, la transformation logit est appliquée aux proportions de survivants de la table-type spéciale qui tient compte du sida, avec une espérance de vie $e_0 = 50$ (colonne 5 du tableau 25.6) ; les résultats apparaissent en colonne 6 du tableau 25.6. Comme on peut le voir sur la figure 25.2, la table-type qui tient compte du sida n'ajuste pas bien les données, mais elle marche mieux que toute autre table-type qui ne tient pas compte de l'impact du sida sur la mortalité.

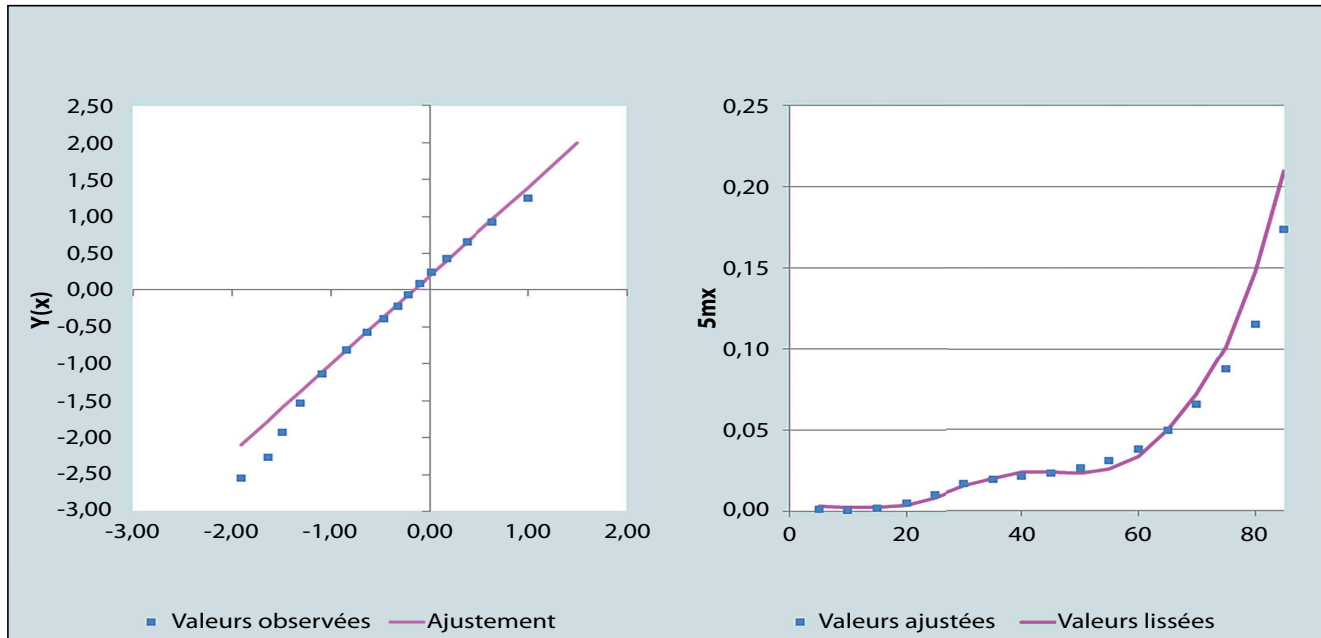
Les coefficients, α et β , sont déterminés comme la pente et l'ordonnée à l'origine de la droite qui ajuste au mieux les transformations logits, qui figurent en colonnes 4 et 6 du tableau 25.6, dans l'intervalle d'âge choisi par l'utilisateur (entre 45 et 80 ans dans cet exemple), soit ici 0,1928 et 1,2008 respectivement.

On applique ensuite ces coefficients aux logits de la table de mortalité conditionnelle, afin de produire les logits lis-

TABLEAU 25.6 Calcul de la table de mortalité conditionnelle, lissée à l'aide du système relationnel logit et d'une table-type de mortalité, Afrique du sud, sexe masculin, 2001-2007

| Âge | ${}_5q_x$ | l_x/l_5 | $Y(x)$ obs. | Sida $l^s(x)$ | Logit standard $Y^s(x)$ | Logit lissé $Y(x)$ | Surviv. lissé $l(x)$ | $T(x)$ | $e(x)$ | Taux lissé ${}_5m_x$ |
|-----|-----------|-----------|----------------|------------------|-------------------------------|--------------------------|----------------------------|--------|--------|----------------------------|
| 0 | | | | | | | | | | |
| 5 | 0,0061 | 1 | | 1,0000 | | | 1 | 51,206 | 51,2 | 0,0030 |
| 10 | 0,0044 | 0,9939 | - 2,5433 | 0,9785 | - 1,9081 | - 2,0984 | 0,9852 | 46,243 | 46,9 | 0,0028 |
| 15 | 0,0100 | 0,9894 | - 2,2705 | 0,9632 | - 1,6326 | - 1,7676 | 0,9717 | 41,351 | 42,6 | 0,0024 |
| 20 | 0,0240 | 0,9796 | - 1,9350 | 0,9512 | - 1,4853 | - 1,5907 | 0,9601 | 36,521 | 38,0 | 0,0041 |
| 25 | 0,0503 | 0,9560 | - 1,5395 | 0,9324 | - 1,3120 | - 1,3827 | 0,9408 | 31,769 | 33,8 | 0,0086 |
| 30 | 0,0797 | 0,9079 | - 1,1444 | 0,8969 | - 1,0818 | - 1,1062 | 0,9014 | 27,164 | 30,1 | 0,0152 |
| 35 | 0,0919 | 0,8356 | - 0,8128 | 0,8420 | - 0,8365 | - 0,8116 | 0,8352 | 22,822 | 27,3 | 0,0200 |
| 40 | 0,1003 | 0,7588 | - 0,5731 | 0,7794 | - 0,6311 | - 0,5650 | 0,7559 | 18,845 | 24,9 | 0,0235 |
| 45 | 0,1091 | 0,6827 | - 0,3831 | 0,7148 | - 0,4593 | - 0,3588 | 0,6721 | 15,275 | 22,7 | 0,0239 |
| 50 | 0,1243 | 0,6082 | - 0,2199 | 0,6560 | - 0,3228 | - 0,1948 | 0,5962 | 12,104 | 20,3 | 0,0230 |
| 55 | 0,1418 | 0,5326 | - 0,0653 | 0,6048 | - 0,2127 | - 0,0626 | 0,5313 | 9,285 | 17,5 | 0,0255 |
| 60 | 0,1732 | 0,4571 | 0,0861 | 0,5530 | - 0,1064 | 0,0650 | 0,4676 | 6,788 | 14,5 | 0,0335 |
| 65 | 0,2191 | 0,3779 | 0,2493 | 0,4918 | 0,0163 | 0,2124 | 0,3954 | 4,631 | 11,7 | 0,0502 |
| 70 | 0,2802 | 0,2951 | 0,4354 | 0,4119 | 0,1781 | 0,4066 | 0,3072 | 2,874 | 9,4 | 0,0718 |
| 75 | 0,3586 | 0,2124 | 0,6553 | 0,3178 | 0,3819 | 0,6513 | 0,2137 | 1,572 | 7,4 | 0,1013 |
| 80 | 0,4475 | 0,1362 | 0,9235 | 0,2173 | 0,6408 | 0,9622 | 0,1274 | 0,719 | 5,6 | 0,1480 |
| 85 | #N/A | 0,0753 | 1,2542 | 0,1201 | 0,9959 | 1,3887 | 0,0586 | 0,255 | 4,3 | 0,2097 |

FIGURE 25.2 Logits observés et logits lissés des probabilités de survie et des taux de mortalité, lissage à l'aide de la table-type de mortalité tenant compte du sida, Afrique du sud, sexe masculin, 2001-2007



sés (colonne 7 du tableau 25.6). Par exemple, pour l'âge de 20 ans, le logit lissé se calcule comme suit :

$$\gamma_{20}^{fitted} = 0,1928 + 1,2008 \times (-1,4853) = -1,5907.$$

Ces valeurs sont ensuite prises pour calculer la table de mortalité lissée (colonne 8 du tableau 25.6). Par exemple, la probabilité de survie à 20 ans se calcule comme suit :

$$l_{20}^{fitted} = \frac{1}{1 + \exp(2 \times (-1,5907))} = 0,9601.$$

Les personnes-années vécues cumulées au-delà de l'âge x , soit T_x , figurent en colonne 9 du tableau 25.6, et se calculent à partir de la table de mortalité lissée. Ces valeurs sont alors utilisées pour calculer les taux de mortalité lissés, qui figurent en colonne 10 du tableau 25.6. Par exemple, à l'âge de 80 ans :

$$T_{80} = 0,255 + \frac{5}{2}(0,1274 + 0,0586) = 0,719$$

$${}_5m_{80}^{fitted} = \frac{0,1274 - 0,0586}{0,719 - 0,2655} = 0,1480.$$

Les espérances de vie, qui apparaissent en colonne 10 du tableau 25.6, se calculent selon la formule classique, en divisant les T_x , qui figurent en colonne 9, par les l_x , qui figurent en colonne 8 du tableau 25.6. Par exemple, l'espérance de vie à 65 ans se calcule comme suit :

$$\frac{4,631}{0,3954} = 11,71.$$

DIAGNOSTICS, ANALYSE ET INTERPRÉTATION

Contrôles et validation

L'estimation de la complétude est de 94 %. La première vérification à faire sur ce résultat est la comparaison avec le résultat correspondant pour l'autre sexe. Ainsi, en appliquant la même méthode que pour les hommes aux données concernant les femmes au cours de la même période (voir la feuille de calcul : *SEG_South Africa_females*) donne une estimation de la complétude de 93 %. Des recherches antérieures (Dorrington, Moultrie and Timæus 2004) ont montré que l'on peut s'attendre à des résultats voisins pour

les deux sexes, et le fait que ces résultats le sont en effet valide par conséquent ces estimations

Une seconde vérification des résultats consiste à les comparer avec ceux obtenus par la Méthode généralisée de la balance de l'accroissement démographique (voir la feuille de calcul : *SEG_South Africa_males*). Cette méthode donne une estimation de la complétude de 92 % sur l'intervalle d'âge de 5 à 84 ans, ce qui aussi suffisamment proche pour valider les résultats précédents

Une troisième vérification consiste à comparer diverses estimations de la mortalité avec d'autres sources, telles que des estimations antérieures pour le pays, ou les « *Perspectives de la population mondiale* » (UN Population Division 2011). L'estimation du quotient de mortalité de 15 à 60 ans, ${}_{45}q_{15}$, est de 519 pour 1 000 après correction pour tenir compte de la sous-déclaration, une valeur proche des 529 pour 1 000 retenue par les « *Perspectives de la population mondiale* » pour la période 2000-2005, ce qui une fois encore montre qu'il y a peu de raisons de mettre en question les résultats obtenus par la méthode.

Il est intéressant de remarquer que l'application de la Méthode de Preston et Coale à ces mêmes données, en prenant pour population moyenne la population au milieu de l'intervalle entre les deux recensements, donne une estimation de 84 % lorsqu'on utilise le même intervalle d'âge. Si on augmente l'âge minimal à 35 ans pour calculer *delta*, on augmente l'estimation à 86 %, toujours un peu plus faible que l'estimation de 94 % produite par la méthode utilisée ci-dessus.

Interprétation

L'analyse des estimations de la complétude (voir la figure 25.1) suggère que la complétude de l'enregistrement des décès semble assez peu variable selon l'âge, du moins au-delà de l'âge de 50 ans. Cette observation est cohérente avec le fait que de nombreux travailleurs migrent aux âges de la retraite des zones urbaines vers les zones rurales, là où la déclaration des décès est plus faible. Comme ces estimations de la complétude sont produites avec un *delta* qui provient de l'application de la méthode généralisée de la balance de l'accroissement démographique, et qu'il est très vraisemblable que nombre de personnes ont effectivement migré de l'urbain vers le rural, réduire la valeur de *delta* pour obtenir des niveaux plus réguliers ne serait pas approprié dans ce cas.

Comme on a déjà pris en compte les migrations, la baisse de l'estimation de la complétude aux jeunes âges est

probablement due à l'effet contraire de celui qui se produit aux âges élevés : les jeunes adultes migrent des zones rurales vers les zones urbaines, pour y trouver un emploi. Il serait donc erroné de laisser les estimations correspondant aux âges de ces jeunes adultes influencer l'estimation finale de la complétude.

Les irrégularités visibles dans la série des rapports des effectifs, ${}_5N_x$, provient des populations au recensement et à l'enquête de communauté, et non des décès. Le caractère erratique de cette série indique probablement des erreurs dans les décomptes de certains groupes d'âge ou bien des erreurs sur les âges dans les populations du recensement ou de l'enquête.

PROBLÈMES D'INTERPRÉTATION SPÉCIFIQUES DE LA MÉTHODE

Source des données sur les décès déclarés

On trouve en général deux types de problèmes avec les données concernant les décès : ceux qui conduisent à un biais de déclaration (sous- ou sur-déclaration) qui est constant par âge, ce qui est précisément ce que la méthode vise à corriger, et ceux qui conduisent à des biais de déclaration qui varient selon l'âge, qui peuvent entraîner une distorsion dans les estimations finales. Bien que l'approche générale reste essentiellement la même quelle que soit la source des données, des sources différentes peuvent être sujettes à des biais spécifiques, qui peuvent avoir un effet sur l'interprétation des résultats. On va illustrer ces cas par des exemples concrets, et d'une manière générale, l'analyse devra prêter attention aux biais suivants dans les données concernant les décès.

1) Décès déclarés à l'état civil

Si la répartition de la population entre les zones urbaines et rurales (ou une division équivalente) diffère significativement par âge, et que la complétude de la déclaration des décès en milieu urbain est significativement supérieure à celle du milieu rural, alors l'hypothèse selon laquelle la complétude est indépendante de l'âge sera vraisemblablement violée, et la complétude va tendre à baisser avec l'âge au-delà de 50 ans si la proportion de personnes qui migrent de l'urbain vers le rural au moment de leur retraite. Si on l'ignore, cette violation de l'hypothèse produira probablement une sous-estimation du niveau moyen de la complétude.

2) Décès déclarés par les ménages

Ces données sont sujettes à quatre problèmes potentiels :

- Si une proportion significative de ménages sont dissous après le décès d'une personne clé (comme la seule personne à subvenir aux besoins du ménage), alors le décès de cette personne risque de ne pas être déclaré, ce qui conduit à violer l'hypothèse que la complétude est constante selon l'âge. Si une proportion significative de décès dans certains groupes d'âge concerne des individus qui ne vivent pas dans des ménages ordinaires (par exemple, ceux qui vivent dans des maisons pour personnes âgées), la violation de l'hypothèse peut même être encore plus sévère. Mais ce n'est pas encore un problème dans la plupart des pays en développement.
- Dans les situations où les jeunes adultes quittent leur famille lorsqu'ils grandissent, pour aller vivre en ville, il est possible qu'ils soient considérés comme membres de plusieurs ménages (ou d'aucun ménage), et leurs décès peuvent être déclarés plus d'une fois (ou jamais), ce qui conduit encore à une violation de l'hypothèse de déclaration des décès constante selon l'âge. Dans ce cas, on peut en limiter l'impact en ignorant les données concernant les décès d'âge inférieur à un âge donné pour estimer la complétude.
- Erreurs sur la période de référence. Comme il arrive fréquemment qu'il y ait confusion sur la période précise au cours de laquelle on doit déclarer les décès, sans même parler d'erreurs sur la date précise du décès, il est possible que cela se traduise par soit une sous-déclaration, soit une sur-déclaration des décès. En supposant que l'on puisse faire l'hypothèse que ces biais sont indépendants de l'âge des décédés, cette distorsion sera prise en compte dans l'estimation de la complétude, et ne constituera pas un problème pour le calcul des taux de mortalité corrigés.
- La période de référence ne couvre qu'une petite partie de la période intercensitaire. Par exemple, on demande fréquemment aux ménages de déclarer les décès qui ont eu lieu au cours des 12 derniers mois précédant le recensement. Non seulement une période très courte comme celle-ci produit de fortes fluctuations aléatoires, mais de plus cela pose le problème supplémentaire de l'estimation de la population au début de cette période de référence. On illustrera ce point dans les exemples ci-dessous, et on expliquera comment résoudre le problème. Par contre, si l'on dis-

pose en plus des décès déclarés par les ménages au premier recensement, on peut utiliser les deux ensembles de données sur les décès pour estimer le nombre de décès qui se sont produits entre les deux recensements, comme cela a été expliqué ci-dessus. Toutefois, comme la question relative aux décès des 12 derniers mois au sein des ménages n'a été posée que relativement peu fréquemment avant le cycle des recensements des années 2010, il se peut que l'on ait un seul ensemble de données sur les décès. Dans ce cas, s'il n'y a pas de raison de présumer que le schéma par âge de mortalité ait changé rapidement au cours de la période, on recommande de calculer les taux de mortalité par âge pour l'année concernée, et d'appliquer ces taux aux personnes-années vécues au cours de l'intervalle intercensitaire, de manière à obtenir une estimation des décès au cours de la période. Mais si on a de bonnes raisons de penser que la mortalité a changé rapidement, par exemple du fait de l'épidémie de VIH/sida, alors cette adaptation va vraisemblablement biaiser les résultats, soit en sous-estimant, soit en surestimant la mortalité, et dans ce cas on ne recommande pas d'utiliser cette méthode de répartition des décès.

3) Décès enregistrés dans les centres de santé

On sait peu de choses à propos de cette source de données, ni comment elle fonctionne. On peut cependant supposer que la complétude dépendra de la répartition géographique des centres de santé auprès desquels les données seront collectées. Dans de nombreux pays en développement, ce genre de services a tendance à être concentré dans les zones urbaines. Dans ce cas, une nouvelle fois, si la proportion de la population vivant en milieu urbain plutôt qu'en milieu rural varie selon l'âge, on ne pourra pas supposer que la complétude est indépendante de l'âge. Il est aussi possible que certaines causes de décès soient plus fréquentes dans les centres de santé, et que si ces causes sont importantes en nombre et qu'elles sont liées à l'âge, tout ceci pourra entraîner une nouvelle violation de l'hypothèse de la complétude constante selon l'âge.

Dans tous les cas mentionnés ci-dessus, on devrait résister à la tentation d'ajuster *delta* de manière à obtenir une série de rapports de même niveau, et on devrait plutôt s'assurer que l'estimation de *c* est calculée à partir d'un intervalle d'âges qui exclut les âges où l'on observe les distorsions.

Interprétation diagnostique générale

En pratique, les deux séries des rapports \widehat{N}_x / N_x et $\widehat{A}N_x / AN_x$ sont susceptibles d'être affectées par des violations des diverses hypothèses. Mais une des forces de cette méthode est que la plupart des violations des hypothèses produit des déviations caractéristiques et distinctes du schéma attendu de points alignés horizontalement, et dans certaines circonstances ces déviations peuvent être interprétées. En voici quelques exemples :

- *Estimation erronée de la couverture des recensements.*
Lorsque delta est trop élevé, la séquence des points est assez linéaire, mais elle tend à augmenter avec l'âge vers le niveau sous-jacent de complétude, et vice-versa, comme on peut s'y attendre d'après l'équation 1 présentée ci-dessous. L'effet est plus fort pour \widehat{N}_x / N_x que pour $\widehat{A}N_x / AN_x$.
- *Exagération de l'âge déclaré.* Il est fréquent que les proches des défunts qui déclarent les décès exagèrent l'âge des décédés, plus que ne le font les personnes vivantes déclarant leur propre âge au recensement. Ce biais produit une séquence de point d'abord régulière, mais qui tend à augmenter rapidement au-delà de l'âge auquel se produit l'exagération. D'après l'équation 1 présentée ci-dessous, l'exagération de l'âge tend à augmenter le nombre de décès aux âges élevés. De plus, les transferts entre les groupes d'âge font que le nombre de décès est multiplié par un coefficient exponentiel qui est plus fort, même si ce second effet a beaucoup moins d'impact. En théorie, un tel biais pourrait aussi être produit par une augmentation de la complétude avec l'âge, au-delà d'un certain âge, mais ce cas ne se produit pas en pratique (Preston, Coale, Trussell *et al.* 1980).
- *Erreurs sur l'âge au recensement, et couverture variable selon l'âge.* Quand ces phénomènes se produisent, on observe une séquence irrégulière des rapports dans l'intervalle d'âge considéré. Comme les effectifs sont de nature cumulative, ils tendent à suivre d'assez près la répartition par âge de la population. Ainsi donc, si on observe des zigzags, il est vraisemblable que les pics soient associés à une surestimation de la population, et les baisses à une sous-estimation. Si ces fluctuations sont indépendantes de l'âge, elles n'introduiront pas de distorsion notable dans les estimations de la complétude. Par contre si elles sont systématiques, par exemple du fait de migrations non-prises en compte en-dessous d'un certain âge, il pourrait être souhaitable d'exclure ces points lors de l'estimation finale de la complétude.

TABEAU 25.7 Décès des 12 derniers mois déclarés au sein des ménages enquêtés au recensement de 2001 et à l'enquête de communauté de 2007, Afrique du sud

| Âge | Recensement, 2001 | | Enquête de communauté, 2007 | |
|-------|-------------------|--------------|-----------------------------|--------------|
| | Sexe masculin | Sexe féminin | Sexe masculin | Sexe féminin |
| 0-4 | 35 873 | 32 096 | 48 322 | 44 418 |
| 5-9 | 3 868 | 3 155 | 4 505 | 5 216 |
| 10-14 | 2 590 | 2 284 | 3 442 | 3 259 |
| 15-19 | 5 628 | 5 122 | 8 246 | 7 878 |
| 20-24 | 10 976 | 13 246 | 16 360 | 21 702 |
| 25-29 | 17 787 | 19 727 | 27 551 | 35 840 |
| 30-34 | 20 038 | 18 292 | 34 832 | 42 576 |
| 35-39 | 19 816 | 15 521 | 38 061 | 34 809 |
| 40-44 | 17 417 | 12 124 | 33 604 | 28 823 |
| 45-49 | 15 840 | 10 105 | 27 829 | 20 973 |
| 50-54 | 15 077 | 9 144 | 28 223 | 18 891 |
| 55-59 | 12 781 | 7 755 | 22 868 | 13 118 |
| 60-64 | 13 428 | 10 367 | 18 775 | 14 912 |
| 65-69 | 11 820 | 10 195 | 17 532 | 14 298 |
| 70-74 | 11 885 | 10 809 | 14 879 | 14 645 |
| 75-79 | 8 794 | 8 393 | 12 966 | 14 151 |
| 80-84 | 7 484 | 9 371 | 9 204 | 12 063 |
| 85+ | 7 115 | 12 389 | 11 735 | 18 178 |

EXEMPLES TIRÉS DES DÉCÈS DÉCLARÉS AU SEIN DES MÉNAGES LORS D'UN RECENSEMENT OU D'UNE ENQUÊTE

Les exemples décrits ci-dessous utilisent les mêmes données que celles qui sont utilisées dans les classeurs *Excel* suivants : *SEG_South Africa_males* et *SEG_South Africa_females*. Mais, au lieu d'utiliser les données des décès déclarés à l'état civil, on partira des décès des 12 derniers mois déclarés au sein des ménages au recensement de 2001 et à l'enquête de communauté de 2007. Ces nombres de décès apparaissent dans le tableau 25.7.

Les nombres de décès qui se sont produits entre la date moyenne du premier recensement (la nuit du 9 au 10 octobre 2001) et la date moyenne de l'enquête de communauté (considérée être la nuit du 14 au 15 février 2007) se calculent en utilisant les classeurs *Excel* suivants : *Estimating deaths_South Africa_males_bhd* et *Estimating deaths_South Africa_females_bhd*.

L'application de la méthode de l'extinction des cohortes synthétiques aux mêmes données concernant le sexe masculin, et en utilisant comme estimation de la couverture relative des recensements, *delta*, celle fournie par la méthode généralisée de la balance de l'accroissement démographique appliquée aux mêmes données, que l'on peut trouver dans le classeur *Excel SEG_South Africa_males_hbd*, suggère que le décompte des décès des 12 dernier mois est pratiquement aussi complet que celui de l'état civil. Cependant, les données relatives aux décès des 12 derniers mois donnent un quotient de mortalité de 15 à 60 ans, $_{45}q_{15}$, égal à 539 pour mille, qui est légèrement supérieur à celui obtenu à partir des données de l'état civil, même s'il reste très proche.

L'application de la méthode de l'extinction des cohortes synthétiques aux mêmes données concernant le sexe féminin, que l'on peut trouver dans le classeur *Excel SEG_South Africa_females_hbd*, suggère que le décompte des décès des 12 dernier mois est nettement moins fiable que celui de l'état civil. Les données des 12 derniers mois donnent un quotient de mortalité de 15 à 60 ans, $_{45}q_{15}$, égal à 493 pour

mille, qui est nettement supérieur à celui obtenu à partir des données de l'état civil (420 pour 1 000), et moins plausible si on le compare au quotient équivalent pour les hommes.

Les raisons de cette performance plus médiocre de la méthode lorsqu'elle est appliquée aux décès féminins déclarés au sein des ménages peuvent être recherchées dans la comparaison des nombres de décès déclarés par au sein des ménages avec les données de l'état civil après correction pour tenir compte du sous-enregistrement, comme cela apparaît dans le tableau 25.8. On peut y voir que le nombre de décès féminins déclarés par les ménages est nettement inférieur à celui de l'état civil au-delà de l'âge de 55 ans, probablement du fait de la dissolution des ménages suite à leurs décès, car beaucoup de ces ménages ont pour chef une femme âgée, susceptible de décéder. On peut aussi y voir une sur-déclaration des décès des hommes de moins de 30 ans et des femmes de moins de 25 ans, ou peut-être une sous-déclaration de ces décès à l'état civil.

Dans le cas où l'on ne dispose des décès des 12 derniers mois que dans le second recensement, on peut encore estimer les décès par âge dans l'intervalle intercensitaire, en

| Âge | Sexe masculin | | | Sexe féminin | | |
|-------|---------------|----------|---------|--------------|----------|---------|
| | Déclarés | Attendus | Rapport | Déclarés | Attendus | Rapport |
| 0-4 | | | | | | |
| 5-9 | 22 683 | 16 979 | 134 % | 22 995 | 14 575 | 158 % |
| 10-14 | 16 462 | 12 224 | 135 % | 15 173 | 10 349 | 147 % |
| 15-19 | 38 013 | 27 784 | 137 % | 35 666 | 26 874 | 133 % |
| 20-24 | 74 934 | 59 946 | 125 % | 95 993 | 84 611 | 113 % |
| 25-29 | 124 403 | 112 129 | 111 % | 152 718 | 154 437 | 99 % |
| 30-34 | 150 792 | 158 796 | 95 % | 166 488 | 170 680 | 98 % |
| 35-39 | 159 016 | 159 137 | 100 % | 137 837 | 141 399 | 97 % |
| 40-44 | 140 172 | 148 269 | 95 % | 111 910 | 115 746 | 97 % |
| 45-49 | 120 016 | 131 988 | 91 % | 85 284 | 93 408 | 91 % |
| 50-54 | 118 989 | 121 242 | 98 % | 76 941 | 81 793 | 94 % |
| 55-59 | 97 977 | 105 641 | 93 % | 57 353 | 72 131 | 80 % |
| 60-64 | 88 088 | 98 089 | 90 % | 69 220 | 78 877 | 88 % |
| 65-69 | 80 451 | 90 359 | 89 % | 67 007 | 86 099 | 78 % |
| 70-74 | 72 827 | 79 663 | 91 % | 69 536 | 93 404 | 74 % |
| 75-79 | 59 632 | 69 665 | 86 % | 61 942 | 88 314 | 70 % |
| 80-84 | 45 365 | 52 533 | 86 % | 58 410 | 77 084 | 76 % |
| 85+ | 51 779 | 50 387 | 103 % | 83 753 | 108 002 | 78 % |

TABEAU 25.8 Rapport des décès des 12 derniers mois déclarés au sein des ménages aux estimations attendues des décès déclarés à l'état civil après correction pour tenir compte du sous-enregistrement, Afrique du sud

partant seulement du second recensement, et en utilisant la formule suivante :

$$\frac{{}_5D_x(t_2)}{{}_5N_x(t_2)}(t_2 - t_1)({}_5N_x(t_1) \times {}_5N_x(t_2))^{1/2}$$

Dans le cas de l'Afrique du sud, en ne considérant que les données des décès des 12 derniers mois fournis par l'enquête de communauté de 2007, on trouve une estimation de la probabilité de décéder de 15 à 60 ans de 581 pour 1 000 pour les hommes et de 556 pour 1 000 pour les femmes. Certes, on a utilisé ici les données de la seconde enquête, à une date où la mortalité avait augmenté du fait de l'épidémie de VIH/sida par rapport à la période antérieure. Il est certain que calculer la mortalité à partir d'une seule année, qui plus est sur un échantillon de taille assez faible, produira vraisemblablement des estimations moins fiables, et dans ce cas surtout pour les femmes. D'autres études (Bradshaw, Dorrington et Laubscher 2012) suggèrent que la probabilité de décès pour l'année 2006 est plus proche de 550 pour 1 000 pour les hommes, et de 450 pour 1 000 pour les femmes.

DESCRIPTION DÉTAILLÉE DE LA MÉTHODE

Exposé mathématique

Le principe de base de la Méthode de l'extinction des cohortes synthétiques fut énoncé par Paul Vincent (1951) : le nombre de personnes qui atteignent leur anniversaire à un instant donné doit être égal aux décès cumulés au-delà de cet âge dans la même cohorte, jusqu'au décès du dernier survivant. Cette méthode est simplement une généralisation de la méthode proposée par Preston, Coale, Trussell *et al.* (1980), qui, elle, exige que la population soit stable. Dans une cohorte suivie en longitudinal on a donc la relation suivante :

$$N(a, t) = \int_0^{\omega - a} D(a + s, t + s) ds$$

où $N(a, t)$ représente le nombre de personnes de la cohorte qui atteignent l'âge a à l'instant t , et $D(a + s, t + s)ds$ représente le nombre de décès au temps $t + s$, qui sont donc d'âge $a + s$, dans l'intervalle de temps infinitésimal ds .

En faisant l'hypothèse que les taux de mortalité sont constants au cours de l'intervalle, et en remarquant que le

nombre de décès à un âge donné doit croître à un taux égal au taux cumulé de croissance de la population du temps t à $t + s$, on peut calculer le nombre de décès dans l'intervalle selon la formule :

$$D(a + s, t + s) = D(a + s, t) e^{\int_0^s r(a + s, t + z) dz}$$

où $r(a + s, t + z)dz$ représente le taux de croissance annuel de la population d'âge $a + s$ dans l'intervalle de temps de $t + z$ à $t + z + dz$. Si la mortalité est constante dans le temps, on peut démontrer que $r(a + s, t + z) = r(a + s - z, t)$ et donc que :

$$N(a, t) = \int_0^{\omega - a} D(a + s, t) e^{\int_0^s r(a + s - z, t) dz} ds$$

Qui peut s'écrire comme :

$$N(a, t) = \int_a^{\omega} D(x, t) e^{\int_a^x r(y, t) dy} dx \quad (1)$$

On peut donc calculer la population attendue à chaque âge en connaissant les décès par âge qui se produisent sur un intervalle de temps et les taux de croissance par âge de la population. La comparaison de cette population attendue avec la population observée à un recensement peut fournir une estimation de la complétude de l'enregistrement des décès par rapport à celle de la population.

Mise en œuvre pratique de la méthode

Comme les données avec lesquelles on travaille sont en général sujettes à des erreurs sur l'âge, on travaille le plus souvent avec des groupes d'âge quinquennaux.

Considérons ici le cas où l'on dispose des données suivantes : ${}_5D_x^r$ le nombre de décès déclarés sur une période de plusieurs années, entre les instants t_1 et t_2 , répartis en groupe d'âge quinquennaux, notés, jusqu'à un intervalle d'âge ouvert commençant à l'âge A , noté ${}_{\infty}D_A^r$; et les effectifs de population aux mêmes instants t_1 et t_2 , répartis selon les mêmes groupes d'âge (provenant de recensements exhaustifs), notés ${}_5N_x^r(t)$ et ${}_{\infty}N_A^r(t)$ (où t désigne soit t_1 soit t_2). On peut utiliser ces données directement pour appliquer la méthode : on calcule ${}_{\infty}D_x^r$ et ${}_{\infty}N_x^r$, on prend comme approximation de N_x^r la quantité : $(t_2 - t_1)({}_5N_{x-5}^r(t_1) \times {}_5N_x^r(t_2))^{1/2} / 5$

et $\int_{t_1}^{t_2} {}_{\infty}N_x^r(t) dt$ pour la quantité :

$$(t_2 - t_1) \left({}_{\infty}N_x^r(t_1) \times {}_{\infty}N_x^r(t_2) \right)^{1/2}$$

Comme cela a été suggéré par Bennett et Horiuchi (1981), on peut calculer l'expression contenue dans l'équation 1 par récurrence, selon la formule suivante :

$$N_x = N_{x+5} \exp[5 \times {}_5r_x] + {}_5D_x \exp[2,5 \times {}_5r_x], \quad (2)$$

où N_x représente le nombre de personnes qui atteignent l'âge x entre les deux recensements, c'est-à-dire entre t_1 et t_2 , et où ${}_5r_x$ représente le taux de croissance annuels de la population qui atteint le groupe d'âge de x à $x + 4$ (âge au dernier anniversaire) entre les instants t_1 et t_2 .

Bennett et Horiuchi (1981, 1984) proposent d'utiliser les taux de croissance par groupe d'âge pour améliorer à la fois \widehat{N}_x aux âges élevés et \widehat{N}_A , où A désigne l'âge au début du dernier intervalle d'âge, l'intervalle ouvert. Ils suggèrent de calculer \widehat{N}_A comme suit :

$$\widehat{N}_A = {}_{\infty}D_A^r \left[\exp({}_{\infty}r_A \times e_A) - ({}_{\infty}r_A \times e_A)^2 / 6 \right]$$

où ${}_{\infty}D_A^r$ représente les décès déclaré à l'âge A et plus, et e_A représente l'espérance de vie à l'âge A .

Pour tenir compte de la courbure aux âges élevés, ils suggèrent aussi de modifier l'équation 2 comme suit :

$$\widehat{N}_x = \widehat{N}_{x+5} \exp[5 \times {}_5r_x] + {}_5\gamma_x \times {}_5D_x^r \exp[2,5 \times {}_5r_x]$$

où :

$${}_5\gamma_x = 1,00 - 2,26 \times {}_5r_x \frac{{}_5D_x^r}{{}_5N_x} + 0,218 \times {}_5r_x - 0,826({}_5r_x)^2.$$

De plus, ils suggèrent de faire une approximation au-delà de l'âge de 60 ans, imposant la courbe d'une population stable sur les intervalles de 5 ans, et d'estimer la quantité : \widehat{N}_x en calculant l'aire située en dessous cette courbe (Bennett and Horiuchi 1981 : 210). Mais en pratique, les données sont rarement suffisamment précises pour permettre un tel raffinement, et ce dernier ne ferait guère de différence sur l'estimation finale de la complétude.

En pratique, pour corriger, au moins en partie, les effets de la préférence pour les chiffres ronds dans les âges déclarés, et aussi pour rendre les données des décès compatibles avec celles de la population pour les groupes d'âge quinquennaux conventionnels, on calcule cette quantité de la manière suivante : $\widehat{N}_x = 2,5(\widehat{N}_x + \widehat{N}_{x+5})$. De plus, comme la séquence des rapports $\widehat{N}_x / {}_5N_x$, et même la séquence des rapports $\widehat{N}_{x-5} / {}_{10}N_{x-5}$ comme le font remarquer

Bennett et Horiuchi (1981), est en général assez erratique du fait des erreurs sur l'âge et des omissions différentielles des personnes dans certains groupes d'âge, il est courant de faire l'hypothèse que le sous-enregistrement est approximativement constant selon l'âge pour tous les âges supérieurs à un âge minimal, comme par exemple 10 ans. On calcule alors cette proportion fixe, c , en prenant soit la moyenne soit la médiane des rapports $\widehat{N}_x / {}_5N_x$ sur un intervalle d'âge donné (on peut aussi, si nécessaire, corriger arbitrairement les taux de croissance par âge, pour tenir compte de la différence de couverture des deux recensements). En fait, la différence de complétude des deux recensements est prise en compte en ajoutant une constante, δ , aux taux de croissance par âge calculés directement à partir des effectifs des populations recensées, δ étant choisi pour produire une séquence de même niveau des rapports : $\widehat{N}_x / {}_5N_x$. Cette relation peut s'établir par les dérivations suivantes. Supposons que : ${}_{\infty}N_x^r(t_1) = k_1 {}_{\infty}N_x(t_1)$ et ${}_{\infty}N_x^r(t_2) = k_2 {}_{\infty}N_x(t_2)$. Alors on a :

$$\begin{aligned} {}_5r_x^r &= \ln \left(\frac{{}_{\infty}N_x^r(t_2)}{{}_{\infty}N_x^r(t_1)} \right) / t = \ln \left(\frac{k_2 {}_{\infty}N_x(t_2)}{k_1 {}_{\infty}N_x(t_1)} \right) / t \\ &= {}_5r_x + \ln \left(\frac{k_2}{k_1} \right) / t = {}_5r_x - \delta \end{aligned}$$

où :

$$\delta = \ln \left(\frac{k_1}{k_2} \right) / t.$$

Alors, $1 - \delta \cdot t$, où t est la durée de l'intervalle intercensitaire, donne une indication de la différence de couverture des deux estimations de la population utilisées pour le calcul des taux de croissance par âge.

En général on donne une représentation graphique des rapports dans les groupes d'âge quinquennaux : $\widehat{N}_x / {}_5N_x$, et des rapports dans les groupes d'âge cumulés : $\widehat{N}_x / {}_A N_x$, définis comme :

$${}_A \widehat{N}_x = \sum_{a=x}^{A-5} {}_5 \widehat{N}_a.$$

Cette dernière série tend à être plus stable, et facilite l'interprétation des données.

Extensions de la méthode

Si les âges sont précis, et si l'hypothèse de la couverture constante par âge des recensements est vérifiée (ce qui est

peu probable), alors la méthode peut être adaptée aux cas où la complétude de la déclaration des décès n'est constante que sur un intervalle d'âges limité, (x to $x + n$). Dans ce cas on applique une version tronquée de la méthode, qui exclut des calculs les décès et la population d'âge supérieur ou égal à $x + n$. Cette adaptation peut s'appliquer aux données de l'état civil, par exemple, dans le cas où la complétude baisse au-delà de l'âge de la retraite, lorsque les personnes

migrent des zones urbaines vers les zones rurales. Elle peut aussi s'appliquer aux décès déclarés par les ménages, lorsque ceux-ci sont sous-estimés car les ménages sont dissous suite au dernier décès d'adulte. Cependant, cette méthode a peu de chance d'être robuste, sauf peut-être quand l'âge $x + n$ est élevé. Par contre, une adaptation analogue de la Méthode généralisée de la balance de l'accroissement démographique est plus simple, et semble un peu plus robuste.

AUTRES LECTURES ET RÉFÉRENCES

L'analyse de la sensibilité de cette méthode aux erreurs de données les plus fréquentes et aux violations des hypothèses sous-jacentes reste assez limitée. Toutefois, on renvoie le lecteur aux documents suivants : à Hill, You et Choi (2009) pour ce qui concerne les hypothèses de la méthode de la répartition des décès dans les populations sans VIH/sida; et à Dorrington et Timæus (2008) pour le cas des populations souffrant d'une forte épidémie de VIH/sida. Par contre, Murray, Rajaratnam, Marcus *et al.* (2010) présentent une analyse de sensibilité utilisant des méthodes stochastiques. Ils concluent que ces méthodes ne sont pas vraiment fiables, mais peut-être cette conclusion est-elle due à des hypothèses irréalistes sur les migrations.

Bennett NG and S Horiuchi. 1981. "Estimating the completeness of death registration in a closed population", *Population Index* **47**(2):207-221. doi: <http://dx.doi.org/10.2307/2736447>

Bennett NG and S Horiuchi. 1984. "Mortality estimation from registered deaths in less developed countries", *Demography* **21**(2):217-233. doi: <http://dx.doi.org/10.2307/2061041>

Blacker J. 1988. *An Evaluation of the Pakistan Demographic Survey*. Karachi: Pakistan Federal Bureau of Statistics. <http://dx.doi.org/10.2307/2061041>

Bradshaw D, RE Dorrington and R Laubscher. 2012. *Rapid Mortality Surveillance Report 2012*. Cape Town: South African Medical Research Council. <http://www.mrc.ac.za/bod/Rapid-Mortality2011.pdf>

Dorrington RE, TA Moultrie and IM Timæus. 2004. *Estimation of mortality using the South African 2001 census data*. Monograph 11. Centre for Actuarial Research, University of Cape Town. <http://www.commerce.uct.ac.za/care/Monographs/Monographs/Mono11.pdf>

Dorrington RE and IM Timæus. 2008. "Death Distribution Methods for Estimating Adult Mortality: Sensitivity Analysis with Simulated Data Errors, Revisited," Paper presented at Population Association of America 2008 Annual Meeting, New Orleans, Louisiana, 17-19 April.

Hill K, D You and Y Choi. 2009. "Death distribution methods for estimating adult mortality: Sensitivity analysis with simulated data error", *Demographic Research* **21**(Article 9):235-254. doi: <http://dx.doi.org/10.4054/DemRes.2009.21.9>

Murray CJL, JK Rajaratnam, J Marcus, T Laakso and AD Lopez. 2010. "What can we conclude from death registration? Improved methods for evaluating completeness", *PLoS Med* **7**(4):e1000262. doi: <http://dx.doi.org/10.1371/journal.pmed.1000262>

Preston SH, AJ Coale, J Trussell and M Weinstein. 1980. "Estimating the completeness of reporting of adult deaths in populations that are approximately stable", *Population Index* **46**:179-202. doi: <http://dx.doi.org/10.2307/2736122>

Timæus IM. 2004. "Impact of HIV on mortality in Southern Africa: Evidence from demographic surveillance," Paper presented at Seminar of the IUSSP Committee "Emerging Health Threats" HIV, Resurgent Infections and Population Change in Africa. Ougadougou, 12-14 February.

UN Population Division. 2011. *World Population Prospects: The 2010 Revision, Volume I: Comprehensive Tables*. New York: United Nations, Department of Economic and Social Affairs, ST/ESA/SER.A/313. http://esa.un.org/unpd/wpp/Documentation/pdf/WPP2010_Volume-I_Comprehensive-Tables.pdf

Vincent P. 1951. « La mortalité des vieillards », *Population* **6**:182-204. doi: <http://dx.doi.org/10.2307/1524149>

CHAPITRE 26 ESTIMATION INDIRECTE DE LA MORTALITÉ ADULTE À PARTIR DES PROPORTIONS D'ORPHELINES ISSUES DE PLUSIEURS ENQUÊTES OU RECENSEMENTS

Ian M Timæus

DESCRIPTION DE LA MÉTHODE

La méthode de base des proportions d'orphelins permet d'estimer indirectement la mortalité des femmes et des hommes adultes à partir des renseignements donnés par les répondants sur la survie de leurs mères et/ou de leurs pères. Les données nécessaires pour appliquer la méthode proviennent en général des recensements ou des enquêtes. Ces données sont collectées à partir des réponses aux questions simples suivantes : « Votre mère est-elle vivante ? » ou « Votre père est-il vivant ? » En utilisant une estimation de l'âge moyen des mères et des pères à la naissance de leurs enfants dans la population étudiée, il est possible de convertir la proportion des répondants de chaque groupe d'âge dont la mère ou le père est vivant lors de l'enquête en probabilité de survie à l'âge adulte (l_{25+n} / l_{25} pour les femmes et l_{35+n} / l_{35} pour les hommes).

Lorsque des données sur les orphelins ont été collectées dans deux enquêtes successives, il est possible de dégager des mesures relatives à la survie des parents entre les deux opérations de collectes, au sein d'une cohorte synthétique, et d'en déduire des paramètres de la table de mortalité pour cette période. En particulier, si les adultes âgés de 15 à 49 ans ont été interrogés sur la survie de leurs parents, on peut estimer les probabilités conditionnelles de survie à partir des données sur les orphelins à l'âge adulte, c'est-à-dire pour les cohortes synthétiques de 20 ans et plus. Les méthodes basées sur des cohortes synthétiques peuvent fournir des estimations de la mortalité adulte pour une période récente et relativement bien définie. Ceci est particulièrement utile dans les pays où l'épidémie du VIH est généralisée et où le niveau de la mortalité des adultes est susceptible d'avoir

changé brusquement au cours des deux dernières décennies. L'approche permet potentiellement de réduire également les biais résultant de la sous-déclaration du statut d'orphelin par ceux qui ont perdus leurs parents à un jeune âge.

Si, dans une seule enquête, une question supplémentaire concernant la date de décès des parents a été posée, on peut utiliser cette information pour reconstruire les proportions de répondants dont la mère ou le père était encore en vie quelques années plus tôt et estimer les probabilités de survie des parents dans la période du passé correspondante, de la même manière que pour les données de plusieurs opérations de collecte successives.

Les méthodes basées sur la survie des parents ont un avantage sur les questions directement posées à propos des décès survenus dans les ménages : la mortalité des adultes peut être estimée de cette manière dans les enquêtes de taille modérée. En revanche, seuls les recensements ou les enquêtes à très large échantillons peuvent donner des estimations directes fondées sur les décès au cours des douze mois précédant la collecte qui sont suffisamment précises pour être utiles. En outre, les méthodes basées sur la survie des parents ne reposent pas sur l'hypothèse selon laquelle la population est fermée à la migration. Cependant, les résultats que l'on obtient ne seront pas représentatifs au niveau des petits états ou des régions d'un pays dans lesquelles une proportion importante de la population aura émigré ou sera composée d'immigrés.

DONNÉES REQUISES

Pour estimer la mortalité des femmes adultes :

- Les proportions de répondants dont la mère est en vie, classées par groupe quinquennal d'âge du répondant, à

deux ou plusieurs dates différentes (ceux qui ne savent pas si leur mère est vivante ou qui n'ont pas répondu à la question devraient être exclus des calculs) ;

- Le nombre de naissances survenues l'année précédant le recensement ou l'enquête classées par groupe quinquennal d'âge de la mère ;
Pour estimer la mortalité des hommes adultes :
- Les proportions de répondants dont le père est en vie, classés par groupe quinquennal d'âge du répondant, à deux ou plusieurs dates différentes (ceux qui ne savent pas si leur père est vivant ou qui n'ont pas répondu à la question devraient être exclus des calculs) ;
- Le nombre de naissances survenues l'année précédant le recensement ou l'enquête classées par groupe quinquennal d'âge de la mère ;
- Une estimation de la différence d'âge entre les hommes et les femmes ayant des enfants, par exemple la différence entre les âges médians des hommes et des femmes actuellement mariés.

Ces tableaux doivent généralement être établis pour les répondants hommes et femmes et les estimations sont faites à partir de deux séries de proportions et pour les deux sexes ensemble.

La méthode des cohortes synthétiques décrite ici produit des estimations de la mortalité adulte à partir des données fournies par les orphelins à l'âge adulte, c'est-à-dire les personnes âgées de 15 ans ou plus. Bien qu'on n'ait pas besoin de données sur les groupes d'âge plus jeunes pour produire les estimations basées sur les cohortes synthétiques, si ces données sont collectées, elles devraient être introduites dans la feuille de calcul en vue de générer les estimations au moyen de la méthode de base des proportions d'orphelins.

Si les poids d'échantillon sont fournis avec les données, il convient de les utiliser de manière appropriée au logiciel statistique utilisé pour produire les tables qui servent d'inputs.

HYPOTHÈSES IMPORTANTES

Une limitation inhérente à la méthode des proportions d'orphelins est que les données relatives à la survie des parents ne peuvent être collectées qu'à partir de leurs enfants survivants au moment de l'enquête. La survie des adultes qui n'ont aucun enfant vivant n'est pas représentée dans les déclarations concernant la survie des parents. En outre, les parents ayant plus d'un enfant survivant sont surreprésentés par rapport à ceux qui ont exactement un enfant survivant, en proportion du nombre de leurs enfants survivants. La méthode produit

donc des résultats non biaisés uniquement si la mortalité des parents n'est pas corrélée au nombre de leurs enfants encore en vie au moment de l'enquête. En général, le biais de sélection qui découle des violations de cette hypothèse est faible (Palloni, Massagli et Marcotte 1984). Dans les populations touchées par les épidémies de VIH, ce biais est susceptible d'être plus sévère.

TRAVAUX PRÉPARATOIRES ET RECHERCHES PRÉLIMINAIRES

Avant de commencer l'analyse, il convient de vérifier le nombre de répondants qui ont déclaré qu'ils ne savaient pas si leur mère ou – plus fréquemment – le père était en vie ou qui ont omis de répondre à la question. Le taux de réponse sur ces questions est généralement très élevé et on peut simplement exclure de l'analyse les répondants qui ont répondu « ne sait pas » ou n'ont pas répondu à la question. Ceci revient à supposer que la proportion des parents décédés de ces répondants est la même que pour les parents de ceux qui ont répondu à la question. Cependant, quelques enquêtes ont collecté suffisamment de données incomplètes pour suggérer que le biais de non-réponse pourrait être un problème sérieux. Par exemple, il est possible que la plupart des gens qui omettent de répondre à la question aient des parents décédés. Si tel est le cas, les orphelins non déclarés pourraient représenter une proportion importante de tous les orphelins, en particulier dans les groupes d'âge plus jeunes ; ce qui conduirait à une sous-estimation importante dans les estimations finales de la mortalité adulte.

Un moyen utile de vérifier la qualité des données concernant les orphelins consiste à comparer les réponses de répondants de sexe masculin avec celles des répondantes de sexe féminin de même âge. On ne s'attend pas que la proportion des parents décédés diffère significativement entre les femmes et les hommes de même âge. Si les proportions divergent parmi les répondants plus âgés, cela pourrait refléter les différences entre hommes et femmes dans les erreurs de déclaration des âges ou indiquer que le genre qui déclare moins le décès des parents (généralement les hommes) est plus susceptible de perdre contact avec leurs familles et de supposer à tort que certains parents décédés sont encore en vie.

Lorsque deux ou plusieurs ensembles de données sur la survie des parents sont disponibles, il convient généralement d'estimer indépendamment la mortalité à partir de chaque série de données, à l'aide de la méthode de base des proportions d'orphelins, et de produire également des estimations à

partir des données sur les orphelins à l'âge adulte organisées dans des cohortes synthétiques, afin de comparer les trois ensembles de résultats. La feuille de calcul Excel associée à ce chapitre produit les deux types d'estimations, la méthode de base des proportions d'orphelins et la méthode basée sur les cohortes synthétiques.

PRÉCAUTIONS ET MISES EN GARDE

- Les estimations tirées de la méthode des proportions d'orphelins sont des probabilités conditionnelles de survie, c'est-à-dire les probabilités de décéder entre deux âges adultes à condition d'être en vie au début de l'intervalle. Pour obtenir une table de mortalité complète, les estimations de la probabilité de survie de la naissance à l'âge adulte doivent être calculées en utilisant une autre source de données sur la mortalité des enfants.
- Dans diverses applications en Afrique de l'Est et ailleurs, la méthode des proportions d'orphelins a donné des résultats qui indiquent des baisses invraisemblablement rapides de la mortalité et des incohérences flagrantes entre les estimations tirées d'enquêtes successives. Cela semble être dû à « l'effet d'adoption », qui consiste en une sous-déclaration du décès des parents lorsque ce décès est survenu quand ils étaient très jeunes (Blacker 1984 ; Blacker and Gapere 1988 ; Hill 1984 ; Timæus 1986). Dans ces cas, les enfants sont souvent élevés par d'autres parents et sont souvent énumérés comme leurs propres enfants. Cela signifie qu'ils sont énumérés comme ayant un parent en vie ; ce qui peut entraîner des estimations très basses de la mortalité. Au fur et à mesure que les répondants vieillissent, la probabilité s'accroît que, en plus du parent biologique, leur parent d'accueil, le parent adoptif ou beau-parent soit également décédé. Cela implique que le biais est plus prononcé pour les jeunes enfants, dont le parent adoptif est probablement vivant. Les procédures permettant d'estimer la mortalité adulte à partir des données de cohortes synthétiques sur les orphelins à l'âge adulte ont été développées spécifiquement pour résoudre ce problème. Toutefois, ces procédures ne peuvent pas complètement éliminer les biais sur les résultats dans les populations où le nombre d'enfants orphelins est sensiblement sous-déclaré.
- Bien qu'il soit possible de dégager des estimations de mortalité en utilisant des données sur les répondants d'une quarantaine d'années, les parents de plusieurs de ces répondants sont très âgés et ont une mortalité très élevée. Cela signifie que la précision avec laquelle on peut estimer

le niveau de mortalité à partir des données sur la survie des parents est par nature beaucoup plus faible qu'elle ne l'est pour les répondants plus jeunes.

- Comme toutes les méthodes qui impliquent l'analyse du changement entre deux enquêtes indépendantes, l'approche basée sur des cohortes synthétiques d'orphelins est vulnérable aux biais résultant de différences dans la qualité des données entre les deux enquêtes. Si les répondants étaient davantage susceptibles de déclarer des parents décédés comme étant vivants dans une des enquêtes, le biais qui résulte de ces erreurs de déclarations sera amplifié dans les estimations obtenues avec la méthode de la cohorte synthétique. La mortalité sera surestimée si trop peu d'orphelins ont été déclarés dans l'enquête réalisée plus tôt et sous-estimée si trop peu d'orphelins ont été déclarés dans l'enquête réalisée plus tard. En outre, les estimations fondées sur la variation de la proportion de parents survivants entre deux enquêtes sont associées à d'erreurs d'échantillonnage plus larges que les deux séries de proportions à partir desquelles elles sont calculées.

APPLICATION DE LA MÉTHODE

Étape 1 : Calculer l'âge moyen des mères et pères à la naissance de leurs enfants

Pour appliquer la méthode des proportions d'orphelins, on doit disposer d'une estimation de l'âge moyen auquel les parents ont eu leurs enfants afin de contrôler les variations dans l'éventail des âges auxquels ils ont été exposés au risque de décès. L'âge moyen des femmes à la naissance de leurs enfants est généralement calculé à partir des données sur les naissances de l'année précédant l'enquête ou le recensement classifiées par groupe quinquennal d'âge des mères. La mesure est simplement l'âge moyen des femmes ayant eu une naissance, selon la formule suivante, qui ne tient pas compte de la structure par âge de la population :

$$\bar{M}^f = \frac{\sum_{x=15}^{45} {}_5B_x(x+2)}{\sum_{x=15}^{45} {}_5B_x}$$

Dans cette équation, ${}_5B_x$ représente les naissances des femmes du groupe d'âge entre x et $x+4$ ans révolus, et $(x+2)$ représente le point médian du groupe d'âge des femmes décalé vers le bas d'une demi-année, pour tenir compte du fait

que les femmes ayant une naissance au cours de l'année précédant l'interview l'ont eu en moyenne 6 mois plus tôt et qu'elles étaient donc alors 6 mois plus jeunes. Ce calcul peut être fait dans le fichier Excel joint. Si les données utilisées pour calculer \bar{M}^f sont compilées par âge des femmes à l'accouchement, le point médian de chaque groupe d'âge deviendrait $x + 2,5$.

Il n'est pas nécessaire d'ajuster les données sur les naissances \bar{M}^f pour tenir compte des erreurs sur la période de référence, d'autant que les estimations de la mortalité ne sont pas très sensibles à ces erreurs. Cependant, si on a des preuves que la structure par âge de naissances a été sévèrement déformée par une mauvaise déclaration des âges, notamment par les femmes qui ont exagéré leur âge, le nombre de naissances selon l'âge pourrait être recalculé à partir d'une structure par âge de la fécondité ajustée avant le calcul de \bar{M}^f .

En principe, l'âge moyen à la maternité devrait se référer à la date à laquelle les répondantes sont nées, qui peut correspondre à n'importe quel moment entre 5 et 45 ans avant la collecte des données sur la survie des parents. Une estimation basée sur les données de fécondité collectées dans la première des deux enquêtes qui ont posé des questions sur la survie des parents devrait être adéquate dans des populations qui, en ce moment-là, n'ont pas encore connu une baisse significative de la fécondité. Si l'on pense que le niveau de la fécondité a baissé, et si des données de recensements ou d'enquêtes antérieures existent, des valeurs de \bar{M}^f pourraient également être calculées à partir de ces données antérieures pour déterminer si cet indicateur a changé. Si c'est le cas, alors, la meilleure façon de déterminer quelles valeurs finales de \bar{M}^f retenir pour l'estimation de la mortalité adulte dépendra des données disponibles et du type de changement de la fécondité. Une autre option pourrait consister à calculer \bar{M}^f à partir de données collectées au moment où la fécondité a commencé à baisser et à utiliser cette valeur pour les groupes de répondantes nées à cette période ou plus tôt, puis à utiliser une interpolation linéaire entre cette valeur et celle de l'enquête en cours pour estimer \bar{M}^f pour les répondantes des plus jeunes groupes d'âge.

L'âge moyen des pères à la naissance de leurs enfants est généralement estimé en ajoutant à l'âge moyen des mères à la naissance de leurs enfants une mesure de la différence d'âge entre les hommes et les femmes ayant des enfants :

$$\bar{M}^m = \bar{M}^f + d$$

Une estimation de cette différence peut être faite à partir des données de recensement, en calculant la différence entre l'âge médian des hommes actuellement mariés et l'âge médian des femmes actuellement mariées. C'est une solution plus appropriée que la différence entre les âges moyens au premier mariage des hommes et des femmes, calculée dans les populations où les dissolutions d'unions et les mariages polygames sont courants. La médiane est utilisée plutôt que la moyenne de sorte que la surestimation différentielle des âges par les répondants âgés, dont la fécondité est vraisemblablement nulle de toute façon, n'affecte pas l'estimation.

Cette méthode d'estimation de l'âge moyen des hommes à la naissance de leurs enfants suppose que les âges des pères d'enfants nés de femmes non mariées sont égaux, en moyenne, aux âges des pères d'enfants nés de mères mariées. Si cette hypothèse n'est pas vérifiée, cette méthode pourrait introduire un biais significatif dans l'estimation de \bar{M}^m dans les populations où la procréation en dehors du mariage est fréquente. Ce problème est difficile à résoudre mais heureusement, les estimations de la mortalité ne sont pas très sensibles aux erreurs sur l'estimation de \bar{M}^m .

Étape 2 : Calculer les mesures relatives à la survie des parents à l'âge adulte pour les cohortes synthétiques

Le classeur Excel contient des feuilles séparées pour le calcul de ces proportions pour les femmes et les hommes adultes. Les valeurs à introduire dans la feuille **Orphelins de mères** sont soit le nombre de répondants par groupe quinquennal d'âge dont la mère est en vie et le nombre d'enquêtés ayant répondu à la question, soit directement les proportions des répondants dont la mère est en vie, calculée à partir de ces nombres. De la même façon, les valeurs à introduire dans la feuille **Orphelins de pères** sont soit le nombre de répondants par groupe quinquennal d'âge dont le père est en vie et le nombre ayant répondu à la question, soit directement les proportions des répondants dont le père est en vie, calculées à partir de ces nombres. Pour la survie des mères comme pour celle des pères, l'ensemble des résultats les plus récents doit être saisi dans le panneau du haut de la feuille de calcul et l'ensemble le plus ancien dans le panneau repris plus bas. Bien que les données puissent concerner les répondantes du sexe féminin, les répondants du sexe masculin ou les répondants des deux sexes, la tabulation doit se faire de la même façon.

Le tableur calcule les paramètres relatifs à la survie des parents dans une cohorte synthétique au moyen de la méthode de « variable r ». Les proportions moyennes des répondants dont les parents sont encore en vie dans la période entre les deux opérations de collecte sont multipliées par l'exponentielle des taux de croissance de ces proportions au cours de la période, qui sont cumulés à partir de l'âge de 20 ans. Ceci « supprime » l'effet de la croissance démographique, produisant ainsi, des proportions qui seraient observées dans une population stationnaire et qui sont relatives à la proportion des répondants avec des parents survivants à partir de l'âge de 20 ans. Ces proportions stationnaires reflètent le taux (ou le rythme) auquel des adultes perdent leurs parents au cours de la période entre les deux opérations de collecte.

La proportion moyenne des répondants d'un groupe d'âge dont la mère (ou le père) est encore en vie au cours de la période entre les deux opérations de collecte est calculée comme ceci :

$${}_5S_x(\bar{t}) = \sqrt{{}_5S_x(t) \times {}_5S_x(t+h)}$$

Où t se rapporte à la première opération de collecte, $t+h$ à la deuxième opération de collecte survenant h années après la première, et une mesure de l'intervalle de temps entre les deux opérations de collecte. Après avoir calculé ces mesures, la proportion moyenne des parents de personnes âgées de 20 ans exacts qui sont vivants au cours de la période peut être estimée comme :

$$S(20, \bar{t}) = \sqrt{{}_5S_{15}(\bar{t}) \times {}_5S_{20}(\bar{t})}$$

Les taux de croissance, entre la première et la deuxième opération de collecte, des proportions de parents encore en vie par groupe d'âge de répondants, sont calculés comme :

$${}_5r_x(\bar{t}) = \frac{\ln({}_5S_x(t+h)) - \ln({}_5S_x(t))}{h}$$

Ensuite, les proportions qui correspondent à une cohorte synthétique d'individus dont les parents vivent toujours, parmi ceux dont les parents étaient vivants quand ces individus avaient 20 ans peuvent être calculées comme :

$$\frac{{}_5S_x(\tau)}{S(20, \bar{t})} = \frac{{}_5S_x(\bar{t})}{S(20, \bar{t})} \cdot e^{\left(\sum_{x=5}^{x=20,5} {}_5r_x(\bar{t}) \right) + 2,5, r_x(\bar{t})}$$

où τ se réfère à la mesure ajustée (c'est-à-dire, la période) du temps \bar{t} pour la cohorte synthétique.

Étape 3a : Calculer les probabilités de survie

conditionnelle de la table de mortalité pour les femmes

La survie des femmes est estimée entre l'âge inférieur de 45 ans et un âge calculé comme $25 + n$ ans, où n est la limite supérieure de chaque groupe d'âges des répondants. L'équation de régression suivante et les coefficients du tableau 26.1 sont utilisés :

$${}_{n-20}p_{45} = a(n) + b(n)\bar{M}^f + c(n){}_5S_{n-5}(\tau) / S(20, \bar{t})$$

Par exemple, lorsque n est égal à 30 ans, les probabilités de survie de la table de mortalité sont estimées dans les intervalles de 10 ans à partir de 45 ans jusqu'à l'âge exact de 55 ans, en utilisant les données sur la survie des mères fournies par les répondants du groupe d'âge de 25 à 29 ans.

TABLEAU 26.1 Coefficients pour estimer les probabilités de survie des femmes à partir des proportions de répondants dont la mère est encore vivante, parmi ceux dont la mère était toujours vivante quand ils avaient 20 ans

| n | $a(n)$ | $b(n)$ | $c(n)$ |
|-----|---------|---------|--------|
| 25 | -0,8623 | 0,00292 | 1,7861 |
| 30 | -0,3822 | 0,00679 | 1,2062 |
| 35 | -0,4355 | 0,01197 | 1,1310 |
| 40 | -0,5995 | 0,01847 | 1,1419 |
| 45 | -0,7984 | 0,02547 | 1,1866 |
| 50 | -0,9360 | 0,03039 | 1,2226 |

Source : Timæus (1991)

Étape 3b : Calculer les probabilités conditionnelles de survie de la table de mortalité pour les hommes

Chaque estimation de la survie des hommes est calculée à partir des données relatives à deux groupes d'âge adjacents de cinq ans, et non à partir d'un seul groupe d'âge. La survie des hommes est calculée entre une borne inférieure de 55 ans et une borne supérieure calculée comme $35 + n$, où n est le milieu de deux groupes d'âge adjacents. On utilise l'équation de régression suivante et les coefficients indiqués dans le tableau 26.2 ci-dessous :

$${}_{n-20}p_{55} = a(n) + b(n)\bar{M}^m + c(n){}_5S_{n-5}(\tau) / S(20, \bar{t}) + d(n){}_5S_n(\tau) / S(20, \bar{t})$$

Par exemple, lorsque n vaut 40 ans, la probabilité de survie est estimée sur un intervalle de 20 ans, à partir de 55 ans jusqu'à l'âge exact de 75 ans, en utilisant les données sur la

survie des pères fournies par les répondants dans les deux groupes d'âge adjacents de 35-39 ans et 40-44 ans.

TABLEAU 26.2 Coefficients pour estimer les probabilités de survie des hommes à partir des proportions de répondants dont le père est encore vivant, parmi ceux dont le père était toujours vivant quand ils avaient 20 ans

| <i>n</i> | <i>a(n)</i> | <i>b(n)</i> | <i>c(n)</i> | <i>d(n)</i> |
|----------|-------------|-------------|-------------|-------------|
| 25 | − 0,0554 | 0,00757 | 0,0239 | 0,8080 |
| 30 | − 0,7539 | 0,01558 | 0,6452 | 0,6498 |
| 35 | − 1,0809 | 0,02273 | 0,9289 | 0,4807 |
| 40 | − 1,1726 | 0,02647 | 0,9381 | 0,4372 |

Source : Timæus (1991)

Étape 4 : Convertir les probabilités de survie en estimations du niveau général de la mortalité

La série des probabilités de survie conditionnelle, notées ${}_np_b$, et obtenues à partir de différents groupes d'âge de répondants, se réfère à l'intervalle entre les deux enquêtes. Ces probabilités représentent chacune une table de mortalité incomplète avec une base à 45 ans pour les femmes et 55 ans pour les hommes. La série sera dans une certaine mesure erratique à la suite des erreurs de déclarations d'âge, des erreurs d'échantillonnage et d'autres erreurs. Elle peut être lissée en ajustant à partir des probabilités un modèle relationnel logit à 2 paramètres associé à des tables de mortalité. Les logits des probabilités de survie conditionnelle sont calculés de la façon suivante :

$$Y_x = \frac{1}{2} \ln \left(\frac{1 - {}_np_b}{{}_np_b} \right)$$

Les logits équivalents de la table de mortalité utilisée comme standard sont exprimés comme ceci :

$$Y_x^s = \frac{1}{2} \ln \left(\frac{l_b - l_{b+n}}{l_{b+n}} \right)$$

Les paramètres α et β du modèle ajusté correspondent à l'ordonnée à l'origine et à la pente de la droite de régression des valeurs de Y_x sur Y_x^s . En principe, les estimations pour les groupes d'âges plus élevés sont moins vulnérables à aux erreurs d'échantillonnage que ceux des groupes d'âges plus jeunes, car elles sont basées sur un nombre plus important de décès de leurs parents. Cependant, ces estimations peuvent indiquer une mortalité plus faible que les estimations faites pour les groupes d'âges plus jeunes ; ceci est probablement

dû au fait que les répondantes plus âgées exagèrent souvent leurs âges. Ainsi, on devrait, à chaque extrémité de la série, exclure les estimations qui sont sans commune mesure avec les autres estimations dans les groupes d'âge utilisés pour estimer α et β .

Après avoir calculé α et β , on peut calculer les estimations lissées des probabilités de survie conditionnelle :

$${}_n\hat{p}_b = \frac{1}{1 + \exp \left(2 \left(\alpha + \beta Y_x^s \right) \right)}$$

Les estimations ajustées des probabilités de survie conditionnelle se réfèrent à une période de temps clairement définie et dépendent peu des hypothèses relatives à la structure par âge de la mortalité sur lesquelles reposent le processus d'estimation. Elles ne seront pas fortement déformées dans les populations avec des structures par âge de la mortalité inhabituelles telles que celles des populations exposées à une épidémie généralisée de VIH. Dès lors, ces probabilités de survie doivent, si possible, être utilisées telles quelles pour les analyses ultérieures. Néanmoins, il est souvent nécessaire de convertir ces estimations relatives aux cohortes synthétiques en un indicateur habituel de mortalité afin de pouvoir mener des comparaisons entre la mortalité des hommes et celle des femmes ou afin de comparer les deux séries d'estimations à d'autres estimations du niveau de la mortalité obtenues à partir de sources alternatives. Cela peut être fait en ajustant une table type de mortalité à 1-paramètre à partir de chaque probabilité de survie conditionnelle pour obtenir l'indicateur désiré à partir du modèle ajusté.

Plusieurs types d'indicateurs du niveau de mortalité ont été utilisés à cette fin, y compris les paramètres de niveau de mortalité de différents systèmes de tables-types de mortalité, des probabilités de survie, des espérances de vie à divers âges entre 5 et 30 ans, et même l'espérance de vie partielle entre 25 et 70 ans, notée ${}_{45}e_{25}$. L'utilisation des paramètres des tables-types a l'avantage de souligner que toute la table de mortalité est estimée par l'ajustement d'un modèle, et non directement mesurée à partir des données observées. Les mesures de l'espérance de vie résument le niveau de mortalité adulte dans son ensemble, alors qu'utiliser les probabilités de survie ou l'espérance de vie entre deux âges adultes permet d'éviter l'extrapolation aux âges élevés des niveaux de mortalité estimés chez les jeunes adultes. De plus en plus, au cours de ces dernières années, les estimations ont été présentées sous forme de probabilités de survie entre les âges exacts 15 et 60 ans, soit par la probabilité ${}_{45}q_{15}$, puisque

cette mesure est préférée par plusieurs organismes internationaux comme indicateur de la mortalité des jeunes adultes et des adultes d'âge moyen.

Dans les applications de la méthode des orphelins présentée ici, les probabilités de survie sont converties en paramètre α d'un système relationnel logit à un paramètre associé à des tables de mortalité, puis en estimations de la probabilité conditionnelle de décéder dans un éventail plus large d'âges adultes. (Il convient de noter que, même si le même standard est utilisé et si la valeur de β est égale à 1, le paramètre α d'un modèle ajusté à partir de la naissance, ne sera pas le même que la valeur de α qui serait calculée dans les modèles qui auraient été ajustés à des mesures de la probabilité conditionnelle de survie à partir des âges 45 ou 55 ans.)

Le tableur Excel calcule la probabilité conditionnelle de survie entre les âges exacts de 30 et 60 ans ($_{30}q_{30}$), 15 et 60 ans ($_{45}q_{15}$), et 50 et 70 ans ($_{20}q_{50}$). Les deux premiers indicateurs du niveau de mortalité adulte sont utiles pour comparer les estimations faites à partir des cohortes synthétiques avec celles obtenues par la méthode de base des proportions d'orphelins (chapitre 22) et d'autres mesures de la mortalité des adultes ; le troisième est le plus utile pour comparer les estimations pour les femmes et les hommes, faites à partir des données sur les orphelins à l'âge adulte ou pour évaluer la cohérence interne d'une série de ces estimations sans extrapolation à partir de la survie des âges adultes moyens aux plus jeunes âges adultes. Le paramètre α des modèles à un paramètre sont calculés à partir des estimations de $_{n-20}p_b$ selon la formule suivante :

$$\alpha = -\frac{1}{2} \ln \left(1 + \frac{\frac{_{n-20}p_b}{l_{b+n-20}^s} - \frac{1}{l_b^s}}{1 - _{n-20}p_b} \right)$$

Où les estimations de $_{n-20}p_b$ sont obtenus à l'étape 2, avec $b = 45$ pour les estimations de la survie des femmes et $b = 55$ pour les estimations de la survie des hommes, et des valeurs de l_x^s qui proviennent d'une table-type de mortalité. Ainsi, on obtient une série d'estimations de α correspondant aux mesures de probabilités conditionnelles de survie calculées à partir des différents groupes d'âge. Des valeurs plus élevées de α correspondant à une mortalité plus élevée. Ensuite, pour chaque α , les indicateurs tels que $_{20}q_{50}$, $_{30}q_{30}$ et $_{45}q_{15}$ peuvent être calculés avec l'équation suivante :

$$_nq_b = 1 - \frac{1 + e^{2(\alpha + Y_i(b))}}{1 + e^{2(\alpha + Y_i(b+n))}}$$

Le tableur peut calculer ces indicateurs en utilisant un standard issu soit des tables-types des Nations Unies, soit de l'une des quatre familles de tables-types de Princeton. La standard choisit doit avoir une structure par âge de la mortalité adulte qui ressemble à celle de la population étudiée. Une autre table de mortalité peut être utilisée comme standard si on a des éléments pour affirmer qu'elle ressemble de plus près à la structure de la mortalité des adultes dans la population étudiée. La table de mortalité la plus appropriée peut ne pas être de la famille de modèles qui traduit le mieux la relation entre la mortalité des enfants et la mortalité des adultes. Si on ne sait rien de la structure par âge de la mortalité à l'âge adulte, l'utilisation du modèle général des Nations Unies ou du modèle Ouest des tables de Princeton est recommandée.

Comme les estimations se rapportent toutes à la même période, il est logique pour produire une estimation finale de la survie pour la période entre les deux enquêtes en calculant la moyenne d'un ensemble contigu d'estimations qui exclut les valeurs aberrantes obtenues à partir des répondants les plus jeunes et les plus âgés. Ces valeurs aberrantes peuvent être identifiées à l'aide d'une représentation graphique des logits des probabilités de survie conditionnelle présentés en vis-à-vis des logits d'une table standard. Si on observe une tendance à la hausse ou à la baisse des α de tous les groupes d'âge à partir des valeurs ajustées sur un modèle à 1-paramètre, la table standard de mortalité à laquelle les évaluations sont appliquées peut être inappropriée. L'analyste devrait probablement adopter soit une autre table standard de mortalité, soit modifier la vitesse à laquelle la mortalité augmente avec l'âge dans le modèle sélectionné en modifiant le paramètre β .

Étape 5 : Calculer la période de référence à laquelle les estimations se rapportent

Chaque probabilité de survie se réfère à la période comprise entre les deux opérations de collecte. On pourrait vouloir leur attribuer une date exacte au sein de cette période, de façon à pouvoir les représenter graphiquement afin de les comparer aux autres estimations de la mortalité adulte. Si l'on suppose un rythme constant d'évolution de la mortalité au cours de la période qui sépare les deux opérations de collecte, on peut considérer que ces probabilités se réfèrent à la moyenne géométrique des dates des deux collectes.

La date de chaque opération de collecte peut être calculée comme la moyenne des dates auxquelles les entretiens ont eu lieu ou le point milieu de la période de travail sur le terrain si les dates exactes des entretiens ne sont pas disponibles.

EXEMPLE

Cet exemple, présenté dans une feuille de calcul disponible sur le site web associé à cet ouvrage, utilise des données sur la survie des mères et des pères, collectées dans les recensements généraux de la population de 1989 et 1999 au Kenya.

Étape 1 : Calculer l'âge moyen des parents à la naissance de leurs enfants

Pour les femmes, l'âge moyen à la naissance des enfants est une moyenne simple de l'âge des femmes ayant donné naissance. Cet âge peut être calculé directement à partir de données individuelles, ou être estimé à partir d'une tabulation de naissances par groupe quinquennal d'âge des mères. Pour cette application, il a été calculé en utilisant les données du recensement général de la population de 1989 (voir tableau 26.3). Au Kenya, on pourrait aussi utiliser les données des recensements précédents de la population afin de vérifier si l'âge à la naissance des enfants a changé au fil du temps :

$$\bar{M}^f = \frac{16585300}{620000} = 26,75.$$

TABLEAU 26.3 Calcul de l'âge moyen des mères à la naissance de leurs enfants, Kenya, 1989

| Groupe d'âge | Naissances de l'année passée $B(i)$ | Milieu du groupe d'âges N | $B(i)*N$ |
|--------------|-------------------------------------|-----------------------------|------------|
| 15-19 | 73 600 | 17 | 1 251 200 |
| 20-24 | 193 400 | 22 | 4 254 800 |
| 25-29 | 170 220 | 27 | 4 595 940 |
| 30-34 | 95 180 | 32 | 3 045 760 |
| 35-39 | 56 340 | 37 | 2 084 580 |
| 40-44 | 23 240 | 42 | 976 080 |
| 45-49 | 8 020 | 47 | 376 940 |
| TOTAUX | 620 000 | | 16 585 300 |

L'âge moyen des hommes à la naissance de leurs enfants est calculé en ajoutant à l'âge moyen des femmes à la naissance de leurs enfants la différence entre les âges médians des

hommes et des femmes actuellement mariées. On peut voir dans le tableau 26.4, que l'âge médian des hommes actuellement mariés se situe entre le point milieu de la tranche d'âge 30-34 et le point milieu de la tranche d'âge 35-39 ans. Par interpolation linéaire entre ces deux points on a :

Âge médian pour les hommes =

$$32,5 + 5 \times \frac{0,5 - 0,3886}{0,5272 - 0,3886} = 36,52$$

Et,

Âge médian pour les femmes =

$$27,5 + 5 \times \frac{0,5 - 0,4234}{0,5597 - 0,4234} = 30,31$$

TABLEAU 26.4 Âge des femmes et des hommes actuellement mariés, Kenya, 1989

| Groupe d'âge | Hommes mariés | Femmes mariées | Proportion cumulée des hommes | Proportion cumulée des femmes |
|--------------|---------------|----------------|-------------------------------|-------------------------------|
| 10-14 | 2 800 | 6 680 | 0,0010 | 0,0019 |
| 15-19 | 18 040 | 212 060 | 0,0071 | 0,0612 |
| 20-24 | 173 840 | 623 040 | 0,0664 | 0,2356 |
| 25-29 | 464 720 | 670 760 | 0,2250 | 0,4234 |
| 30-34 | 479 460 | 487 180 | 0,3886 | 0,5597 |
| 35-39 | 406 000 | 387 000 | 0,5272 | 0,6681 |
| 40-44 | 330 140 | 305 500 | 0,6398 | 0,7536 |
| 45-49 | 250 540 | 243 120 | 0,7253 | 0,8216 |
| 50-54 | 212 820 | 189 240 | 0,7979 | 0,8746 |
| 55-59 | 161 760 | 137 120 | 0,8531 | 0,9130 |
| 60-64 | 135 060 | 113 860 | 0,8992 | 0,9449 |
| 65-69 | 101 860 | 75 540 | 0,9340 | 0,9660 |
| 70-74 | 72 080 | 49 980 | 0,9586 | 0,9800 |
| 75-79 | 56 240 | 30 100 | 0,9778 | 0,9884 |
| 80+ | 65 120 | 41 380 | 1,0000 | 1,0000 |
| TOTAUX | 2 930 480 | 3 572 560 | | |

L'âge moyen des hommes à la naissance de leurs enfants est donc :

$$M^m = M^f + d = 26,75 + (36,52 - 30,31) = 32,96$$

Étape 2 : Calculer les mesures relatives à la survie des parents à l'âge adulte pour les cohortes synthétiques

Les proportions des Kenyans dont la mère était encore en vie dans l'intervalle intercensitaire des recensements de 1989

et de 1999 sont présentées dans la quatrième colonne du tableau 26.5. Ces proportions sont les moyennes géométriques des proportions indiquées dans les deux recensements qui sont présentées dans la deuxième et la troisième colonne du tableau 26.5. Par exemple, dans le groupe d'âge de 25 à 29 ans, on aura :

$${}_5S_{25}(\bar{t}) = \sqrt{0,8839 \times 0,8771} = 0,8805.$$

La proportion des répondants de 20 ans exacts dont la mère est toujours en vie est calculée à partir de ces estimations pour les groupes d'âge de 15-19 ans et 20-24 ans :

$$S(20,\bar{t}) = \sqrt{0,9446 \times 0,9156} = 0,9300.$$

Le jour de référence du recensement en 1999 était le 24 août alors que pour le recensement de 1989 il s'agit du 25 octobre. Ainsi, le taux de croissance au cours de la décennie (1989-1999) de la proportion de mères survivantes pour le même groupe d'âge est :

$${}_5r_{25}(\bar{t}) = \frac{\ln(0,8771) - \ln(0,8839)}{9,83} = -0,00078.$$

Pour le premier groupe d'âge, le taux de croissance cumulé de 20 ans à 22,5 est égal à :

$$2,5{}_5r_{20}(\bar{t}) = -0,00170 \times 2,5 = -0,00425.$$

Pour le deuxième groupe d'âge, le taux de croissance cumulé de 20 à 27,5 ans est égal à :

$$\begin{aligned} 5 \sum_{a=20,5}^{20} {}_5r_a(\bar{t}) + 2,5{}_5r_{25}(\bar{t}) &= \\ 5 \times -0,00170 + 2,5 \times -0,00078 &= -0,01045 \end{aligned}$$

Tandis que pour le troisième groupe d'âge, il est égal à :

$$\begin{aligned} 5 \sum_{a=20,5}^{25} {}_5r_a(\bar{t}) + 2,5{}_5r_{30}(\bar{t}) &= \\ 5 \times (-0,00170 - 0,00078) + 2,5 \times 0,00018 &= -0,01195 \end{aligned}$$

et ainsi de suite.

Les proportions d'individus qui, dans une cohorte synthétique, ont une mère encore en vie parmi ceux qui avaient encore une mère vivante quand ils avaient à 20 ans sont présentées dans la septième colonne du tableau 26.5. Elles sont calculées à partir des proportions moyennes et des taux de croissance dans les quatrième et cinquième colonnes. Par exemple, pour le groupe d'âge de 25 à 29 ans :

$$\frac{{}_5S_x(\tau)}{S(20,\tau)} = \frac{0,8805}{0,9300} \cdot e^{-0,01045} = 0,9369.$$

TABLEAU 26.5 Estimation de la survie des femmes dans l'intervalle entre les deux recensements et estimations correspondantes de α et de ${}_3q_{30}$, à partir des données sur les orphelins de mères à l'âge adulte, Kenya, 1989-1999

| Groupe d'âge | Proportion de non-orphelins 1989 ${}_5S_{n-5}(t)$ | Proportion de non-orphelins 1999 ${}_5S_{n-5}(t+h)$ | Proportion de non-orphelins ${}_5S_{n-5}(\bar{t})$ | Taux d'accroissement | Âge n | Proportion de non-orphelins ${}_5S_{n-5}(\tau)$ | Valeur estimée de $\frac{l(25+n)}{l(45)}$ | Valeur ajustée de $\frac{l(25+n)}{l(45)}$ | Probabilité de décès (${}_3q_{30}$) |
|--------------|--|--|---|----------------------|---------|--|---|---|---------------------------------------|
| 15-19 | 0,9557 | 0,9336 | 0,9446 | | | | | | |
| 20-24 | 0,9233 | 0,9080 | 0,9156 | - 0,00170 | 25 | 0,9804 | 0,9669 | 0,9667 | 0,192 |
| 25-29 | 0,8839 | 0,8771 | 0,8805 | - 0,00078 | 30 | 0,9369 | 0,9295 | 0,9291 | 0,172 |
| 30-34 | 0,8229 | 0,8244 | 0,8236 | 0,00018 | 35 | 0,8751 | 0,8745 | 0,8804 | 0,167 |
| 35-39 | 0,7553 | 0,7691 | 0,7622 | 0,00184 | 40 | 0,8139 | 0,8240 | 0,8145 | 0,140 |
| 40-44 | 0,6258 | 0,6685 | 0,6468 | 0,00671 | 45 | 0,7057 | 0,7203 | 0,7244 | 0,140 |
| 45-49 | 0,5335 | 0,5653 | 0,5492 | 0,00589 | 50 | 0,6184 | 0,6329 | 0,6037 | 0,113 |

Les calculs effectués dans cette étape pour les données sur les orphelins de père sont identiques et sont présentés dans le tableau 26.6.

Étape 3a : Calculer des probabilités de survie conditionnelle de la table de mortalité pour les femmes

Ces probabilités de survie sont présentées dans la huitième colonne du tableau 26.5 et sont calculées à partir des proportions reprises dans la septième colonne en utilisant les coefficients de régression présentés dans le tableau 26.1 et l'estimation de l'âge moyen des mères à la naissance de leurs enfants \bar{M}^f (26,75 ans) obtenus à l'étape 1. Par exemple, pour les répondantes âgées de 25 à 29 ans, on obtient :

$${}_{10}p_{45} = -0,3822 + 0,00679 \times 26,75 + 1,2062 \times 0,9369 = 0,9295.$$

Il est utile de noter que chaque mesure de la table de mortalité est assez proche – en valeur – de la proportion à partir de laquelle elle a été calculée.

Étape 3b : Calculer les probabilités de survie conditionnelle de la table de mortalité pour les hommes

Ces probabilités de survie sont présentées dans la huitième colonne du tableau 26.6 et sont calculées à partir des proportions reprises dans la septième colonne en utilisant les coefficients de régression présentés dans le tableau 26.2 et l'estimation de l'âge moyen des pères à la naissance de leurs enfants \bar{M}^m (32,96 ans) obtenue à l'étape 1. Par exemple,

pour l'estimation finale présentés dans le tableau 26.6, le calcul est le suivant :

$${}_{20}p_{55} = -1,1726 + 0,02647 \times 32,96 + 0,9381 \times 0,6156 + 0,4372 \times 0,4772 = 0,4860.$$

Étape 4 : Convertir les probabilités de survie en estimations du niveau de mortalité

Pour lisser la série d'estimations de probabilités conditionnelles de survie en ajustant un système relationnel logit à 2 paramètres, on calcule d'abord les logits des probabilités. Par exemple, l'estimation de ${}_{10}p_{45}$ pour les femmes obtenue à partir de données sur les répondants âgés de 25-29 est converti de la sorte :

$$Y_x = \frac{1}{2} \ln \left(\frac{1 - 0,9295}{0,9295} \right) = -1,290.$$

La valeur équivalente pour la table-type de mortalité des Nations Unies (modèle Général) correspondant à une espérance de vie à la naissance de 60 ans est :

$$Y_x^s = \frac{1}{2} \ln \left(\frac{0,7887 - 0,7103}{0,7103} \right) = -1,102.$$

En régressant les logit des estimations observées de la survie conditionnelle sur les logits du standards (voir le site web), et en excluant le dernier point (basé sur les répondants âgés de 45-49 ans) qui se situe très bas en dessous de la ligne, on obtient des estimations suivantes :

TABLEAU 26.6 Estimation de la survie des hommes dans l'intervalle entre les deux recensements et estimations correspondantes de α et de ${}_{30}q_{30}$ à partir des données sur les orphelins de père à l'âge adulte, Kenya, 1989-1999

| Groupe d'âge | Proportion de non-orphelins 1989 ${}_5S_{n-5}(t)$ | Proportion de non-orphelins 1999 ${}_5S_{n-5}(t+h)$ | Proportion de non-orphelins ${}_5S_{n-5}(\bar{t})$ | Taux d'accroissement | Âge n | Proportion de non-orphelins ${}_5S_{n-5}(\tau)$ | Valeur estimée de $\frac{l(35+n)}{l(55)}$ | Valeur ajustée de $\frac{l(35+n)}{l(55)}$ | Probabilité de décès (${}_{30}q_{30}$) |
|--------------|--|--|---|----------------------|---------|--|---|---|--|
| 15-19 | 0,8670 | 0,8368 | 0,8518 | | | | | | |
| 20-24 | 0,7971 | 0,7730 | 0,7849 | - 0,00312 | 25 | 0,9525 | 0,9052 | 0,9045 | 0,259 |
| 25-29 | 0,7136 | 0,7055 | 0,7096 | - 0,00117 | 30 | 0,8519 | 0,7816 | 0,7849 | 0,257 |
| 30-34 | 0,6071 | 0,6074 | 0,6073 | 0,00004 | 35 | 0,7270 | 0,6395 | 0,6370 | 0,244 |
| 35-39 | 0,4972 | 0,5198 | 0,5084 | 0,00453 | 40 | 0,6156 | 0,4860 | 0,4678 | 0,225 |
| 40-44 | 0,3729 | 0,3953 | 0,3839 | 0,00592 | | 0,4772 | | | |

$\alpha = -0,3398$ et $\beta = 0,8597$. Le fait que l'estimation de β est inférieure à 1 indique que la mortalité des femmes au Kenya augmente moins rapidement que ce qui est prévu par le modèle de mortalité standard dans la tranche d'âge de 45 à 75 ans.

Ayant obtenu α et β , la probabilité conditionnelle de survie ajustée pour le second groupe d'âge, par exemple, peut être calculée comme suit :

$${}_{10}\hat{p}_{45} = \frac{1}{1 + \exp(2(-0,3398 + 0,8597 \times -1,1016))} \\ = 0,9291.$$

La série complète des probabilités conditionnelles de survie ajustées est présentée dans la neuvième colonne du tableau 26.5 pour les femmes et du tableau 26.6 pour les hommes. En considérant les estimations de ${}_5p_{45}$ et ${}_{25}p_{45}$ du tableau 26.5, la probabilité conditionnelle de décès entre les âges exacts de 50 et 70 ans dans le modèle ajusté à 2 paramètres est égale à : $1 - 0,7244 / 0,9667 = 0,251$.

Les estimations de ${}_{30}q_{30}$, la probabilité de décéder entre 30 et 60 ans exacts calculés par l'ajustement des probabilités conditionnelles de survie estimées au modèle à 1-paramètre, sont présentées dans les colonnes les plus à droite du tableau 26.5 et du tableau 26.6. Par exemple, la valeur de α calculée à partir de l'estimation de ${}_{10}p_{45}$ pour les femmes donne :

$$\alpha = -\frac{1}{2} \ln \left(1 + \frac{\frac{0,9295}{0,7103} - \frac{1}{0,7887}}{1 - 0,9295} \right) = -0,228.$$

En utilisant cette valeur de α , la mesure correspondante de ${}_{30}q_{30}$ est :

$${}_{30}q_{30} = 1 - \frac{1 + e^{2(-0,228-0,8655)}}{1 + e^{2(-0,228-0,3064)}} = 0,172.$$

Étape 5 : Calculer la période de référence à laquelle les estimations se rapportent

Les estimations de la cohorte synthétique sont des mesures de la mortalité des adultes au cours de la période entre les deux enquêtes. Leur date de référence peut être calculée en faisant la moyenne géométrique des dates de collecte de données sur le terrain des deux enquêtes. Ainsi, en l'appliquant aux recensements de 1989 et de 1999 au Kenya, on obtient :

$$T = \sqrt{1989,81 \times 1999,64} = 1994,72.$$

DIAGNOSTICS, ANALYSE ET INTERPRÉTATION

Contrôles et validation

Le nombre de répondants qui ont déclaré qu'ils ne savaient pas si leur mère ou leur père était en vie ou qui n'ont pas répondu aux questions doit être vérifié avant d'écarter ces répondants de l'analyse. Si ce nombre est important, les données fournies peuvent ne pas être représentatives de la population dans son ensemble. De plus, un niveau élevé de non-réponse pourrait indiquer que le personnel de terrain ou les enquêtes ont éprouvé des difficultés avec la question sur la survie des parents. Dans ce cas, il est possible que les réponses, même quand elles ont été fournies, ne soient pas fiables. S'il y a un niveau élevé de non-réponse sur une question, il serait utile de déterminer si ces non-réponses sont concentrées sur une minorité d'enquêteurs ou sur certaines catégories de répondants.

Si l'information sur la survie des mères et des pères a été obtenue à la fois de *répondants* hommes et femmes dans un recensement ou une grande enquête avec des marges d'erreur étroites, il est possible d'établir les proportions de pères et de mères encore en vie séparément pour les répondants de chaque sexe en vue de comparer la cohérence des déclarations. La cohérence des déclarations ne garantit pas leur exactitude, mais des différences statistiquement significatives entre les proportions calculées pour les répondants hommes et femmes impliquent qu'au moins un des deux sexes, et peut-être les deux, donnent des réponses inexactes.

Il est courant de constater que les femmes déclarent des proportions de parents vivants plus basses que les hommes. Certains analystes pensent que c'est parce que les femmes restent en contact plus étroit que les hommes avec leurs parents et que certains hommes disent que leurs parents sont en vie parce qu'ils ne savent pas qu'ils sont en fait décédés. Si cela est vrai, les données obtenues des femmes sont plus exactes que celles obtenues des hommes. Mais il n'y a pas de preuve convaincante qui confirme cette interprétation et d'autres facteurs, en particulier les différences entre hommes et femmes dans les erreurs sur la déclaration de leur âge, peuvent aussi être à l'origine des incohérences dans les proportions déclarées par les hommes et les femmes. De plus, dans les enquêtes ne portant que sur quelques milliers de ménages, les marges d'erreur des estimations pour les premiers groupes d'âge peuvent être larges par rapport aux proportions de mères ou de pères décédés. S'il n'y pas de raison particulière pour privilégier les réponses des femmes, nous

conseillons donc d'établir les estimations finales de la mortalité adulte sur les réponses combinées des deux sexes.

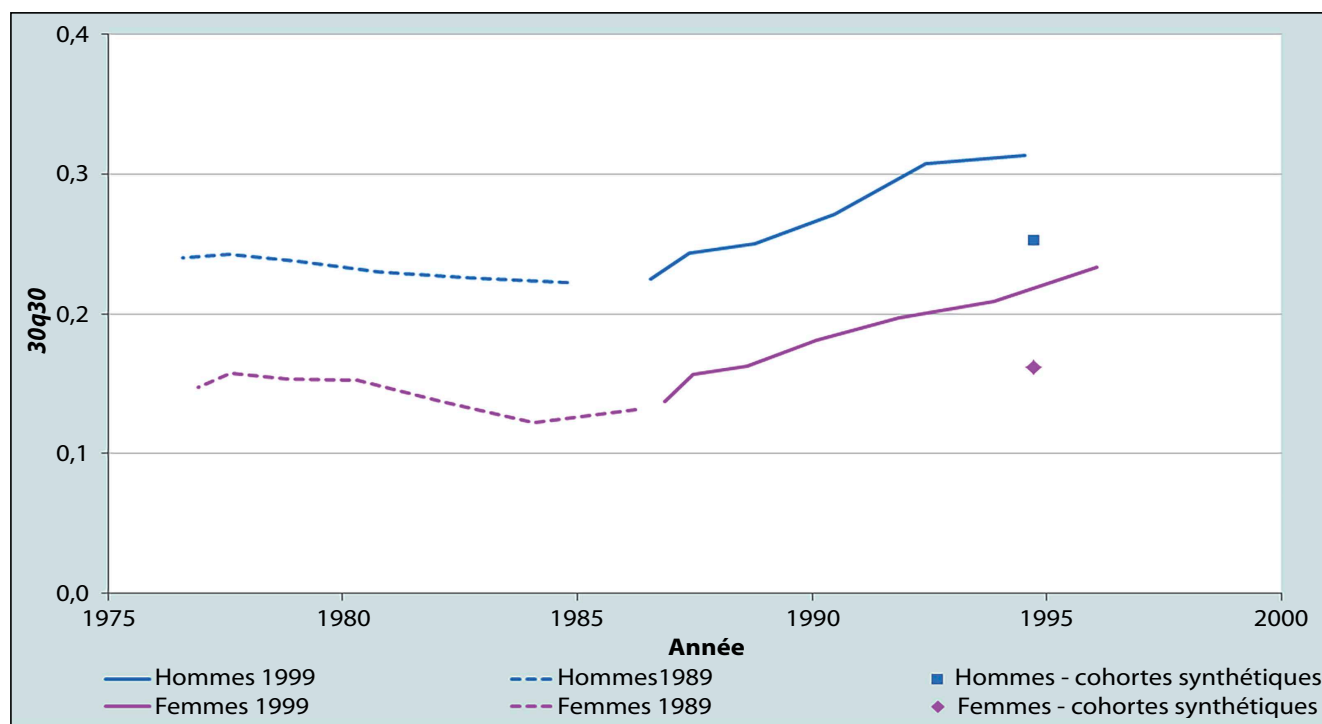
Interprétation

Les résultats de l'analyse de données sur les orphelins collectées lors des recensements du Kenya de 1989 et 1999 sont représentés dans la figure 26.1 ci-dessous. Selon les données du recensement de 1989, la mortalité des adultes au Kenya avait connu un lent déclin dans les années 1970 et au début des années 1980. Les probabilités de décès étaient assez faibles et un écart important existait entre la mortalité des hommes et des femmes. En revanche, les données du recensement de 1999 indiquent que la mortalité a augmenté de façon régulière pour les hommes et les femmes et a atteint un niveau relativement plus élevé entre la fin des années 1980 et le milieu des années 1990.

Une caractéristique rassurante de ces résultats est que les estimations de la mortalité pour 1985-1986 obtenues à partir des deux recensements sont cohérentes. Celles du

recensement 1989 (les points les plus récents présentés sur les lignes en pointillés) sont basées sur les déclarations relatives à la survie des parents des répondants qui sont encore des enfants. Les estimations obtenues pour une période légèrement plus récente à partir du recensement de 1989 (les premiers points sur les lignes solides) sont estimées à partir des déclarations qui concerne la survie des parents des répondants étaient trentenaires en 1999. Bien que cette cohérence entre les estimations faites à partir des probabilités de survie des parents des répondants de différents âges dans différentes enquêtes ne garantisse pas leur exactitude, les résultats obtenus suggèrent que l'on peut exclure certaines erreurs (mais pas toutes), y compris le biais résultant de l'effet d'adoption. Ce biais est plus sévère pour les estimations effectuées à partir des données sur la survie des parents des jeunes enfants. Plus l'enfant est jeune au moment du décès de ses parents, plus il est probable qu'à la question relative à la survie des parents, il ait été répondu en faisant référence à un parent adoptif ou à un autre parent qui l'a

FIGURE 26.1 Tendances de la probabilité conditionnelle de décéder entre 30 et 60 ans exacts, estimée à partir des données sur la survie des parents aux recensements du Kenya de 1989 et de 1999



élevé. Quand les répondants avancent en âge, les mauvaises déclarations deviennent proportionnellement moins importantes par rapport à l'augmentation rapide du nombre de décès des parents qui se produisent à la fois à mesure que les répondants et leurs parents vieillissent. Ainsi, si l'effet d'adoption était un problème au Kenya, on s'attendrait à ce que les estimations de la mortalité des adultes relatives au milieu des années 1980 soient plus faibles quand elles sont déduites du recensement de 1989 que lorsqu'elles sont dégagées des données du recensement de 1999.

Si les estimations de la mortalité obtenues auprès des plus jeunes répondants par la méthode basique des orphelins avaient été biaisées à la baisse, on s'attendrait à ce que les estimations tirées des proportions d'orphelins parmi les jeunes adultes organisées en cohortes synthétiques soient plus élevées que celles fournies par la méthode basique des orphelins pour les mêmes périodes. Cela n'est toutefois pas le cas au Kenya – elles sont plus faibles.

Ce schéma où les estimations déduites des jeunes adultes et obtenues de cohortes synthétiques sont plus faibles que celles dégagées des déclarations des jeunes enfants de façon classique est inhabituel. Cela reflète probablement l'importance croissante de la mortalité due au sida au Kenya au cours des années 1990. Les estimations des cohortes synthétiques sont principalement basées sur l'expérience des parents âgés de 50 ans et plus, qui n'ont pas été touchés fortement par l'épidémie du sida. Ainsi, en utilisant des tables-types de mortalité classiques pour déterminer ${}_{30}q_{30}$ à partir des différentes probabilités, on sous-estime la mortalité adulte, car, la mortalité dans les groupes d'âge de 30 à 50 ans est plus faible dans ces tables qu'au Kenya.

En revanche, les parents des jeunes répondants, dont on déduit les estimations les plus récentes par la méthode de base des proportions d'orphelins, ont pour la plupart entre 30 à 40 ans, ce qui correspond au groupe d'âge où la mortalité par sida connaît son niveau le plus élevé. Ainsi, utiliser des tables-types de mortalité classiques pour déterminer ${}_{30}q_{30}$ à partir de ces données produit une surestimation, car la mortalité des 50 à 60 ans supposée dans ces tables est plus élevée qu'au Kenya.

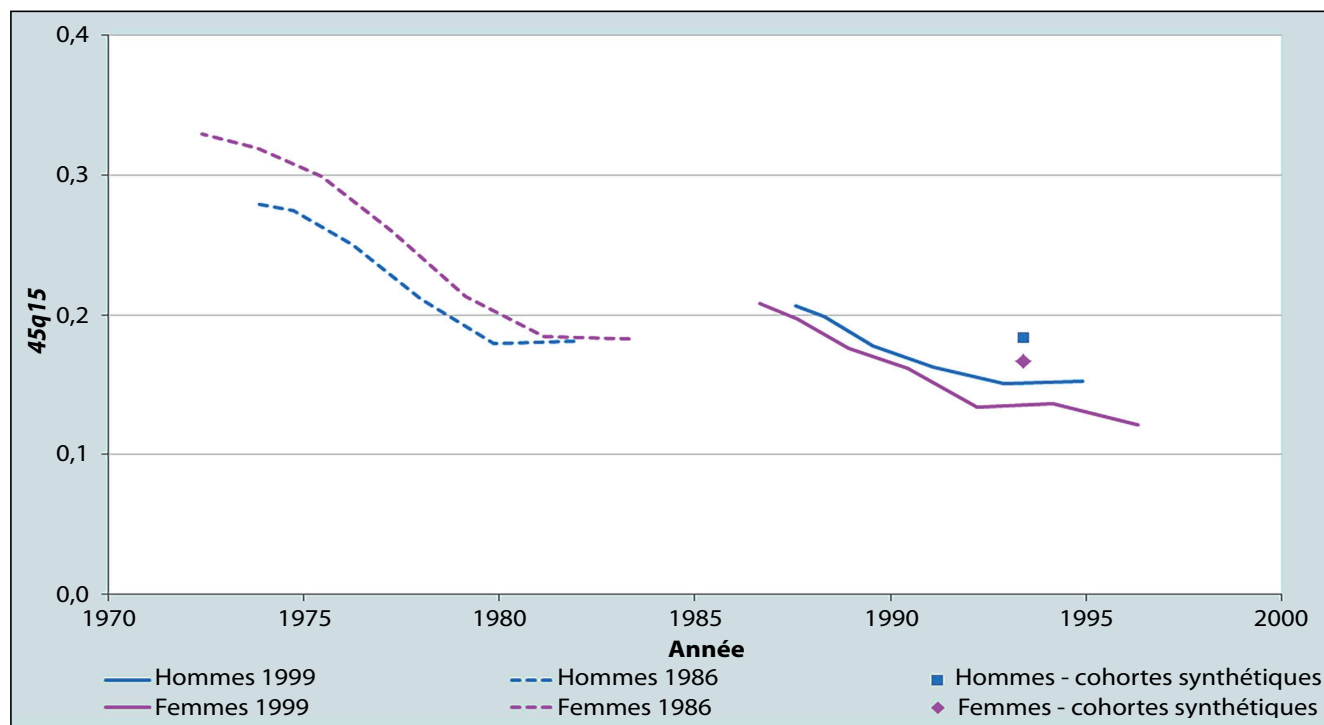
Les estimations des cohortes synthétiques appuient cette interprétation : dans cette application, la probabilité conditionnelle de décès baisse entre 30 et 60 ans quand l'âge des enquêtes augmente aussi bien pour les femmes que pour les hommes (voir le tableau 26.5 et le tableau 26.6). Ceci suggère que la mortalité est relativement élevée chez les plus

jeunes dans la tranche d'âge de 45 à 75 ans et relativement faible aux vieux âges au Kenya par rapport au modèle général des tables-types des Nations Unies. Ceci pourrait aussi indiquer une mortalité élevée liée au sida chez les plus jeunes adultes. La valeur réelle de la probabilité ${}_{30}q_{30}$ vers le milieu des années 1990 se situe probablement quelque part entre les estimations produites par les deux variantes de la méthode des orphelins. En utilisant un modèle standard de mortalité par âge qui augmente plus lentement avec l'âge, en fixant le paramètre β de la table type de mortalité à 0,7 et en recalculant α , on produit un ensemble d'indicateurs de mortalité adulte pour les deux sexes ayant une meilleure cohérence interne. Il permet également de réduire les incohérences entre les estimations de cohortes synthétiques et la plupart des estimations les plus récentes obtenues à partir des données du recensement de 1999 en utilisant la méthode de base des proportions d'orphelins. Ainsi, la probabilité de décéder entre 30 et 60 ans au Kenya dans les années 1990, toujours conditionnée au fait d'avoir survécu jusqu'à 30 ans, était probablement d'environ 20 pour cent pour les femmes et 30 pour cent pour les hommes.

La figure 26.2 présente une seconde application de l'approche reposant sur des cohortes synthétiques pour analyser deux ensembles de données sur les orphelins recueillies dans deux opérations de collecte successives. Il s'agit des données des recensements des Îles Salomon de 1986 et de 1999. Dans cette application, l'indice de mortalité présenté sur le graphique est la probabilité de décéder entre 15 et 60 ans, conditionné au fait d'avoir survécu jusqu'à 15 ans (${}_{45}q_{15}$). Un contraste évident apparaît immédiatement entre ces séries d'estimations et celles du Kenya ou celles des pays arabes dont les résultats sont présentés dans la discussion sur la méthode de base des proportions d'orphelins (chapitre 22). En effet, on remarque que les inégalités par sexe de mortalité adulte sont faibles aux Îles Salomon.

L'application de la méthode fournit une preuve évidente de problèmes de la qualité des données recueillies dans le recensement de 1986. Premièrement, les estimations suggèrent que la mortalité a connu une baisse très rapide, mais les estimations les plus récentes de la série la plus ancienne, déduites des données qui concernent les enfants, indiquent une mortalité sensiblement plus faible que les estimations faites pour quelques années plus tard à partir des données recueillies auprès des enquêtés plus âgés dans le recensement de 1989. Les incohérences de ce genre indiquent généralement que les estimations sur les données du pre-

FIGURE 26.2 Tendances de la probabilité de décéder entre 15 et 60 ans exacts, estimés à partir des données sur la survie des parents, Îles Salomon, 1986 et 1999 recensements



mier recensement (1986) sont trop basses parce que l'effectif des enfants orphelins est sous-estimé en raison de l'effet d'adoption. Dans la mesure où la tendance à sous-déclarer les enfants orphelins peut être une caractéristique persistante de la culture d'une population, ces incohérences jettent ainsi un doute sur les estimations les plus récentes faites à partir des données collectées au recensement suivant. Le deuxième problème avec les estimations du recensement de 1986 aux Îles Salomon est qu'elles suggèrent que les femmes ont une mortalité adulte plus élevée que les hommes. C'est une situation très atypique.

Les estimations faites pour le début des années 1990 à partir des proportions d'orphelins parmi les jeunes adultes organisées dans des cohortes synthétiques appuient l'idée selon laquelle les estimations calculées à partir des données recueillies sur les enfants en 1999 sont également trop faibles. Les estimations obtenues à l'aide des cohortes synthétiques sont probablement les plus fiables puisqu'elles sont basées exclusivement sur les données des orphelins adultes. Ainsi, on peut provisoirement conclure que la probabilité

de décéder entre 15 et 60 ans aux Îles Salomon est passée de 30 pourcents environ au début des années 1970 à 17,5 pourcents sur l'espace de deux décennies.

DESCRIPTION DÉTAILLÉE DE LA MÉTHODE

Introduction

Des méthodes simples et robustes pour estimer la mortalité à partir des données relatives à des cohortes d'orphelins et recueillies dans une seule opération de collecte ont d'abord été publiées par Brass and Hill (1973). Zlotnik and Hill (1981) ont été les premiers à souligner qu'à partir du moment où les questions « Votre mère est en vie ? » ou « Votre père est vivant ? » ont été posées au sein d'une même population dans deux enquêtes successives, il devient possible de calculer des mesures relatives à la survie des parents au sein de cohortes synthétiques et obtenir ainsi des indices qui reflètent la mortalité des adultes dans la période entre les deux enquêtes. Puisque ces estimations sont faites à partir des changements survenus au niveau de la survie des parents entre les deux

enquêtes, les indicateurs de mortalité adultes qui en découlent sont vulnérables aux différences entre enquêtes en matière d'erreurs de déclaration et d'erreurs d'échantillonnage. Toutefois, la période de référence des mesures est généralement plus récente que celle des deux autres séries d'estimations simplement basées sur les proportions d'orphelins observées dans les cohortes prises individuellement.

Les données des cohortes synthétiques ont un autre avantage : si les décès survenus au cours de la période entre les deux enquêtes sont tous complètement déclarés, les omissions de décès plus éloignés dans le passé n'auront pas d'impact sur les résultats. Ainsi, les données des cohortes synthétiques sur la survie des parents sont potentiellement moins vulnérables à « l'effet d'adoption » que les données des cohortes simples prises individuellement, à cause de la sous-déclaration des très jeunes enfants orphelins par les enquêtés. Ceci est important, car, l'effet d'adoption est le biais majeur affectant la méthode des proportions d'orphelins ; ce biais explique les résultats invraisemblables et les incohérences entre enquêtes successives observées dans un certain nombre d'applications de cette méthode. La façon la plus simple de corriger ces biais est d'analyser uniquement les données sur les orphelins adultes en les organisant dans des cohortes synthétiques (Timæus 1991). On peut le faire en construisant une cohorte synthétique qui débute à 20 ans exacts et combine les données sur la survie des parents collectées au cours de deux opérations successives. Cette cohorte indique la proportion de la population adulte dont les mères ou les pères resteraient en vie s'ils étaient exposés à la mortalité du moment, parmi ceux qui, à 20 ans exacts, avaient une mère ou un père encore vivant. Une telle cohorte synthétique peut être construite uniquement à partir des données relativement fiables fournies par les jeunes adultes.

Timæus (1991) propose d'utiliser 20 ans comme base de la cohorte pour deux raisons. Tout d'abord, ce choix minimise la possibilité de sous-estimer les proportions d'orphelins à cet âge de base et par conséquent, il permet d'éviter une surestimation des proportions d'orphelins et de la mortalité des adultes aux âges suivants. Deuxièmement, parce que l'information sur deux groupes d'âge est nécessaire pour estimer la survie des parents à l'âge exact séparant les groupes, cette approche permet d'appliquer la méthode aux données recueillies dans les enquêtes dans lesquelles seules les femmes âgées de 15 à 49 ans sont interrogées sur la survie de leurs parents.

La généralisation à toutes les populations des relations établies par la théorie des populations stables (Preston et

Coale 1982) entre les structures par âge de la population, l'accroissement démographique et la mortalité offre un moyen pratique pour construire de telles cohortes synthétiques. Les mesures sur la survie des parents organisées dans des cohortes synthétiques et correspondantes à une population stationnaire peuvent être obtenues en ajustant les proportions observées des parents survivants en utilisant les taux de croissance des proportions par âge, pour éliminer l'impact des tendances passées de la mortalité. Lorsque les données proviennent de deux enquêtes, l'ajustement en utilisant les taux d'accroissement de la survie des parents par âge a un avantage par rapport aux méthodes qui combinent les changements survenues au niveau des cohortes ; elles sont faciles à appliquer lorsque l'intervalle entre les enquêtes ou recensements ne sont pas de cinq ou dix ans.

Exposé mathématique

Preston et Coale (1982) montrent que dans une population fermée répartie par âge, on peut établir l'équivalence suivante :

$$N(0,t) = N(a,t) \cdot e^{\int_0^a \mu(z,t) dz} \cdot e^{\int_0^a r(z,t) dz} \quad (1)$$

où $N(a, t)$ est le nombre d'individus d'âge a à l'instant t , et $\mu(z, t)$ et $r(z, t)$ représentent respectivement l'intensité de mortalité et le taux de croissance à l'âge z et au temps t . L'attrition qui caractérise la population d'individus dont la mère ou le père est vivant, notée NO , peut être décomposée en deux composantes : la mortalité des parents et la mortalité de la population elle-même (Preston and Chen 1984 ; Timæus 1986) :

$$NO(0,t) = NO(a,t) \cdot e^{\int_0^a \mu(z,t) dz} \cdot e^{\int_0^a \pi(z,t) dz} \cdot e^{\int_0^a r^{NO}(z,t) dz}$$

où $\pi(z, t)$ représente le taux instantané de perte des parents, et $r^{NO}(z, t)$ le taux de croissance de la population dont les parents survivent toujours, à l'âge z et au temps t . En supposant que les orphelins et le reste de la population soient exposés aux mêmes risques de mortalité et en utilisant le fait que $N(0,t) \equiv NO(0,t)$, le rapport de l'expression pour les non-orphelins sur celle pour la population totale donne :

$$e^{-\int_0^a \pi(z,t) dz} = \frac{NO(a,t)}{N(a,t)} \cdot e^{\int_0^a r^{NO}(z,t) - r(z,t) dz} \quad (2)$$

Dans cette équation, le terme de gauche représente la probabilité stationnaire qu'une personne d'âge a ait une mère ou un père vivant, noté $S(a, \tau)$, tandis que $NO(a, t)/N(a, t)$ est égale à la proportion non corrigée équivalente, $S(a, t)$. Avec des données d'enquêtes, il est plus commode de travailler avec le taux de croissance de la proportion de la population dont les parents sont vivants, $r^s(z, t)$, qu'avec son équivalent, $r^{NO}(z, t) - r(z, t)$, qui est la différence entre les taux d'accroissement pour les non-orphelins et ceux de la population totale.

Au-dessus de n'importe quel âge, la population peut être considérée comme étant autonome et la relation entre la structure par âge, le taux de croissance, et l'intensité de la mortalité indiquée dans l'équation (1) sera toujours valable pour ces populations définies par un âge de départ. Ainsi, en utilisant les notations déjà établies, on a :

$$\frac{NO(20, t)}{N(20, t)} e^{-\int_{20}^a \pi(z, t) dz} = \frac{NO(a, t)}{N(a, t)} \cdot e^{-\int_{20}^a r^s(z, t) dz}$$

pour $a > 20$. Lorsque les deux côtés sont divisés par $S(20, t)$, cela donne :

$$\frac{S(a, \tau)}{S(20, \tau)} = \frac{S(a, t)}{S(20, t)} \cdot e^{-\int_{20}^a r^s(z, t) dz}$$

Sous sa forme discrète, pour les groupes d'âge x à $x + 5$, on a :

$$\frac{{}_5S_x(\tau)}{S(20, \tau)} = \frac{{}_5S_x(t)}{S(20, t)} \cdot e^{\left(\sum_{z=20,5}^{x-5} {}_5r_z(t) \right) + 2,5, {}_5r_x(t)}.$$

Mise en œuvre de la méthode

Afin de simplifier l'estimation des mesures de la table de mortalité à partir de ces proportions (Timæus 1991) a développé des modèles de régression pour la mortalité des hommes et des femmes, en estimant les coefficients à partir des données sur la survie des parents dans le même ensemble de populations simulées que celles utilisées dans la méthode de base des proportions d'orphelins (Timæus 1992).

La proportion de personnes âgées d'âge a qui ont une mère survivante, notée $S(a)$, peut être calculée comme la moyenne des probabilités de survie chez les mères qui ont donné naissance à chaque âge y , pondéré par la proportion de naissances qui se produisent à l'âge y (Brass and Hill 1973) :

$$S(a) = \frac{\int_s^{\omega} e^{-r^y} f(y) l(y + a) dy da}{\int_s^{\omega} e^{-r^y} f(y) l(y) dy}$$

où l'intégrale couvre toute la population en âge de procréer de s à ω ans. La division de $S(a)$ par $S(20)$ pour $a > 20$, annule le dénominateur. Ainsi, la proportion d'individus d'un groupe d'âge quinquennal dont les mères sont encore vivantes, parmi ceux qui avaient une mère survivante quand ils avaient 20 ans exactement s'exprime comme :

$$\frac{{}_5S_x}{S(20)} = \frac{\int_x^{x+5} e^{-ra} l(a) \int_s^{\omega} e^{-r^y} f(y) l(y + a) dy da}{\int_x^{x+5} e^{-ra} l(a) da \int_s^{\omega} e^{-r^y} f(y) l(y + 20) dy} \quad (3)$$

pour $x \geq 20$. La proportion équivalente d'individus dans chaque groupe d'âge dont les pères survivent est :

$$\frac{{}_5S_x}{S(20)} = \frac{\int_{x+5}^x e^{-ra} l(a) \int_s^{\omega} e^{-r^y} f(y) l(y) l(y + a) / l(y - 0,75) dy da}{\int_{x+5}^x e^{-ra} l(a) da \int_s^{\omega} e^{-r^y} f(y) l(y) l(y + 20) / l(y - 0,75) dy} \quad (4)$$

où $f(y)$ représente le calendrier de la fécondité par âge et $l(a)$ correspond aux survivants de la table de mortalité des hommes plutôt que celle des femmes, les âges de procréation s et ω étant aussi ceux des hommes.

Les équations 3 et 4 peuvent être évaluées numériquement en utilisant des tables-types de mortalité et des schémas de fécondité par âge. On peut ensuite ajuster un modèle de régression qui prédit des probabilités de survie à partir des données simulées sur la survie des parents. L'équation d'estimation utilisée pour les orphelins de mère après 20 ans est analogue à celle proposée pour les orphelins depuis la naissance (Timæus 1991, 1992). Cette équation est basée sur l'observation selon laquelle, la proportion de répondants dont les mères sont vivantes, équivaut à une probabilité de survie $l(B + N) / l(B)$, où N est l'âge des répondants et B est proche de l'âge moyen à la naissance des enfants, mais dépend aussi de N (Brass and Hill 1973). Pour les applications pratiques cependant, il est plus commode d'ajus-

ter légèrement les proportions sur base de l'âge moyen à la naissance des enfants et d'estimer les probabilités de survie à partir d'un âge de base arrondi, b , situé près de B , et pour une durée d'exposition n , qui est un multiple de cinq ans. En outre, pour les orphelins après l'âge de 20 ans, l'exposition commence 20 ans après l'âge B . Ainsi, la probabilité de survie est estimée à partir d'un âge de référence de 45 ans et l'équation utilisée pour faire les estimations prend la forme suivante :

$${}_{n-20}p_{45} = a(n) + b(n)\bar{M}^f + c(n) {}_5S_{n-5}(\tau) / S(20, \bar{\tau})$$

L'équation équivalente donne parfois des résultats médiocres pour les hommes. Des estimations plus précises peuvent être obtenues si l'information sur la survie des pères dans deux groupes d'âge adjacents est utilisée pour estimer la mortalité. Si les schémas par âge de la mortalité et de la fécondité diffèrent des schémas généraux reflétés dans les coefficients de régression, les proportions de répondants dont le père est vivant dans le groupe d'âge supérieur et celles du groupe d'âge inférieur sont altérées dans des directions qui se compensent en comparaison à la proportion de pères survivant pour l'âge qui sépare les deux groupes d'âge (Timaues 1992). Si l'on estime les mesures de la table de mortalité à partir des données sur deux groupes d'âge, pour une durée d'exposition égale à l'âge qui les sépare, on réduit la sensibilité des résultats aux variations de la pente de la relation entre la survie des parents et les survivants de la table de mortalité. L'âge moyen à la naissance des enfants des hommes dans les pays en développement est légèrement inférieur à 35 ans en moyenne. Ainsi, les probabilités de survie peuvent être estimées à partir des orphelins après 20 ans en utilisant 55 ans comme âge de référence et un modèle de la forme :

$${}_{n-20}p_{55} = a(n) + b(n)\bar{M}^m + c(n) {}_5S_{n-5}(\tau) / S(20, \bar{\tau}) \\ + d(n) {}_5S_n(\tau) / S(20, \bar{\tau})$$

Les coefficients pour les différents groupes d'âge définis par n sont présentés dans le tableau 26.1 et le tableau 26.2.

AUTRES LECTURES ET RÉFÉRENCES

La méthode de base des proportions d'orphelins est présentée dans tous les manuels classiques d'estimation indirecte (Sloggett, Brass, Eldridge *et al.* 1994 ; Division de la Population des NU 1984) mais, à l'exception de celui des Nations unies sur l'estimation de la mortalité adulte (UN Population

EXTENSIONS DE LA MÉTHODE

En plus de l'analyse des données sur les orphelins issues de deux opérations de collecte en utilisant les cohortes synthétiques, Chackiel et Orellana (1985) soulignent que l'on peut recueillir des données dans une seule enquête pour produire des estimations récentes de la même manière. En plus des données habituelles sur la survie des parents, il est pour cela nécessaire de disposer d'informations sur les dates de décès des parents. Par exemple, l'enquête pourrait poser des questions sur l'année et le mois où le parent est décédé, ou sur le nombre d'années écoulées depuis son décès. Si les dates de décès des parents sont rapportées avec une précision raisonnable, cette information peut être utilisée pour reconstruire la proportion des enquêtés qui avaient des parents vivants cinq et dix ans plus tôt. À partir de ces données transversales successives, on peut construire des indices relatifs à la survie des parents organisés dans des cohortes synthétiques formellement identiques à celles générées à partir des données recueillies dans une série d'opérations de collecte distinctes. Ces indices peuvent ainsi être analysés en utilisant la procédure d'estimation de la mortalité à partir des données de plusieurs enquêtes sur les orphelins à l'âge adulte décrite ici.

Assez peu d'enquêtes ou de recensements ont tenté de recueillir des informations sur la date de décès des parents. La qualité des réponses a été très faible pour certaines d'entre elles, mais dans d'autres enquêtes, les dates de décès pour la décennie précédant l'enquête est du plus grand intérêt analytique, car elles semblent avoir été très bien déclarées. Une approche alternative pour distinguer les parents décédés récemment et qui pourrait générer de meilleurs résultats consiste à demander si les parents sont décédés avant ou après un autre événement du cycle de vie, tel que le mariage ou la naissance du premier enfant du répondant.

Les méthodes pour produire des estimations de mortalité à partir de données de ce type sont décrites avec d'autres méthodes d'analyse de données sur les orphelins obtenues d'une seule enquête ou recensement (chapitre 22).

Division 2002), ces manuels s'attachent à l'ancienne variante de la méthode, qui recourt à des facteurs de pondération pour obtenir des indices de tables de mortalité, plutôt qu'à la méthode fondée sur la régression, couramment utilisée aujourd'hui. Les méthodes fondées sur la régression ont été

proposées précédemment pour les femmes (Hill and Trussell 1977; Palloni and Heligman 1985), mais elles ont été développées pour les hommes plus récemment par Timæus (1992). Ce dernier article passe aussi en revue des contributions plus anciennes sur le sujet et présente les bases théoriques de la méthode.

Les méthodes permettant d'organiser les données sur la survie des parents dans des cohortes synthétiques et d'estimer la mortalité adulte sur cette base ont été proposées dans les années 1980 (Chackiel et Orellana 1985 ; Timæus 1986 ; Division de la Population des NU 1984 ; Zlotnik et Hill 1981). La version de cette approche qui met l'accent sur les orphelins après 20 ans et qui est décrite ici a été proposée par Timæus (1991).

- Blacker JGC. 1984. "Experiences in the use of special mortality questions in multi-purpose surveys: the single-round approach," in *Data Bases for Mortality Measurement*. New York: United Nations, pp. 79-89. http://www.un.org/esa/population/publications/UN_1984_Data_Bases_for_Mortality_Measurement/UN1984_Data_Bases.htm
- Blacker JGC and JM Gapere. 1988. "The indirect measurement of adult mortality in Africa: results and prospects," in *African Population Conference, Dakar, 1988*. Liège: International Union for the Scientific Study of Population, Vol. 2:3.2.23-38.
- Brass W and K Hill. 1973. "Estimating adult mortality from orphanhood," in *International Population Conference, Liège, 1973*. Liège: International Union for the Scientific Study of Population, Vol. 3:111-123.
- Chackiel J and H Orellana. 1985. "Adult female mortality trends from retrospective questions about maternal orphanhood included in censuses and surveys," in *International Population Conference, Florence, 1985*. Liège: International Union for the Scientific Study of Population, Vol. 4:39-51.
- Division de la Population des NU. 1984. *Manuel X. Techniques indirectes d'estimation démographique*. New York : Nations Unies, Département des affaires économiques et sociales internationales, ST/ESA/SER.A/81. http://unstats.un.org/unsd/demographic/standmeth/handbooks/Manuel_X-fr.pdf
- Hill K. 1984. "An evaluation of indirect methods for estimating mortality," in Vallin, J, Pollard John H and L Heligman (eds). *Methodologies for the Collection and Analysis of Mortality Data*. Liège: Ordina, pp. 145-176.
- Hill K and TJ Trussell. 1977. "Further developments in indirect mortality estimation", *Population Studies* **31**(2):313-334. doi: <http://dx.doi.org/10.1080/00324728.1977.10410432>
- Palloni A and L Heligman. 1985. "Re-estimation of structural parameters to obtain estimates of mortality in developing countries", *Population Bulletin of The United Nations* **18**:10-33.
- Palloni A, M Massagli and J Marcotte. 1984. "Estimating adult mortality with maternal orphanhood data: analysis of sensitivity of the techniques", *Population Studies* **38**(2):255-279. doi: <http://dx.doi.org/10.1080/00324728.1984.10410289>
- Preston SH and N Chen. 1984. *Two Census Orphanhood Methods for Estimating Adult Mortality, with Applications to Latin America*.
- Preston SH and AJ Coale. 1982. "Age structure, growth, attrition and accession: A new synthesis", *Population Index* **48**(2):217-259.
- Sloggett A, W Brass, SM Eldridge, IM Timæus, P Ward and B Zaba. 1994. *Estimation of Demographic Parameters from Census Data*. Tokyo, Japan: United Nations Statistical Institute for Asia and the Pacific.
- Timæus I. 1986. "An assessment of methods for estimating adult mortality from two sets of data on maternal orphanhood", *Demography* **23**(3):435-450. doi: <http://dx.doi.org/10.2307/2061440>
- Timæus IM. 1991. "Estimation of mortality from orphanhood in adulthood", *Demography* **28**(2):213-227. doi: <http://dx.doi.org/10.2307/2061276>
- Timæus IM. 1992. "Estimation of adult mortality from paternal orphanhood: a reassessment and a new approach", *Population Bulletin of The United Nations* **33**:47-63.
- UN Population Division. 2002. *Methods for Estimating Adult Mortality*. New York: United Nations, Department of Economic and Social Affairs, ESA/P/WP.175. http://www.un.org/esa/population/techcoop/DemEst/methods_adultmort/methods_adultmort.html
- Zlotnik H and KH Hill. 1981. "The use of hypothetical cohorts in estimating demographic parameters under conditions of changing fertility and mortality", *Demography* **18**(1):103-122. doi: <http://dx.doi.org/10.2307/2061052>

Les données d'enquêtes
et les méthodes directes

Chapitre 27 ESTIMATION DE LA MORTALITÉ ADULTE À PARTIR DE LA SURVIE DES FRÈRES ET SŒURS

Ian M Timæus

DESCRIPTION DE LA MÉTHODE

Cette méthode calcule la mortalité adulte directement à partir de données fournies par les adultes sur leurs fratries (c'est-à-dire, leurs frères et sœurs). Elle ne peut être appliquée que si une enquête a recueilli des histoires de fratries complètes. Ces histoires sont obtenues en demandant à chaque répondant le nom, le sexe, l'âge, le statut de survie et, s'ils sont décédés, l'âge et l'année du décès de chacun de ses frères et sœurs nés de la même mère. Les informations sur les frères sont utilisées pour estimer la mortalité des hommes et les informations sur les sœurs pour estimer la mortalité des femmes. Plusieurs enquêtes recueillent les histoires de fratries uniquement à partir des informations fournies par les femmes ; mais les histoires de fratries collectées auprès des répondants de sexe masculin peuvent être analysées en utilisant exactement les mêmes méthodes. Comme les répondants et leurs frères et sœurs ont en moyenne le même âge, les informations sur les fratries des répondants peuvent être utilisées pour mesurer la mortalité dans les mêmes groupes d'âge que ceux des répondants.

La collecte des histoires de fratries est un processus complexe qui nécessite une formation et une supervision rigoureuses du personnel de terrain pour être exécutée correctement. Ce n'est pas une méthode de collecte dont l'utilisation est appropriée dans un recensement. De nombreuses Enquêtes Démographiques et de Santé (EDS) collectent des histoires de fratries (le terme utilisé dans les EDS est le « Module mortalité maternelle »). Alors que la plupart de ces enquêtes ont recueilli ces informations seulement auprès des femmes âgées de 15 à 49 ans, car c'est pour ces femmes qu'est complété le questionnaire individuel détaillé, certaines EDS ont aussi collecté de la même façon des histoires de fratries auprès des hommes.

Un avantage que les méthodes basées sur la fratrie ont sur les questions relatives au décès survenu dans les ménages est que seuls les recensements ou exceptionnellement les grandes enquêtes peuvent collecter des informations sur un nombre relativement important de décès survenus dans les ménages au cours de l'année précédant l'enquête pour produire des estimations de mortalité qui soient suffisamment précises pour être utiles. Parce que les répondants donnent en moyenne des informations sur plusieurs frères et sœurs, et qu'on peut calculer des taux sur plusieurs années d'exposition, les estimations peuvent être faites à partir d'histoires de fratries collectées dans de plus petites enquêtes. Néanmoins, toutes les méthodes permettant d'estimer la mortalité des adultes nécessitent que soient collectées des données sur plusieurs milliers de ménages. Un autre avantage de la méthode est qu'elle repose sur une procédure d'estimation dans lequel peu d'hypothèses interviennent, en particulier, elle ne suppose pas que la population étudiée soit fermée à la migration. Les résultats ne seront toutefois pas représentatifs concernant des petits États ou des régions dans lesquels une proportion substantielle de la population aura émigré ou sera composée d'immigrés.

CONTEXTE

Les méthodes d'origine élaborées pour estimer la mortalité à partir des informations sur la survie des frères et sœurs étaient des méthodes indirectes fondées sur l'idée que l'âge moyen des frères et sœurs des répondants de tout âge est proche de l'âge de ces répondants. La proportion des frères et sœurs encore en vie d'un répondant est donc un bon estimateur de la probabilité de survivre (dans une table de mortalité) jusqu'à l'âge du répondant (Hill and Trussell 1977 ; Division de la Population des NU 1984). Malheureusement, sur

le terrain, l'application de cette approche a montré que la qualité des données recueillies sur les frères et sœurs était souvent assez pauvre, parce que les frères ou les sœurs qui décèdent avant ou peu après la naissance du répondant sont souvent omis par les répondants qui peuvent tout ignorer d'eux (Blacker and Brass 1983 ; Zaba 1986).

L'intérêt pour l'estimation de la mortalité à partir des données sur les fratries a été relancé par le développement de la méthode des sœurs pour mesurer la mortalité maternelle (Graham, Brass and Snow 1989). Cette méthode suppose que l'on dispose de données sur le nombre de sœurs du répondant qui ont survécu jusqu'à l'âge de 15 ans, le nombre d'entre elles qui sont décédées par la suite, et parmi les sœurs décédées, le nombre de celles qui étaient enceintes au moment du décès, ou qui étaient enceintes dans les 6 à 8 semaines qui ont précédé le décès. En limitant l'analyse aux frères et sœurs à qui ont survécu jusqu'à l'âge de 15 ans, on exclut de fait les frères et sœurs qui sont décédés avant 15 ans et, par conséquent, peuvent avoir été ignorés ou oubliés par le répondant. Les réponses fournies par les répondants aux deux premières questions dans chaque groupe d'âge quinquennal permettent de calculer les proportions de sœurs encore en vie parmi celles qui ont survécu à 15 ans. La méthode indirecte de la survie des frères et sœurs adultes (chapitre 23) a ensuite été développée pour estimer la mortalité des femmes adultes toutes causes confondues à partir des mêmes proportions des femmes encore en vie. Des données comparables sur les frères des répondants peuvent être utilisées pour estimer la mortalité des hommes.

La méthode permettant d'estimer directement la mortalité adulte à partir des informations sur l'histoire des fratries a été lancée par le programme d'Enquêtes Démographiques et de Santé (EDS), en s'inspirant des recherches antérieures détaillées plus haut (Rutenberg and Sullivan 1991). Cette technique est plus ambitieuse en termes de quantité d'information qui est recueillie auprès des répondants ; elle est également plus demandeuse du point de vue des répondants et du personnel de terrain qui mène les interviews. Cependant, en remplaçant les estimations indirectes fondées sur des modèles démographiques par des mesures directes des taux de mortalité des adultes, on réduit le nombre d'hypothèses nécessaires pour produire ces estimations, et de façon plus importante encore, on peut ainsi distinguer les décès survenus dans un passé récent et les décès survenus plus loin dans le passé, qui sont connus pour être rapportés avec moins de précision.

DONNÉES NÉCESSAIRES ET HYPOTHÈSES

Tabulation de données nécessaires

Le calcul direct des taux de mortalité à partir des histoires de fratries implique à peu près les mêmes démarches que le processus plus familier de calcul des taux de mortalité infanto-juvénile à partir des données d'histoires génésiques. En effet, une histoire complète des frères et sœurs du répondant, quand elles incluent le répondant, est en fait une histoire génésique de leur mère. Cependant, en comparaison avec les estimations de mortalité faites pour les enfants, les estimations de mortalité adulte sont entachées d'importantes erreurs d'échantillonnage. Cela reflète le fait que les taux de mortalité sont beaucoup plus bas à l'âge adulte que dans l'enfance et que, dans une population en croissance, le nombre de frères et sœurs exposés aux risques de mortalité est plus petit que le nombre d'enfants déclarés par les mères. Par conséquent, dans toutes les enquêtes auprès des ménages, on observe beaucoup moins de décès parmi les frères et sœurs adultes que parmi les enfants.

Le calcul des mesures de mortalité qui se réfèrent à des cohortes à partir d'histoires de frères et sœurs a peu d'intérêt au vu de la facilité avec laquelle les logiciels modernes d'exploitation et d'analyse de données d'enquêtes permettent de calculer les durées d'exposition au risque de décès. Ainsi, ce document met l'accent sur le calcul des taux de mortalité par âge et par période à partir des données sur l'histoire des fratries et sur la façon avec laquelle on peut en déduire d'autres indicateurs de mortalité.

Pour calculer la mortalité des femmes, on a besoin des données suivantes :

- Le nombre de décès des sœurs des répondants selon la période et le groupe quinquennal d'âge des sœurs au moment de leur décès.
- Le nombre d'années d'exposition vécues par les sœurs et distribuées par période et par groupe quinquennal d'âge de ces sœurs (au moment de l'exposition au risque).

Pour calculer le taux de mortalité des hommes, on a besoin des données suivantes :

- Le nombre de décès des frères des répondants selon la période et le groupe quinquennal d'âge des frères au moment de leur décès.
- Le nombre d'années d'exposition vécues par les frères et distribuées par période et par groupe quinquennal d'âge de ces frères (au moment de l'exposition au risque).

Le calcul des durées d'exposition doit généralement exclure le répondant lui-même ou la répondante elle-même (par définition, ceci est fait quand on analyse les données sur les sœurs récoltés par les hommes et *vice versa*). Cette condition est expliquée dans la discussion sur les hypothèses importantes formulées pour cette méthode.

Les tableaux sur la fratrie de même sexe que le répondant (c'est-à-dire les sœurs des femmes et les frères des hommes) doivent être pondérés seulement par les pondérations d'échantillonnage fournies avec les données. Les tableaux sur la fratrie de sexe opposé (c'est-à-dire les frères des femmes et les sœurs des hommes) doivent en outre être pondérés par l'inverse de la taille de la fratrie survivante de même sexe que le répondant. Cette nécessité sera aussi expliquée dans la discussion sur les hypothèses importantes faites par la méthode.

La période de temps pendant laquelle l'exposition au risque de mortalité est mesurée peut être définie en termes de dates du calendrier ou par rapport à la date à laquelle le répondant a été interviewé. Cette dernière approche permet une utilisation optimale des données, car elle garantit assure que le temps vécu par les fratries des répondants au cours des années incomplètes dans lequel les entretiens ont eu lieu est tout de même inclus dans l'analyse. Alternative-ment, le calcul des taux de mortalité sur des années civiles particulières a l'avantage de donner des résultats qui correspondent à un temps précis et comparable avec celui associé aux estimations provenant d'autres sources. Cette approche implique généralement de ne pas prendre en compte les données de l'année où l'enquête a eu lieu sur le terrain (bien qu'on puisse choisir d'inclure ces données dans la période la plus récente si le plus gros des interviews a eu lieu à la fin de cette année).

La feuille de calcul disponible sur <http://demographi-cestimation.iussp.org/fr> est conçue pour calculer les taux de mortalité pour deux périodes successives antérieures à la date de la collecte des données sur le terrain. Beaucoup de rapports d'EDS fournissent des taux pour la période de sept ans précédant l'enquête (soit, 0-6 années complètes) et une façon d'utiliser ce fichier Excel serait de calculer les taux de mortalité pour les trois années précédant l'enquête et ensuite pour les quatre années qui précèdent ces trois années. Alternative-ment, on pourrait calculer les taux de mortalité pour deux périodes de quatre ans et donc pour une période totale de huit ans précédant la collecte des données sur les histoires des fratries. L'expérience montre que l'exhaustivité de la déclaration des frères et sœurs décédés et la précision avec

laquelle leurs âges et leurs dates de décès sont déclarés se dégradent souvent rapidement pour des événements survenus loin dans le passé. En outre, en utilisant des périodes de quatre ans plutôt que de cinq ans, on minimise les erreurs qui résultent de l'attribution par erreur de dates de décès à cinq et dix ans avant la collecte des données en raison d'une tendance à arrondir les années.

Les groupes d'âge de frères et sœurs pour lesquels les données sont collectées doivent correspondre globalement à l'âge des répondants qui ont procuré les informations. Par exemple, afin de mesurer la mortalité entre 15 et 60 ans, dans les meilleures de cas, les données sur la survie de frères et sœurs devraient être recueillies auprès des répondants âgés de 15 à 59 ans. Si les données sont collectées uniquement auprès d'adultes, peu d'entre eux ont des frères et sœurs qui sont encore de jeunes enfants, et, même si les données sur les enfants ne sont pas biaisées par l'omission de certains frères et sœurs décédés, les estimations sont susceptibles d'être entourées par d'importantes erreurs d'échantillonnage et elles ne seront pas représentatives de tous les jeunes enfants. Ainsi, pour les EDS et les autres enquêtes qui recueillent des données sur les histoires de fratries auprès de répondants âgés de 15 à 49 ans, l'indice de mortalité préféré est ${}_{35}q_{15}$, c'est-à-dire la probabilité qu'un individu âgé de 15 ans décède avant d'atteindre son 50^e anniversaire.

Bien que la probabilité conditionnelle de décéder entre 15 et 60 ans (${}_{45}q_{15}$) soit largement utilisée par les agences internationales et d'autres organisations comme indice de prédilection de mortalité des adultes, le nombre de décès déclarés par les répondants âgés de 15 à 49 pour la tranche d'âge 55-59 ans sera plus faible que celui rapporté pour les groupes d'âge plus jeunes. Ainsi, plutôt que de calculer directement ${}_{45}q_{15}$, il est préférable de le faire en ajustant une table de mortalité standard aux estimations obtenues pour les âges de 15 à 54 ans, puis en extrapolant à partir de ce modèle pour obtenir la mortalité dans la dernière classe d'âge de cinq ans. C'est l'approche qui est mise en œuvre dans le fichier Excel (voir le site web).

Pour éliminer les ambiguïtés résultant des mariages polygames et des remariages, il est généralement indiqué aux enquêteurs que la « fratrie » concerne les enfants nés de la même mère. Que cela ait été fait ou non, les réponses peuvent généralement être acceptées telles qu'elles sont. Aussi longtemps que les répondants ont le même groupe de parents à l'esprit quand ils listent les frères et sœurs en vie, d'une part, et d'autre part, ceux sont décédés, il importe peu

pour l'estimation de la mortalité de savoir exactement qui sont les parents des frères et sœurs.

Si les histoires de fratries ont été recueillies auprès des hommes et des femmes, leurs réponses devraient généralement être exploitées séparément de sorte que les deux ensembles de données puissent être pondérés de manière appropriée et qu'il soit possible de vérifier la cohérence de l'une série par rapport à l'autre.

Hypothèses importantes

Une limitation inhérente aux méthodes de mesure de la mortalité adulte fondées sur les fratries est qu'elles sous-estiment la mortalité dans la mesure où cette mortalité se concentre au sein de certaines fratries (c'est-à-dire au sein de certains ensembles de frères et/ou de sœurs nés de la même mère). Cette concentration (*clustering* en anglais) survient chaque fois que les décès sont plus concentrés dans une petite fraction des fratries que sous le seul effet du hasard et elle résulte de l'hétérogénéité entre fratries face au risque de décès des individus (Zaba and David 1996). Il en résulte un biais vers le bas des estimations de mortalité, simplement parce que les membres des fratries à forte mortalité sont moins nombreux que les membres de fratries à faible mortalité à survivre pour répondre aux questions sur leurs frères et sœurs. Ce biais ne peut pas être complètement corrigé car, à la limite, les fratries dont les membres sont tous décédés ne sont déclarées par personne. Il n'est pas possible de savoir combien de ces fratries ont existé ou quelles étaient leurs tailles, bien qu'il soit possible de faire des suppositions à leur égard.

Les schémas par âge estimés et les tendances de la mortalité seront biaisées si l'effet de *clustering* de la mortalité au sein des fratries varie avec l'âge. En effet, si les caractéristiques partagées par les frères et sœurs (par exemple, les facteurs génétiques, les premières expériences de la vie, le statut socio-économique, les styles de vie, et le lieu de résidence) influencent fortement la mortalité des adultes d'âge moyen, alors que la mortalité avant 40 ans a une grande composante aléatoire, les estimations de mortalité pour les adultes plus âgés seront davantage sous-estimées que celles faites pour les jeunes adultes.

La question du biais résultant des réponses multiples aux questions sur les fratries a donné lieu à une abondante littérature. Le problème existe aussi bien dans les données d'enquêtes que de recensements, car plus un individu est cité dans un recensement, plus il a de chances d'avoir un

frère ou une sœur qui le cite dans un échantillon probabiliste.¹ En outre, même dans les enquêtes, il peut arriver que des réponses multiples soient données concernant un même individu. Par exemple, si deux sœurs de la même mère sont interrogées dans le même ménage, il y aura des déclarations multiples concernant les autres membres de leur fratrie. La procédure d'analyse standard utilisée, par exemple, dans les rapports EDS s'appuie sur les événements et les durées d'exposition des frères et sœurs cités, en laissant de côté la durée d'exposition de la répondante (survivante) elle-même. Les événements et la durée d'exposition sont pondérés seulement par les pondérations d'échantillonnage du répondant, sans prendre en compte les nombres de répondants survivants potentiels dans la fratrie.

Trussell et Rodriguez (1990) ont démontré mathématiquement que, pour des fratries dont tous les membres font face à un risque identique de décéder, le calcul standard qui exclut aussi le répondant du dénominateur des mesures donne des estimations non biaisées de la mortalité. En effet, la réduction – au numérateur – du nombre de mentions d'un frère ou d'une sœur décédés, du fait que les personnes décédées ne peuvent pas se citer l'une l'autre, et l'exclusion des répondants vivants du dénominateur se compensent mutuellement et donnent le risque correct concernant chaque fratrie en tant que groupe.

La question des biais qui pourraient résulter de différences de mortalité en fonction de la taille de la fratrie est liée à la question de la déclaration multiple de certains frères ou sœurs. Elle a suscité beaucoup d'intérêt chez les chercheurs car, à la différence d'autres facteurs qui jouent sur le risque au sein des fratries classées par sexe et âge du répondant, la taille de la fratrie de chaque répondant est connue. Si la mortalité ne varie pas avec la taille de la fratrie, les estimations standard sont les mêmes à la fois pour chaque taille de fratrie, y compris les fratries d'une personne qui sont exclues de l'analyse parce que le répondant n'a personne à déclarer, et pour la population dans son ensemble. Même si la mortalité varie avec la taille de la fratrie, les estimations standard restent sans biais pour chaque taille de la fratrie, comme l'a montré Masquelier (2013). Mais pour obtenir des esti-

1 La question discutée ici est celle du biais dans l'estimation centrale de la mortalité. Elle ne doit pas être confondue avec la question du calcul correct de l'erreur d'échantillonnage des estimations de mortalité, dans les enquêtes où certains individus sont cités plus d'une fois et où la mortalité constitue des grappes parmi les fratries.

mations de la mortalité pour la population, il convient de repondérer les estimations relatives aux fratries de différentes tailles en fonction de la distribution des tailles des fratries observée dans la population. Quand les répondants citent les personnes de même sexe, on peut le faire en divisant la proportion de frères ou sœurs dans les fratries survivantes de même taille par la fréquence des fratries de cette taille dans la population. Pour les fratries d'une seule personne, leur mortalité doit être extrapolée à partir de la mortalité dans les fratries de plus grandes tailles.

Gakidou et King (2006) soutiennent que, à l'inverse de la démarche standard, les fratries doivent inclure le répondant survivant, mais elles doivent toujours être pondérées en plus par la probabilité qu'elles soient déclarées – c'est-à-dire par l'inverse du nombre de répondants potentiels qui survivent dans la fratrie. Comme dans la démarche de Masquelier, un ajustement supplémentaire doit aussi être fait pour les fratries non déclarées du fait du décès de tous leurs membres. Dans une analyse des histoires complètes de fratries tirées d'enquêtes EDS, Obermeyer, Rajaratnam, Park *et al.* (2010) estiment que l'absence d'ajustement pour tenir compte des chances variables d'être mentionné par plusieurs frères et sœurs peut biaiser de – 20 % les estimations de la mortalité toutes causes confondues.

Toutefois, Masquelier (2013) soutient que Obermeyer *et al.* ont repondéré les données de façon inappropriée et qu'il en est résulté une surestimation du biais. Il souligne que, si on doit repondérer, il importe d'ajuster seulement pour tenir compte de la chance d'être mentionné plusieurs fois, mais uniquement par les frères et sœurs qui ont survécu à l'âge initial à partir duquel la mortalité est mesurée. Mais il s'interroge en outre sur la réalité de la variation de la mortalité par taille de la fratrie. Il pourrait s'agir d'un artefact dû à l'omission plus fréquente des frères et sœurs décédés dans les histoires relatives à des fratries de plus grandes tailles. C'est pourquoi Masquelier recommande d'utiliser l'approche standard, sans essayer de repondérer les données relatives à chaque taille de fratrie. À partir de là, soit la mortalité est estimée pour chaque taille de la fratrie, et une estimation repondérée est obtenue pour la population de la même façon que l'approche décrite dans les précédents paragraphes, soit les estimations ne font l'objet d'aucune pondération supplémentaire. Comme il est nécessaire, pour obtenir des estimations séparées pour chaque taille de fratrie, que l'on dispose d'un très large échantillon ou que l'on utilise des modèles de régression pour ajuster les données, c'est la

seconde approche (l'absence de pondération spécifique pour les tailles de fratries) qui est retenue ici.

La problématique est différente quand on analyse les déclarations qui portent sur le sexe opposé (par exemple, les réponses faites par des femmes au sujet de leurs frères). Dans ce cas, le répondant ne fait pas partie du groupe exposé au risque de décéder. Toutefois, le calcul standard donnera encore des résultats biaisés pour l'ensemble de la population si la mortalité des frères est associée au nombre de sœurs, dans l'exemple où c'est l'une d'entre elles qui est interrogée. Concernant les déclarations qui portent sur le sexe opposé, il est donc clair qu'il faut pondérer chaque déclaration par l'inverse du nombre de frères ou sœurs du même sexe que le répondant, comme suggéré par Gakidou et King (2006). Bien entendu, les questions sur les membres de la fratrie de sexe opposé ne peuvent évidemment donner aucune information sur les fratries au sein desquelles aucun membre survivant n'est de même sexe que le répondant. L'adoption de cette démarche revient donc à supposer que la mortalité des individus dans de telles fratries est égale à la mortalité du reste de la population. Mais dans les enquêtes qui recueillent des données auprès des deux sexes, chacun d'entre eux fournit cette information sur l'autre, de sorte qu'on peut pondérer les décès et l'exposition au risque déclarés par les répondants par l'inverse de la probabilité que les frères et sœurs dans chaque groupe d'âge soient mentionnés.

TRAVAUX PRÉPARATOIRES ET RECHERCHES PRÉLIMINAIRES

La première étape dans l'analyse des données sur l'histoire des fratries devrait consister à évaluer l'ampleur de la non-déclaration et de l'incomplétude des informations dans l'ensemble de données, en particulier le nombre de répondants qui ont déclaré qu'ils ne savaient pas si une sœur ou un frère particulier était encore en vie ou ont simplement omis de répondre à la question. Il est également important d'évaluer la qualité de données sur les dates naissance des frères et sœurs et de vérifier si, l'âge à l'enquête ou leur âge au décès sont manquants ou ont été imputés. Si beaucoup de répondants n'ont pas répondu à ces questions, les données observées peuvent ne pas être représentatives de la population dans son ensemble. En plus, un niveau élevé de non-réponse peut indiquer que le personnel de terrain ou les répondants ont eu quelques difficultés avec les questions ; il peut être éclairant de déterminer si ces problèmes sont observés chez une minorité des enquêteurs ou au sein d'une catégorie

spécifique de répondants. Une concentration des âges ou des années sur certains chiffres particuliers reflète également une mauvaise qualité de l'information recueillie. Si les données sur l'âge et les dates semblent être de mauvaise qualité, on peut potentiellement obtenir de meilleurs résultats en analysant les données avec la méthode indirecte relative à la survie des frères et des sœurs à l'âge adulte (chapitre 23).

Si les femmes et les hommes ont été interrogés, un moyen utile de vérifier l'exhaustivité des données est d'évaluer combien de frères et sœurs sont rapportées en moyenne par les répondants du sexe opposé et si le rapport de masculinité à la naissance change nettement à mesure que l'âge du répondant ou l'intervalle de temps depuis la naissance du frère ou de la sœur augmente. Il est assez fréquent d'observer que les enquêtés d'un des deux genres (généralement des hommes) rapporte moins de frères et sœurs, et en particulier moins de frères et sœurs décédés. Dans d'autres enquêtes, les hommes et les femmes peuvent déclarer un même nombre de frères et sœurs de chaque sexe (après avoir ajusté les données pour tenir compte du fait que les répondants ne renseignent pas sur eux-mêmes), mais des nombres différents de frères et sœurs qui sont survivants. Le premier type d'écarts pourrait résulter de déclarations d'âge erronées, ce qui n'est pas le cas du second type d'incohérences.

Tout biais associé à l'effet de *clustering* de la mortalité au sein de certaines familles résulte se traduit par une sous-estimation. En outre, il semble peu probable que les répondants inventent les frères et sœurs qui n'existent pas ou déclarent que leurs frères et sœurs vivants sont décédés. Ainsi, l'analyse devrait probablement se concentrer sur les données fournies par les répondants du sexe qui renseignent le plus de frères et sœurs et la plus grande proportion de frères et sœurs décédés.

PRÉCAUTIONS ET MISES EN GARDE

- Les seules méthodes faisant usage de données sur les fratries qui peuvent être recommandées produisent des probabilités de survie au-delà de 15 ans, conditionnées au fait d'avoir au moins survécu à 15 ans exacts. En théorie, il est possible de collecter et d'analyser des données sur les décès de frères et sœurs quand ils sont encore des enfants. Malheureusement, ces données sont souvent très incomplètes, en particulier pour les frères et sœurs qui sont décédés avant ou peu après la naissance du répondant. Ainsi, la plupart des applications de la méthode cherchent uniquement à mesurer la mortalité des frères et sœurs à 15 ans ou plus. Pour produire une table de morta-

lité complète, il faut estimer des probabilités de survie de la naissance à 15 ans en faisant recours à d'autres sources de données.

- Même dans une grande enquête, le nombre de frères et sœurs qui décèdent chaque année dans chaque groupe d'âge reste faible. Dans la plupart des applications de la méthode, les décès et la période d'exposition doivent être regroupés sur plusieurs années afin d'estimer des taux de mortalité qui sont assez précis pour être utiles. Ainsi, la méthode est peu susceptible d'être utile pour détecter des variations brusques ou des fluctuations rapides de la mortalité chez les adultes.
- À mesure que l'on exploite les données en remontant dans le temps avant la période de collecte, le nombre de frères et sœurs exposés au risque de décéder diminue rapidement, en particulier aux âges élevés. En outre, les répondants ont tendance à omettre certains frères et sœurs décédés, en particulier les décès qui sont survenus dans un passé lointain. Ainsi, les histoires de fratrie conduisent généralement à sous-estimer la mortalité et ce biais s'aggrave pour les estimations qui sont éloignées de la période de collecte. Ainsi, la méthode directe relative aux frères et sœurs ne devrait pas être utilisée pour estimer les taux de mortalité pour une période éloignée de plus de 10 ans de l'enquête. Souvent, seules les données portant sur les sept dernières années sont analysées.
- Étant donné que la probabilité d'omission des frères et sœurs décédés augmente à mesure que l'intervalle de temps écoulé depuis leurs décès et l'enquête s'agrandit, le calcul des taux de mortalité pour deux périodes précédant l'enquête sert surtout à un examen de la plausibilité des estimations. Il permet à l'analyste de vérifier si les données indiquent une augmentation invraisemblable de la mortalité. Il faut être prudent avant d'extrapoler les tendances de la mortalité adulte à partir des estimations faites à partir d'un seul ensemble de données sur la fratrie.
- La procédure directe d'estimation de la mortalité adulte à partir de l'information sur les frères et sœurs adultes n'implique pas l'hypothèse d'une population fermée à la migration. Il peut néanmoins être difficile d'interpréter des estimations fondées sur la fratrie concernant la mortalité d'unités géographiques infranationales, comme des zones urbaines et rurales ou des districts, ou des groupes de répondants ayant des caractéristiques socioéconomiques spécifiques. En effet, si les frères et sœurs partagent généralement la même identité ethnique, il arrive

fréquemment que les frères et sœurs vivent ailleurs que les répondants et aient des caractéristiques socioéconomiques différentes d'eux. Les estimations relatives aux populations infranationales risquent aussi d'avoir de très larges intervalles de confiance.

APPLICATION DE LA MÉTHODE

Que l'on analyse les données sur les frères, les sœurs, ou les deux sexes combinés, et que l'on utilise les informations procurées par les hommes, les femmes ou les deux sexes, la procédure d'estimation des taux de mortalité à partir des histoires de fratries est la même. Le fichier Excel associé à ce chapitre est conçu pour calculer les taux de mortalité pour les frères et pour les sœurs pour deux périodes précédant une enquête et pour l'ensemble de la période couverte par la combinaison de ces deux périodes. Des feuilles de calcul séparées sont fournies pour l'analyse des données fournies par les hommes et les femmes enquêtées. Le classeur peut produire des estimations pour des périodes de durée variable et pour les données exploitées par année avant l'enquête ou pour des années calendrier qui correspondent globalement à ces années avant l'enquête.

Deux tabulations sont nécessaires pour calculer les taux de mortalité pour les frères et sœurs de chaque sexe à partir des informations fournies par les répondants de chaque sexe, donnant potentiellement lieu à quatre paires de tabulations. Un tableau croisé devrait porter sur le nombre de décès de frères et sœurs par âge pour chaque année civile et un autre tableau devrait concerner les personnes-années d'exposition au risque de décès par âge pour chaque année civile. On peut produire les données pour les groupes d'âge et les périodes qui vont être utilisées dans l'analyse. Alternativement, on peut produire les tableaux par année d'âge et par année calendrier de sorte que les données sur les décès et les personnes-années puissent être ensuite agrégées dans un ensemble d'intervalles plus larges jugées pertinentes.

Comme on peut supposer que les dates de naissance ou de décès de certains frères et sœurs n'auront pas été déclarées, la manière la plus satisfaisante pour corriger ce problème dans les données consiste à imputer des dates exactes à l'aide de nombres aléatoires afin d'attribuer à un frère ou à une sœur une date exacte à laquelle l'événement a pu se produire (Stanton, Abderrahim and Hill 1997). Les organismes qui ont conduit les enquêtes, tels que MeasureDHS, ont peut-être procédé à de telles imputations avant de diffuser les données provenant des enquêtes qu'elles ont réalisées.

Pour que l'imputation et l'analyse des données portant sur certains frères et sœurs soient possibles, on a besoin de connaître soit leur date de naissance ou leur âge actuel s'ils sont en vie, ainsi que leur année de décès, leur âge au décès ou le temps écoulé en année depuis leur décès s'ils sont décédés. Si les deux dates de naissance et de décès sont toutes deux incomplètes, on peut généralement attribuer au hasard à la personne une date exacte de naissance avant de lui assigner une date de décès précise. Des précautions doivent être prises pour enregistrer la clé (appelée *seed* en anglais) utilisée pour générer des nombres aléatoires, de sorte que les dates imputées puissent être reproduites exactement si on a besoin de recréer les fichiers de données utilisées pour estimer les taux de mortalité à partir des données originales.

Les détails de la procédure qui devrait être utilisée pour procéder à l'imputation des dates exactes dépendent des questions précises posées aux répondants notamment sur les âges et les dates de décès de leurs frères et sœurs. Quelques exemples suffiront pour illustrer les principes des calculs. Si le répondant a été interviewé le 23/11/2011 et a indiqué que l'un de ses frères et sœurs était âgé de 33 ans, la date de naissance de cette personne doit se situer entre le 24/11/1977 et le 23/11/1978. Si un frère est né en octobre 1972 et décédé à 17 ans, il faudrait d'abord lui attribuer au hasard une date exacte de naissance, peut-être le 14/10/1972, puis lui attribuer une date exacte de décès entre le 14/10/1989 et le 13/10/1990. Si le répondant a donné une information sur l'année de décès de son frère, la période dans laquelle on choisit au hasard une date de décès devrait être limitée à l'année correcte en question.

Il faut s'assurer que la date de décès qui a été assignée à partir d'un nombre d'années écoulées depuis la naissance du frère ou de la sœur décédé n'excède pas la date de l'interview du répondant. Par exemple, si le répondant a été interviewé le 28/2/2003 et a déclaré que son frère est né en 1980 et décédé à 23 ans, alors on considérera que cette personne est née dans les deux premiers mois de 1980 et que son décès a eu lieu en 2003, mais, au plus tard le 28 février 2003. La procédure d'imputation doit également veiller à ce que les dates de naissance imputées se produisent dans un ordre temporel correct.

Une fois qu'une date exacte de naissance a été attribuée à chaque frère et sœur et, qu'une date exacte de décès a été assignée à ceux qui sont décédés, il est facile d'identifier les groupes d'âge où les décès se sont produits et de répartir les durées vécues par chaque personne entre les groupes d'âge

et les périodes d'exposition (Stanton, Abderrahim and Hill 1997). Les logiciels modernes d'analyse de données d'enquêtes contiennent souvent des commandes qui font ce calcul de façon semi-automatisée.

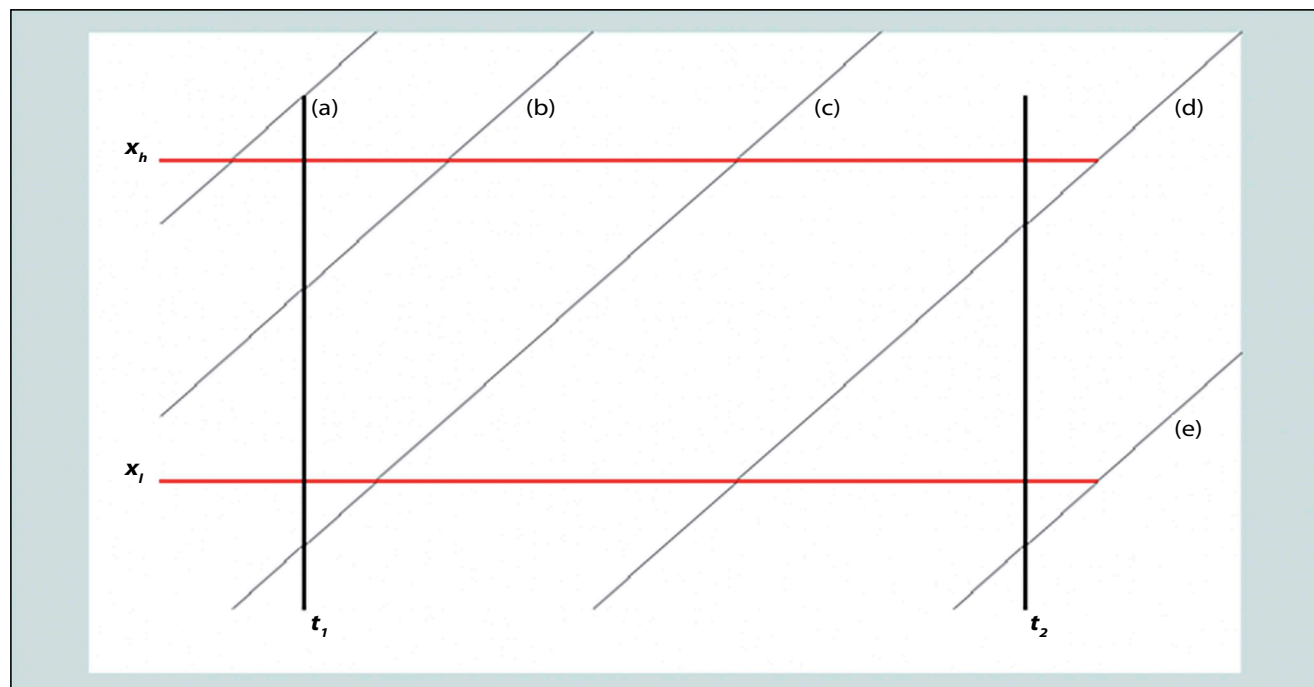
Graphiquement, la mortalité est mesurée pour le groupe d'âge et la période de temps définie par les lignes épaisses dans la figure 27.1. Le parcours de vie d'un individu est représenté par âge et période par les lignes diagonales (telles qu'il le serait dans un diagramme de Lexis classique). Le groupe d'âge pour lequel la mortalité est calculée est défini par un intervalle ayant x_l comme limite inférieure, et x_h comme limite supérieure. La période de temps pendant laquelle la mortalité est calculée est $(t_2 - t_1)$, où t_2 est la date de fin de la période et t_1 la date du début de la période. Ainsi, toutes les personnes d'âge x au temps t_1 qui ne décèdent pas avant t_2 seront âgées de $x_{t2} = x_{t1} + (t_2 - t_1)$ au temps t_2 . Pour les statistiques sur la mortalité des adultes, l'âge et les périodes de temps sont presque toujours mesurés en années.

Il convient de noter que la durée d'exposition vécue par chaque frère ou sœur dans une année donnée est presque toujours divisée entre deux âges adjacents. Cinq scénarios

possibles sont représentés sur la figure 1, marqué de (a) à (e). En notant x_d l'âge au décès des personnes qui décèdent dans un groupe d'âge donné et une période donnée, le calcul pour chaque scénario des personnes-années d'exposition vécues par les frères et sœurs des répondants dans ce groupe d'âge et cette période peut être fait à partir des formules reprises dans le tableau 27.1. En additionnant les durées d'exposition de l'ensemble des frères et sœurs, on obtient le dénominateur pour le taux de mortalité, soit le total d'années d'exposition au risque dans le groupe d'âge au cours de la période. La somme de décès survenus dans la même tranche d'âge et la même période fournit le numérateur du taux.

Une fois que les tabulations des décès et des durées d'exposition ont été produites, diverses mesures de mortalité peuvent être générées en utilisant les formules standards de construction d'une table de mortalité. Dans la feuille de calcul Excel (voir le site web), ces calculs sont effectués pour les données présentées par groupes quinquennaux d'âge. Le taux de mortalité par âge, ${}_5M_x$, est obtenu en divisant les décès dans un groupe d'âge quinquennal au cours d'une année spécifique (ou d'un ensemble d'années) par les per-

FIGURE 27.1 Diagramme de Lexis illustrant le calcul de l'exposition au risque



sonnes-années d'exposition au risque de mortalité dans ce groupe et durant cette période :

$${}_5M_x = \frac{{}_5D_x}{{}_5N_x}$$

La probabilité de décéder dans un groupe d'âge quinquennal, ${}_5q_x$ peut être calculée pour les années concernées à partir du taux de mortalité correspondant à l'aide de la formule standard, qui suppose que les décès sont répartis uniformément à l'intérieur du groupe d'âge :

$${}_5q_x = \frac{5 \times {}_5M_x}{(1 + 2,5 \times {}_5M_x)}$$

La probabilité de survie dans un groupe d'âge quinquennal, ${}_5p_x$, est $1 - {}_5q_x$.

À partir de la série d'estimations de ${}_5p_x$, on peut calculer la probabilité cumulée de décéder entre 15 ans et 50 ans pour la période (${}_{35}q_{15}$), en multipliant les probabilités intermédiaires de survie dans les intervalles de cinq ans pour obtenir la probabilité de survie de 15 à 50 (${}_{35}p_{15}$), et en soustrayant cette probabilité de 1 pour obtenir son complément :

$${}_{35}q_{15} = 1 - {}_5p_{15} \times {}_5p_{20} \times {}_5p_{25} \dots \times {}_5p_{45}$$

Les intervalles de confiance à 95 pour cent qui entourent ces mesures synthétiques de la mortalité des adultes sont calculées dans la feuille de calcul en utilisant la formule de Greenwood. Cette formule suppose que les données sont générées à partir d'un échantillon aléatoire simple et suresti-

ment ainsi la précision des indicateurs basés sur des données provenant d'enquêtes par grappes.

La feuille de calcul fournit une représentation graphique des logits des probabilités conditionnelles de survie de 15 ans à chacun des âges ultérieurs, rapportés aux valeurs équivalentes des logits des survivants d'une table type de mortalité. Ces diagrammes sont utiles pour évaluer la qualité des estimations. Les erreurs de données apparaissent habituellement sous la forme d'irrégularités dans la série ou sur la forme d'une courbe qui fléchit vers le bas dans les groupes d'âge les plus élevés. Ce dernier trait est caractéristique d'une sous-estimation de la mortalité due à l'exagération des âges à l'enquête et des âges au décès.

Enfin, la feuille de calcul ajuste un modèle relationnel à 2-paramètre associé à une table type de mortalité sur la série de valeurs ${}_n p_{15}$ au moyen d'une régression linéaire simple sur tous les groupes d'âge de 15 à 55 ans. Ceci va atténuer l'ampleur de certaines erreurs dans la série. Les valeurs ajustées des probabilités ${}_{35}q_{15}$ et ${}_{45}q_{15}$ sont extraites de cette table de mortalité ajustée. La feuille de calcul peut calculer ces indicateurs de mortalité en utilisant soit le modèle Général des tables types des Nations Unies (UN Population Division 1982) ou l'une des quatre familles des tables types de mortalité de Princeton (Coale, Demeny and Vaughan 1983). La table type de mortalité choisie devrait avoir une structure par âge de la mortalité au sein des âges adultes qui ressemble à celle de la population étudiée. Une autre table de mortalité peut être utilisée comme standard s'il y a de bonnes raisons de penser qu'elle se rapproche davantage de la structure de la mortalité des adultes dans la population étudiée. La table de mortalité la plus appropriée peut ne pas appartenir à la

TABLEAU 27.1 Algorithme pour déterminer la durée d'exposition au risque de mortalité

| Scénario | Description | Règles définissant la durée d'exposition | Période d'exposition vécue par les survivants | Période d'exposition des décédés (si le décès a lieu dans la période) |
|----------|--|--|---|---|
| (a) | Plus âgé que x_b à t_1 | $x_{t1} > x_b$ | 0 | 0 |
| (b) | Age entre x_l et x_b à t_1 . Atteint l'âge x_b dans la période | $x_l < x_{t1} < x_b$ $x_{t1} + (t_2 - t_1) > x_b$ | $x_b - x_{t1}$ | $x_d - x_{t1}$ |
| (c) | Atteint x_l et x_b dans la période | $x_l > x_{t1}$ $x_{t1} + (t_2 - t_1) > x_b$ | $x_b - x_l$ | $x_d - x_l$ |
| (d) | Atteint x_l dans la période. La période se termine avant x_b | $x_l > x_{t1}$ $x_l < x_{t1} + (t_2 - t_1) < x_b$ | $x_{t1} + (t_2 - t_1) > x_l$ | $x_d - x_l$ |
| (e) | N'atteint pas x_l dans la période | $x_{t1} + (t_2 - t_1) < x_l$ | 0 | 0 |

famille de modèles qui saisit le mieux la relation entre la mortalité des enfants et celle des adultes. Si on ne sait rien de la structure par âge de la mortalité à l'âge adulte, l'utilisation du modèle général des Nations Unies ou du modèle Ouest de Princeton est recommandée.

EXEMPLE

Les calculs à la base de la méthode sont illustrés dans le tableau 27.2 à l'aide des données sur la survie des sœurs recueillies auprès des femmes interrogées dans l'enquête sur la mortalité maternelle du Bangladesh (*Maternal Mortality Survey of Bangladesh 2001*), disponible sur le site de MeasureDHS. Cette enquête menée au Bangladesh disposait d'un échantillon exceptionnellement large. Les décès et les durées d'exposition des sœurs des répondantes appartenant aux différents groupes d'âge quinquennaux ont été cumulés sur la période de sept ans précédant l'enquête. Après le calcul des taux et des probabilités de survie de la table de mortalité, ces dernières ont été lissées par ajustement d'une table type de mortalité à l'aide d'un modèle relationnel à 2-paramètres et en utilisant le modèle Sud des tables de mortalité de Princeton comme standard.

La figure 27.2 présente les estimations en vis-à-vis du standard de mortalité, en distinguant les périodes entre 0 à 2 ans et entre 3 à 6 ans avant l'enquête. Les estimations équivalentes pour les frères des répondants sont présentées dans la figure 27.3. Pour ces données particulières, les estimations ajustées de ${}_{35}q_{15}$ et ${}_{45}q_{15}$ sont presque identiques aux estimations calculées directement à partir des données observées soit 90 pour 1000 et 167 pour 1 000 respectivement.

DIAGNOSTICS, ÉVALUATION ET INTERPRÉTATION

Contrôles et validation

Si des histoires de fratries ont été recueillies à la fois auprès des hommes et des femmes dans un recensement ou dans une enquête de grande envergure, il est recommandé de les analyser séparément par sexe des répondants, afin d'évaluer la cohérence de leurs déclarations. La mortalité des individus d'un sexe déterminé, tel que rapporté par leurs frères, devrait être le même que la mortalité des mêmes individus tels que rapportés par leurs sœurs. Le cas contraire laisserait supposer la présence d'un biais important dans les estimations pour un ou pour les deux sexes. La cohérence dans les données collectées ne garantit bien sûr pas leur exactitude, mais des différences statistiquement significatives entre les estima-

tions obtenues à partir des répondants d'un des deux sexes pris séparément signifient qu'au moins un des deux sexes, et peut-être les deux, a donné des réponses inexactes aux questions. Ce type de contrôle sur les résultats ne pourrait pas être effectué sur les données de l'enquête sur la mortalité maternelle au Bangladesh 2001, car les histoires de fratries n'ont été recueillies qu'auprès des femmes dans cette étude.

Interprétation

Les résultats de l'application de la méthode directe de calcul des taux de mortalité à partir des histoires de fratries collectées dans l'enquête sur la mortalité maternelle au Bangladesh en 2001 sont présentés dans les figures 27.2 et 27.3. Ils sont encourageants. Les points relevés ne sont pas erratiques et la courbe ne fléchit pas vers le bas aux âges plus élevés. Il y a une certaine courbure dans la série masculine, en particulier pour la période antérieure, mais ceci couvre toute la série de données de 20 ans jusqu'à 60 ans, et ne concerne pas uniquement les âges plus élevés. Ceci indique probablement que la structure par âge de la mortalité dans cette population diffère de celle de la table-type de mortalité qui a été utilisée.

Les estimations relatives aux deux périodes sont cohérentes l'une par rapport à l'autre pour les deux sexes, suggérant que la mortalité adulte a baissé de façon substantielle dans les années 1990. Ainsi, on estime que la probabilité de décéder entre 15 et 50 ans pour les hommes a chuté de 104 pour 1 000 à 76 pour 1 000 entre la période située de 3 à 6 ans avant l'enquête et la période située de 0 à 2 ans avant l'enquête. La probabilité de décéder entre 35 et 50 ans a chuté chez les femmes de 107 pour mille à 73 pour mille au cours de ces deux mêmes périodes.

Bien que la probabilité globale de décéder entre 15 et 50 ans au Bangladesh soit très similaire pour les hommes et les femmes et ait diminué presque au même rythme pour les deux sexes, les figures 27.2 et 27.3 révèlent que les structures par âge de la mortalité aux âges adultes des hommes et des femmes sont très différentes dans ce pays. La mortalité augmente beaucoup plus fortement avec l'âge pour les hommes que pour les femmes. Le paramètre β de table de mortalité féminine modèle ajustée aux données de 1994 à 2001 est de 0,95 tandis qu'il est de 1,14 pour les hommes. Ainsi, si l'on examine les taux de mortalité pour les groupes d'âge quinquennaux, ils montrent qu'au Bangladesh, dans le groupe d'âge de 15 à 40 ans les femmes ont une mortalité plus élevée que les hommes, alors que dans le groupe 40 à 50 ans les hommes ont une mortalité plus élevée que les femmes.

TABLEAU 27.2 Calcul direct des taux de mortalité des femmes par âge et des quotients de mortalité entre 15 et 50 ans ainsi que 15 et 60 ans, Bangladesh, 1994-2001

| Groupe d'âge x à $x+4$ | Décès des sœurs | Personnes- années d'exposition | Taux de mortalité ${}_5M_x$ | Probabilité de survivre sur cinq ans ${}_5p_x$ | Probabilité cumulée de survivre ${}_{x-10}p_{15}$ | Logits ${}_{x-10}Y_{15}$ | Logits ajustés ${}_{x-10}Y_{15}$ |
|-----------------------------|--------------------|--------------------------------------|-----------------------------------|---|--|-----------------------------|-------------------------------------|
| 15-19 | 350,3 | 211 840,6 | 0,00165 | 0,9918 | 0,9918 | - 2,3956 | -2,4318 |
| 20-24 | 436,8 | 241 208,5 | 0,00181 | 0,9910 | 0,9828 | - 2,0235 | -2,0109 |
| 25-29 | 488,0 | 241 111,4 | 0,00202 | 0,9899 | 0,9729 | - 1,7909 | -1,7758 |
| 30-34 | 455,0 | 210 963,3 | 0,00216 | 0,9893 | 0,9625 | - 1,6225 | -1,5978 |
| 35-39 | 417,5 | 160 378,1 | 0,00260 | 0,9871 | 0,9500 | - 1,4727 | -1,4472 |
| 40-44 | 377,5 | 97 268,6 | 0,00388 | 0,9808 | 0,9318 | - 1,3072 | -1,3020 |
| 45-49 | 242,0 | 50 456,1 | 0,00480 | 0,9763 | 0,9097 | - 1,1550 | -1,1581 |
| 50-54 | 169,4 | 19 621,2 | 0,00863 | 0,9577 | 0,8713 | - 0,9561 | -1,0002 |
| 55-59 | 56,7 | 6 276,6 | 0,00904 | 0,9558 | 0,8328 | - 0,8027 | -0,8286 |
| ${}_{35}q_{15}$ (95% CI) | | | 0,090 | (0,086-0,094) | | ($\alpha = - 0,191$) | 0,090 |
| ${}_{45}q_{15}$ (95% CI) | | | 0,167 | (0,155-0,179) | | ($\beta = 0,949$) | 0,160 |

FIGURE 27.2 Logits des probabilités de survie à partir de 15 ans des femmes, comparés aux logits du modèle Sud des tables-types de mortalité de Princeton, Bangladesh, 1994-2001

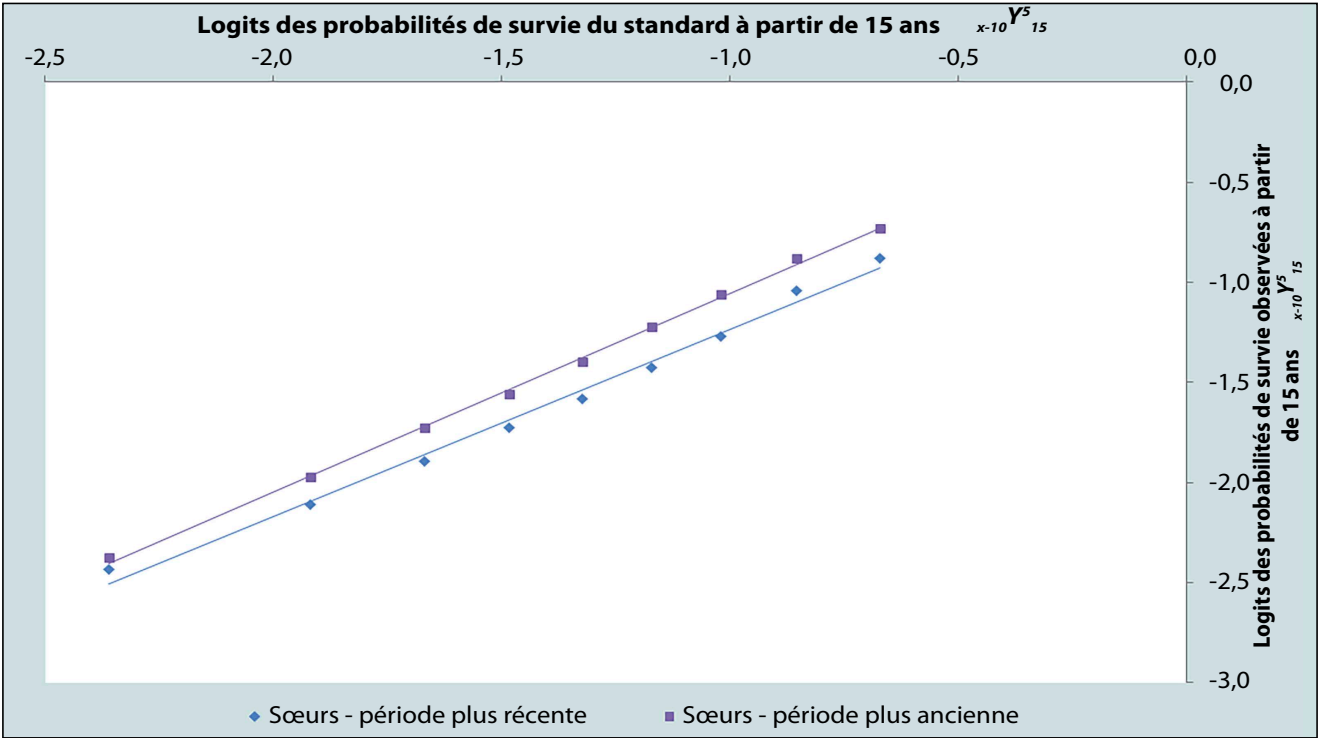
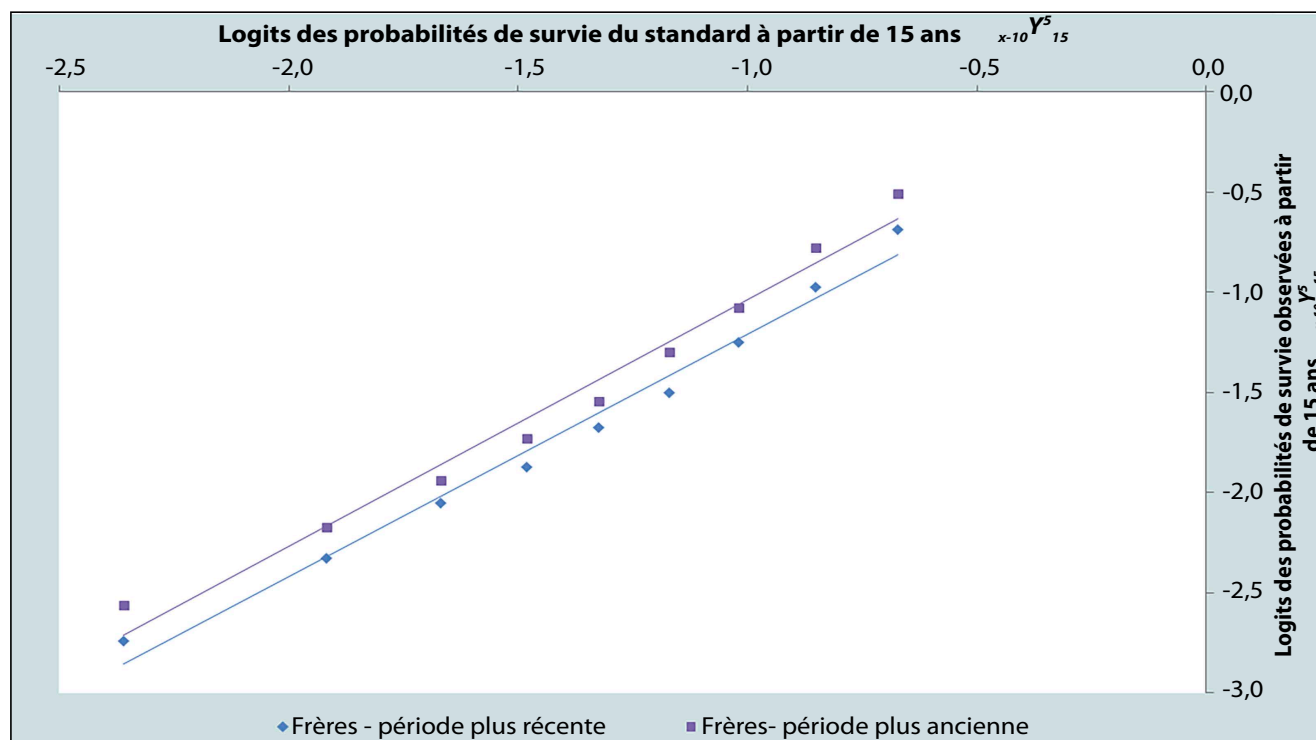


FIGURE 27.3 Logits des probabilités de survie à partir de 15 ans des hommes, comparés aux logits du modèle Sud des tables-types de mortalité de Princeton, Bangladesh, 1994-2001



La régularité interne de chacune des quatre séries d'estimations, la cohérence des estimations pour les deux périodes précédant l'enquête, et la plausibilité de la structure par âge de la mortalité évaluée par rapport aux standards externes, prouve que la méthode a bien fonctionné dans cette enquête du Bangladesh. Le trait le plus surprenant de ces résultats est la très forte baisse de la mortalité des adultes qui est suggérée dans la deuxième moitié des années 1990 au Bangladesh.

Fonctionnement dans les populations subissant une épidémie généralisée de VIH

L'épidémie de VIH pose deux problèmes pour les méthodes d'estimation de la mortalité basée sur la survie des apparentés (UN Population Division 1982). Premièrement, les deux modes de transmission (sexuelle et verticale) du virus créent un biais de sélection significatif dans les données recueillies sur la survie des parents. Deuxièmement, l'incidence de l'infection au VIH est concentrée chez les jeunes adultes. Ainsi, les populations qui font face à une mortalité importante causée par le sida ont des structures de mortalité

par âge très différentes des autres populations et des systèmes de tables-types de mortalité existants.

Un avantage majeur que les méthodes basées sur la survie de la fratrie ont sur les méthodes de survie des parents pour mesurer la mortalité des adultes est qu'elles ne sont pas concernées par les biais de sélection découlant de la transmission directe du virus. Certains biais résiduels subsistent toutefois en raison de la concentration de la mortalité due au sida au sein des fratries. Tous les enfants nés d'une femme après qu'elle ait été infectée sont exposés au risque d'infection par transmission verticale. En outre, le risque d'infection par le VIH a tendance à varier considérablement entre les localités or, les frères et sœurs vivent souvent à proximité les uns des autres. L'effet de la proximité sera cependant relativement faible par rapport aux biais qui affectent les données que les parents ont fournies à propos de leurs enfants ou vice versa. En outre, les estimations directes de la mortalité basée sur la survie des frères et sœurs ont un avantage sur la méthode indirecte d'estimation également basée sur les fratries dans les populations à forte mortalité liée au sida en ce qu'elles mesurent la structure par âge de la mortalité directement – aucune hypothèse ne doit être faite à ce propos.

EXTENSIONS ET VARIANTES DE LA MÉTHODE

Afin d'extraire le maximum d'information utile des données sur les fratries en présence d'erreurs de déclaration et d'échantillonnage, les analystes ont eu recours à des modèles de régression appliqués aux données individuelles (au niveau des frères et sœurs) issues de plusieurs enquêtes et plusieurs pays pour contraindre quelque peu les résultats (par exemple, Obermeyer, Rajaratnam, Park et al 2010 ; Timæus and Jasseh 2004). Par exemple, Timæus et Jasseh intègrent un schéma par âge standard

de mortalité à 2 paramètres dans leur modèle de régression du logarithmique des taux de mortalité afin de lisser les données. Ils font varier le coefficient de régression du standard (qui détermine la structure par âge de la mortalité) entre les pays, mais sans le faire changer au fil du temps. En outre, ils supposent que la vitesse de baisse de la mortalité pour des causes autres que le sida varie selon les pays mais qu'elle suit une tendance log-linéaire dans tous les pays. D'autres analystes ont des contraintes différentes sur leurs estimations.

AUTRES LECTURES ET RÉFÉRENCES

La méthode directe de calcul de la mortalité adulte à partir des histoires de fratries n'est pas abordée dans les manuels classiques sur l'estimation indirecte. Bien que leur rapport soit principalement axé sur la mesure de la mortalité maternelle, Stanton, Abderrahim et Hill (1997) discutent en détail d'un certain nombre de questions importantes relatives à l'estimation de la mortalité toutes causes confondues à partir des histoires de fratries, y compris l'imputation des dates exactes de naissance et de décès et le calcul des durées d'exposition. Les biais liés à la mortalité différentielle selon la taille des fratries et les problèmes associés au fait que certaines fratries sont mentionnées plusieurs fois sont discutés par Gakidou et King (2006), Masquelier (2013), et d'autres.

- Blacker JGC and W Brass. 1983. "Experience of retrospective enquiries to determine vital rates," in Moss, L and H Goldstein (eds). *The Recall Method in Social Surveys*. London: University of London Institute of Education, pp. 48-61.
- Coale AJ, P Demeny and B Vaughan. 1983. *Regional Model Life Tables and Stable Populations*. London: Academic Press.
- Division de la Population des NU. 1984. *Manuel X. Techniques indirectes d'estimation démographique*. New York : Nations Unies, Département des affaires économiques et sociales internationales, ST/ESA/SER.A/81. http://unstats.un.org/unsd/demographic/standmeth/handbooks/Manuel_X-fr.pdf
- Gakidou E and G King. 2006. "Death by survey: estimating adult mortality without selection bias from sibling survival data", *Demography* **43**(3):569-585. doi: <http://dx.doi.org/10.1353/dem.2006.0024>
- Graham W, W Brass and RW Snow. 1989. "Estimating maternal mortality: The sisterhood method", *Studies in Family Planning* **20**(3):125-135. doi: <http://dx.doi.org/10.2307/1966567>
- Hill K and TJ Trussell. 1977. "Further developments in indirect mortality estimation", *Population Studies* **31**(2):313-334. doi: <http://dx.doi.org/10.2307/2173920>

- Masquelier B. 2013. "Adult mortality from sibling survival data: A reappraisal of selection biases", *Demography*, **50**(1):207-228. doi: <http://dx.doi.org/10.1007/s13524-012-0149-1>
- Obermeyer Z, JK Rajaratnam, CH Park, E Gakidou et al. 2010. "Measuring adult mortality using sibling survival: a new analytical method and new results for 44 countries, 1974-2006", *PLoS Medicine* **7**(4):e1000260. doi: <http://dx.doi.org/10.1371/journal.pmed.1000260>
- Rutenberg N and JM Sullivan. 1991. "Direct and indirect estimates of maternal mortality from the sisterhood method," Paper presented at Demographic and Health Surveys World Conference, August 5-7, 1991, Washington, D.C. Columbia. Macro International. Vol. 3:1669-1696.
- Stanton C, N Abderrahim and K Hill. 1997. *DHS Maternal Mortality Indicators: An Assessment of Data Quality and Implications for Data Use*. Calverton: Macro International.
- Timæus IM and M Jasseh. 2004. "Adult mortality in Sub-Saharan Africa: evidence from Demographic and Health Surveys", *Demography* **41**(4):757-772. doi: <http://dx.doi.org/10.1353/dem.2004.0037>
- Trussell J and G Rodriguez. 1990. "A note on the sisterhood estimator of maternal mortality", *Studies in Family Planning* **21**(6):344-346. doi: <http://dx.doi.org/10.2307/1966923>
- UN Population Division. 1982. *Model Life Tables for Developing Countries*. New York: United Nations, Department of Economic and Social Affairs, ST/ESA/SER.A/77. http://www.un.org/esa/population/techcoop/DemMod/model_lifetabs/model_lifetabs.html
- UN Population Division. 1983. *Manual X: Indirect Techniques for Demographic Estimation*. New York: United Nations, Department of Economic and Social Affairs, ST/ESA/SER.A/81. <http://www.un.org/esa/population/techcoop/DemEst/manual10/manual10.html>
- Zaba B. 1986. *Measurement of Emigration using Indirect Techniques: Manual for the Collection and Analysis of Data on Residence of Relatives*. Liège: Ordina.
- Zaba B and PH David. 1996. "Fertility and the distribution of child mortality risk among women", *Population Studies* **50**(2):263-278. doi: <http://dx.doi.org/10.1080/0032472031000149346>

La mortalité maternelle

Chapitre 28 INTRODUCTION À L'ANALYSE DE LA MORTALITÉ MATERNELLE

Kenneth Hill

CONTEXTE

La mortalité maternelle est depuis longtemps un problème sérieux pour les services nationaux de santé, mais son importance s'est encore accrue depuis une vingtaine d'années avec la définition d'objectifs quantitatifs. Pendant cette période, la communauté internationale a régulièrement énoncé des objectifs de réduction de la mortalité maternelle en termes de Rapport de Mortalité Maternelle (RMM), soit le nombre de décès maternels pour 100 000 naissances vivantes. En 1990, le Sommet mondial pour les Enfants a proposé comme objectif la réduction de moitié du RMM entre 1990 et 2000. En 1994, la Conférence internationale sur la Population et le Développement (ICPD) a confirmé cette résolution et y a ajouté l'objectif supplémentaire, à plus long terme, de réduire encore le RMM de moitié entre 2000 et 2015. En 2000, le Sommet du Millénaire a fait de la résolution de l'ICPD le cinquième Objectif du Millénaire pour le Développement (réduction de la mortalité maternelle). L'ambition était ainsi de réduire le RMM de trois quarts entre 1990 et 2015. Le rapport 2011 de la Commission d'information et de responsabilisation pour la santé de la femme et de l'enfant, instituée par le Secrétariat général des Nations Unies, a réaffirmé qu'il était important de faire régulièrement le point sur le RMM, retenu parmi les onze indicateurs de la santé de la mère, du nouveau-né et de l'enfant. Il est donc clair que la mesure de la mortalité maternelle bénéficie d'une très haute priorité. Ce chapitre en présente les principales modalités.

DÉFINITION

La 10^e Révision de la Classification internationale des maladies (CIM-10) définit ainsi le décès maternel : « le décès d'une femme survenu au cours de la grossesse ou dans un délai de 42 jours après sa terminaison, quelles qu'en soient

la durée ou la localisation, pour une cause quelconque déterminée ou aggravée par la grossesse ou les soins qu'elle a motivés, mais ni accidentelle, ni fortuite » La mesure des décès maternels suppose donc la détermination de leurs causes, question qui n'est pas évoquée ailleurs dans ce manuel. Les décès maternels se répartissent en décès dus à une cause obstétrique directe (principalement l'hémorragie, l'arrêt de progression du travail, l'éclampsie, la septicémie et les complications liées à l'avortement) et décès dus à une cause obstétrique indirecte (décès associés à la grossesse chez des femmes présentant un problème de santé, préexistant ou apparu récemment, aggravé par la grossesse ou par l'accouchement).

La mesure de la mortalité maternelle constitue un problème majeur dans les pays où l'enregistrement des naissances et des décès n'est pas suffisamment complet (Graham, Ahmed, Stanton *et al.* 2008), non seulement parce que certains décès ne sont pas enregistrés, mais aussi parce qu'on a besoin d'établir la cause du décès (voir, par exemple, Mathers, Fat, Inoue *et al.* 2005). La cause d'un décès est déterminée de préférence par un médecin présent au moment de la mort ou peu après, mais bien des décès surviennent en l'absence d'un médecin. De plus, même certifié par un médecin, un décès survenu hors de la salle d'accouchement peut être erronément attribué à une cause non-maternelle. Certains progrès ont été obtenus ces dernières années avec le développement et la mise en œuvre des techniques d'autopsie verbale, par lesquelles on demande aux membres de la famille de décrire les signes et les symptômes observés autour du décès, mais il y a encore beaucoup de controverses sur le bon fonctionnement de ces méthodes (Chandramohan, Rodrigues, Maude *et al.* 1998). La description des outils de l'autopsie verbale et de son analyse dépasse le cadre de ce manuel.

En partie à cause de la difficulté d'identifier correctement les décès maternels, la CIM-10 définit aussi le « décès lié à la grossesse » comme tout décès survenu au cours de la grossesse, lors de l'accouchement ou dans les 42 jours suivant la fin de la grossesse, quelle qu'en soit la cause. La rubrique « décès liés à la grossesse » inclut donc tous les décès maternels plus les décès dus à des causes accidentelles ou accessoires, qui sont exclus de la rubrique « décès maternels ». La catégorie « décès liés à la grossesse » présente l'avantage d'être plus facile à mettre en œuvre, car elle requiert seulement que l'on situe le décès par rapport à la grossesse sans devoir en établir exactement la cause. L'inconvénient est que la mesure de la mortalité liée à la grossesse est souvent considérée à tort comme une mesure de la mortalité maternelle, et que toute évolution de la mortalité liée à la grossesse qui n'est pas due aux causes spécifiquement maternelles limite l'intérêt de cet indicateur pour observer l'impact des politiques de réduction de la mortalité maternelle. En général, dans les enquêtes démographiques, on enregistre les décès liés à la grossesse et on évite de devoir en déterminer les causes.

On discute activement, et il n'y a pas consensus, sur le rapport existant entre le nombre déclaré de décès liés à la grossesse et le nombre, inconnu, de véritables décès maternels. Du fait même de leurs définitions, le nombre réel de décès liés à la grossesse doit être égal ou supérieur au nombre réel de décès maternels, puisque ces derniers sont tous associés à la grossesse, tandis que les décès liés à la grossesse ne sont pas tous des décès maternels. Mais, en pratique, les choses sont moins nettes à cause d'éventuelles erreurs de déclaration. Certains estiment que les décès liés à la grossesse déclarés sont plus nombreux que les décès maternels réels (Garenne, McCaa et Nacro 2008 ; Stecklov, 1995), tandis que d'autres prétendent que les décès liés à la grossesse sont probablement sous-évalués parce que, par exemple, le déclarant peut ne pas savoir que la femme décédée était enceinte au moment du décès (Shahidullah, 1995 ; Wilmoth, 2009). La relation pourrait donc aller dans un sens comme dans l'autre. Dans ce manuel, nous ne prendrons pas formellement parti pour l'une ou l'autre thèse, mais nous soulignons plutôt la nécessité d'employer la dénomination correcte des indicateurs. Ainsi, un indice basé sur l'enregistrement des décès liés à la grossesse doit être présenté comme une mesure de la mortalité liée à la grossesse, tandis qu'un indice fondé sur ce que l'on estime être de véritables décès maternels (grâce à une autopsie verbale, par exemple) doit être présenté comme une mesure de la mortalité maternelle.

INDICES DE BASE

DE LA MORTALITÉ MATERNELLE

Il existe deux indices courants de mortalité maternelle (et corrélativement deux indices de mortalité liée à la grossesse). Ce sont le Rapport de Mortalité Maternelle (RMM ; *en anglais : Maternal Mortality Ratio*), couramment appelé à tort en français taux de mortalité maternelle, qui sert à définir l'OMD-5 – et le Taux de Mortalité Maternelle (TMM ; *en anglais : Maternal Mortality Rate*). On rencontre assez souvent deux autres indices : la part des décès maternels parmi les décès de femmes en âge de procréer (PMDF ; *en anglais : proportion of maternal deaths among deaths of women of reproductive age*) et le risque de décès maternel au cours de la vie (LTR ; *en anglais : lifetime risk*). On utilise surtout ce dernier indice à l'appui d'un plaidoyer.

Le Rapport de Mortalité Maternelle

Le RMM est, sur une période déterminée, le nombre de décès maternels pour 100 000 naissances vivantes. Soulignons l'utilisation comme dénominateur du nombre de naissances vivantes et non du nombre de grossesses. Le RMM mesure donc essentiellement un risque obstétrique, en gros le risque de décès pour 100 000 événements à risque.

Le Taux de Mortalité Maternelle

Le TMM est un taux de mortalité par cause. C'est le nombre de décès maternels enregistrés sur une période déterminée pour 1 000 personnes-années vécues par la population féminine en âge de procréer (habituellement de 15 à 49 ans).

Le RMM et le TMM ont le même numérateur et la relation qui les lie est simple :

$$\begin{aligned} RMM &= \frac{MD}{LB} \times 100\,000 = \frac{MD}{FPRA} \times 1000 \times \frac{FPRA}{LB} \times 100 \\ &= TMM \times \frac{100\,000}{GFR} \end{aligned}$$

où, sur une période donnée, MD (*maternal deaths*) représente le nombre de décès maternels, LB (*live births*) le nombre de naissances vivantes, FPRA (*female population of reproductive age*) le nombre de personnes-années vécues par la population féminine en âge de procréer, et GFR (*general fertility rate*) le taux de fécondité générale exprimé pour 1 000 femmes en âge de procréer.

Part des décès maternels parmi les décès de femmes en âge de procréer

L'indice PMDF est défini par MD/FDRA, où FDRA est le nombre de décès de femmes en âge de procréer. On l'utilise essentiellement dans des exercices de modélisation (voir, par exemple, Hill, Thomas, AbouZahr *et al.* 2007 ; Hogan, Foreman, Naghavi *et al.* 2010 ; Wilmoth, Zureick, Mizoguchi *et al.* 2010 ; et Wilmoth, Mizoguchi, Oestergaard *et al.* 2012), mais il peut également servir à l'évaluation de la qualité des données (voir plus loin).

Le risque sur la vie entière

L'indice LTR est habituellement défini comme le risque de décès maternel depuis l'âge de 15 ans jusqu'à la fin de la vie. Wilmoth (2009) suggère de le calculer sur la base de 1 000 femmes atteignant l'âge de 15 ans :

$$LTR = \frac{(T_{15} - T_{50})}{l_{15}} \times TMM$$

où T_{15} et T_{50} sont les nombres de personnes-années vécues au-delà des âges de 15 ans et 50 ans respectivement, et l_{15} est le nombre de survivantes au 15^e anniversaire dans une table de mortalité adaptée à la population en question.

À chacun de ces quatre indices correspond un indice homologue en termes de décès liés à la grossesse, calculé en remplaçant le nombre de décès maternels par le nombre de décès liés à la grossesse.

SOURCES DES DONNÉES

Outre l'exploitation de l'état civil, il existe deux modes largement répandus de collecte des données nécessaires pour mesurer la mortalité liée à la grossesse : l'histoire complète des frères et sœurs (*en anglais : full sibling history, FSH*) et un recensement ou une grande enquête auprès des ménages recueillant des données sur les décès survenus récemment dans le ménage (*en anglais : household survey collecting data on recent household deaths, HSHD*). L'histoire résumée des frères et sœurs est aujourd'hui peu utilisée (Graham, Brass et Snow 1989), en partie parce qu'elle fournit des estimations moyennes sur de très grands intervalles de temps.

L'histoire complète des frères et sœurs

L'histoire complète des frères et sœurs (FSH) suppose la collecte de données complexes et détaillées, qui nécessite une solide formation et une supervision étroite du personnel de terrain. C'est pourquoi cette méthodologie n'est pas

adaptée au cas d'un recensement. La FSH a été fréquemment introduite dans les EDS sous la forme du « Module Mortalité Maternelle » à partir de 1991, ainsi que dans quelques autres enquêtes auprès des ménages. Elle consiste à recueillir des informations auprès de certaines personnes éligibles. Dans la plupart des EDS, les femmes éligibles pour l'histoire génésique sont également éligibles pour la FSH, mais, dans certaines enquêtes, on a aussi considéré des hommes comme éligibles pour la FSH. Il s'agit de recueillir des données sur tous les frères et sœurs nés de la même mère que la personne interrogée. La FSH peut donc être vue comme l'histoire génésique complète de la mère de la personne interrogée, mais en excluant cette personne elle-même. Dans les EDS, on demande, pour chaque frère ou sœur : le nom, le sexe, s'il/elle est vivant/e, si oui son âge en années, sinon à combien d'années remonte son décès et son âge au moment du décès. En ce qui concerne les sœurs décédées alors qu'elles étaient en âge de procréer, des questions complémentaires permettent de savoir si elles sont décédées (i) enceintes, (ii) lors d'un accouchement ou (iii) dans les 42 jours (ou 2 mois) de la fin d'une grossesse.

De ce qui précède, il résulte clairement que les indices calculés à partir d'une FSH portent sur la mortalité liée à la grossesse et non sur la mortalité maternelle. L'histoire des frères et sœurs ne se prête pas à l'utilisation d'une autopsie verbale (qui s'imposerait pour la mesure de la mortalité maternelle), car le décès d'une sœur en âge de procréer peut s'être produit dans un ménage autre que celui de la personne interrogée, laquelle n'aura donc guère de connaissance directe des signes et symptômes entourant le décès. Il n'est généralement pas possible d'essayer d'identifier le ménage dans lequel a eu lieu le décès pour procéder à une autopsie verbale avec l'aide d'un membre de ce ménage.

La FSH fournit des informations sur les décès liés à la grossesse et sur l'exposition des femmes à ce risque ; elle constitue donc une base pour l'estimation des taux de mortalité liée à la grossesse. Si on veut calculer des rapports de mortalité liée à la grossesse, on doit disposer aussi de données sur les naissances vivantes. Une EDS classique recueillant aussi bien une histoire génésique complète qu'une FSH, cela n'est généralement pas un problème.

L'histoire résumée des frères et sœurs

Ce sont Graham, Brass et Snow (1989) qui ont été les premiers à proposer l'utilisation de données sur la survie des frères et sœurs pour estimer la mortalité maternelle. Ils suggé-

raient d'employer une histoire résumée des frères et sœurs. Dans le cadre de cette histoire résumée, on recueille, séparément pour chaque sexe, des données sur les nombres de frères et sœurs qu'a eus la personne interrogée, les nombres de ceux qui ont survécu jusqu'à 15 ans (ou jusqu'au premier mariage) et – pour les sœurs décédées au-delà de 15 ans – on demande si elles sont décédées enceintes, ou lors d'un accouchement ou dans les 42 jours de la fin d'une grossesse. Mais cette méthode n'est pas conseillée. Il peut y avoir jusqu'à trente ans d'écart d'âge entre la personne interrogée et une de ses sœurs, ce qui fait que les décès de sœurs peuvent s'étaler sur une très longue période antérieure à l'enquête. Les dates de référence des estimations de mortalité maternelle obtenues par le biais de l'histoire résumée des frères et sœurs se situent alors assez loin dans le passé (en moyenne pas moins d'une douzaine d'années avant l'enquête), et cela en limite l'intérêt pratique de la méthode. Nous n'irons donc pas plus loin dans la description de cette méthode.

Recensement ou grande enquête recueillant des données sur les décès survenus récemment dans le ménage

Dans les pays qui ne disposent pas d'un état civil complet, les recensements des années 1970 et 1980 ont souvent recueilli des données sur les décès survenus récemment dans le ménage, en général dans les 12 derniers mois. Les doutes sur la qualité de ces données en ont réduit l'utilisation dans les recensements des années 1990 et 2000, mais le regain d'intérêt pour la mortalité adulte, et en particulier la mortalité maternelle, a entraîné un net retour de cette pratique depuis 2010. En général, il s'agit de demander si un membre habituel du ménage est décédé au cours des 12 derniers mois (parfois, la question porte sur une période différente en faisant référence à une date fixe ou à un événement mémorable). En cas de réponse positive, on enregistre le nom, le sexe et l'âge au décès de la personne décédée. S'il s'agit d'une femme en âge de procréer, on pose une ou plusieurs questions supplémentaires sur le moment du décès en rapport avec une éventuelle grossesse : la femme est-elle décédée enceinte, ou lors d'un accouchement ou dans les 6 semaines (parfois 2 mois) de la fin d'une grossesse ? Cette méthodologie est présentée par Stanton, Hobcraft, Hill *et al.* (2001), et les résultats obtenus sont examinés par Hill, Queiroz, Stanton *et al.* (2007) et Hill, Queiroz, Wong *et al.* (2009).

Telles qu'on les utilise habituellement, ces questions sur les décès récents permettent d'identifier les décès liés à la grossesse. Cependant, il y a eu quelques tentatives d'appro-

fondissement des renseignements sur les décès de femmes en âge de procréer (ou sur un échantillon de ces décès) au moyen d'une autopsie verbale, afin d'identifier les véritables décès maternels. C'est ce qui a été fait dans un certain nombre de très grandes enquêtes auprès des ménages, par exemple, dans la Bangladesh Maternal Morbidity and Mortality Survey 2001 (Hill, El-Arifeen, Koenig *et al.* 2006), ainsi qu'à la suite d'au moins deux recensements (en Iran en 1986 et au Mozambique en 2007).

Un recensement ou une grande enquête auprès des ménages qui recueille des données sur les décès survenus récemment dans le ménage commence toujours par dresser un tableau des membres du ménage par sexe et âge. Ce tableau fournit les dénominateurs des Taux de Mortalité liée à la Grossesse. On a besoin de données supplémentaires sur la fécondité récente pour calculer les Rapports de Mortalité liée à la Grossesse. On obtient habituellement ces informations en demandant aux femmes en âge de procréer si elles ont eu une naissance vivante au cours de l'année précédant l'enquête, ou en leur demandant la date de leur dernière naissance vivante. Il faut également recueillir des données sur la fécondité totale des femmes pour pouvoir évaluer, et éventuellement corriger, les indices de fécondité (voir le chapitre 7 sur l'estimation de la fécondité au moyen d'un modèle relationnel de Gompertz).

Données recueillies auprès des établissements de santé

Une part importante du budget d'une enquête est constituée par les frais engagés pour amener l'enquêteur jusqu'au ménage qu'il doit interroger. La plus grande part de ces frais peut être éliminée en profitant d'occasions où ce sont les personnes à interroger qui viennent jusqu'à l'enquêteur, par exemple quand une femme vient à l'hôpital pour accoucher. De plus, il est probable que les établissements de santé enregistrent les naissances et les décès survenus sur place, ainsi que les causes des décès, dans le cadre de leur système habituel d'information et de gestion. Mais ces données souffrent d'un biais de sélection : on ne peut jamais être sûr que les femmes qui accouchent dans un établissement de santé constituent un échantillon représentatif de l'ensemble des mères. Afin d'améliorer la couverture de la collecte, on mène certaines expériences pour voir si le personnel des dispensaires et d'autres agents de santé des communautés locales pourraient recueillir de bonnes données sur les naissances et les décès. Une telle approche s'apparente à un système d'état civil par échantillonnage.

Une stratégie d'analyse de ces données hospitalières a été proposée, mais pas encore mise en œuvre. Partant de l'hypothèse que les femmes qui accouchent à l'hôpital (ou y viennent en consultation pour une raison quelconque liée à leur grossesse) constituent un échantillon biaisé de l'ensemble des mères (on ne sait pas quel est le sens du biais : surreprésentation des femmes qui ont beaucoup de problèmes de grossesse ou de celles qui en ont peu), si on pouvait évaluer correctement les probabilités de sélection, on pourrait corriger le biais des chiffres recueillis. Par exemple,

on pourrait demander aux femmes qui viennent à l'hôpital leur âge, le nombre d'enfants qu'elles ont eus, le nombre de ceux qui sont encore vivants, plus quelques questions complémentaires sur leur situation socio-économique. On pourrait alors relier, dans un modèle, le nombre d'enfants et le nombre d'enfants survivants à la structure socio-économique de la population totale, fournie, par exemple, par un recensement. À notre connaissance, cette approche n'a jamais été testée.

RÉFÉRENCES

- Chandramohan D, LC Rodrigues, GH Maude and RJ Hayes. 1998. "The validity of verbal autopsies for assessing the causes of institutional maternal death", *Studies in Family Planning* **29**(4):414-422. doi: <http://dx.doi.org/10.2307/172253>
- Garenne M, R McCaa and K Nacro. 2008. "Maternal mortality in South Africa in 2001: From demographic census to epidemiological investigation", *Population Health Metrics* **6**:4. doi: <http://dx.doi.org/10.1186/1478-7954-6-4>
- Graham W, S Ahmed, C Stanton, C Abou-Zahr and O Campbell. 2008. "Measuring maternal mortality: An overview of opportunities and options for developing countries", *BMC Medicine* **6**:12. doi: <http://dx.doi.org/10.1186/1741-7015-6-12>
- Graham W, W Brass and RW Snow. 1989. "Estimating maternal mortality: The sisterhood method", *Studies in Family Planning* **20**(3):125-135. doi: <http://dx.doi.org/doi:10.2307/1966567>
- Hill K, S El-Arifeen, M Koenig, A Al-Sabir, K Jamil and H Raggars. 2006. "How should we measure maternal mortality in the developing world? A comparison of household deaths and sibling history approaches", *Bulletin of the World Health Organization* **84**(3):173-180. doi: <http://dx.doi.org/10.2471/BLT.05.027714>
- Hill K, BL Queiroz, C Stanton and C AbouZahr. 2007. "Measuring maternal mortality via the population census: Experience from Africa," Paper presented at 5th African Population Conference. Arusha, Tanzania, 10-14 December 2007.
- Hill K, BL Queiroz, L Wong, J Plata *et al.* 2009. "Estimating pregnancy-related mortality from census data: Experience in Latin America", *Bulletin of the World Health Organization* **87**(4):288-295. doi: <http://dx.doi.org/10.2471/BLT.08.052233>
- Hill K, K Thomas, C AbouZahr, N Walker *et al.* 2007. "Estimates of maternal mortality worldwide between 1990 and 2005: An assessment of available data", *The Lancet* **370**(9595):1311-1319. doi: [http://dx.doi.org/10.1016/S0140-6736\(07\)61572-4](http://dx.doi.org/10.1016/S0140-6736(07)61572-4)
- Hogan MC, KJ Foreman, M Naghavi, SY Ahn *et al.* 2010. "Maternal mortality for 181 countries, 1980-2008: A systematic analysis of progress towards Millennium Development Goal 5", *The Lancet* **375**(9726):1609-1623. doi: [http://dx.doi.org/10.1016/S0140-6736\(10\)60518-1](http://dx.doi.org/10.1016/S0140-6736(10)60518-1)
- Mathers CD, DM Fat, M Inoue, C Rao and AD Lopez. 2005. "Counting the dead and what they died from: An assessment of the global status of cause of death data", *Bulletin of the World Health Organization* **83**(3):171-177.
- Shahidullah M. 1995. "The sisterhood method of estimating maternal mortality: The Matlab experience", *Studies in Family Planning* **26**(2):101-106. doi: <http://dx.doi.org/10.2307/2137935>
- Stanton C, J Hobcraft, K Hill, N Kodjogbé *et al.* 2001. "Every death counts: Measurement of maternal mortality via a census", *Bulletin of the World Health Organization* **79**(7):657-664.
- Stecklov G. 1995. "Maternal mortality estimation: Separating pregnancy-related and non-pregnancy-related risks", *Studies in Family Planning* **26**(1):33-38. doi: <http://dx.doi.org/10.2307/2138049>
- Wilmoth J, S Zureick, N Mizoguchi, M Inoue and M Oestergaard. 2010. *Levels and Trends of Maternal Mortality in the World: The Development of New Estimates by the United Nations*. Geneva: WHO, UNICEF, UNFPA and the World Bank. http://www.who.int/reproductivehealth/publications/monitoring/MMR_technical_report.pdf
- Wilmoth JR. 2009. "The lifetime risk of maternal mortality: Concept and measurement", *Bulletin of the World Health Organization* **87**(4):256-262. doi: <http://dx.doi.org/10.2471/BLT.07.048280>
- Wilmoth JR, N Mizoguchi, MZ Oestergaard, L Say *et al.* 2012. "A new method for deriving global estimates of maternal mortality", *Statistics, Politics, and Policy* **3**(2):Article 3. doi: <http://dx.doi.org/10.1515/2151-7509.1038>

Chapitre 29 ESTIMATION DE LA MORTALITÉ LIÉE À LA GROSSESSE À PARTIR DE LA SURVIE DES FRÈRES ET SŒURS

Kenneth Hill

DESCRIPTION DE LA MÉTHODE

L'évaluation et l'utilisation des données de l'histoire complète des frères et sœurs (*en anglais : full sibling history, FSH*) pour estimer la mortalité adulte dans son ensemble sont décrites dans le chapitre 27. L'histoire complète des frères et sœurs est analogue à une histoire génésique complète : on questionne la personne interrogée (habituellement une femme en âge de procréer) à propos de chacun de ses frères et sœurs nés de la même mère. Pour les frères et sœurs survivants, on demande le sexe et l'âge en années révolues ; pour les frères et sœurs décédés, on enregistre le sexe, l'âge au moment du décès en années révolues et l'année du décès. En ajoutant les données de la femme interrogée, on obtient l'histoire génésique de la mère. Comme l'histoire génésique complète, l'histoire des frères et sœurs permet de situer dans le temps les événements (les décès) et les durées d'exposition au risque, et donc de calculer des taux transversaux de mortalité par âge. Pour estimer la mortalité liée à la grossesse, on recueille des informations supplémentaires sur les décès de sœurs en âge de procréer, en demandant si elles sont décédées alors qu'elles étaient enceintes, lors d'un accouchement ou dans les 42 jours (parfois 2 mois) de la fin d'une grossesse. Ce chapitre est consacré à l'analyse des données sur les sœurs en âge de procréer.

Nous allons aborder ici un important problème touchant les histoires complètes des frères et sœurs. Dans les EDS, l'histoire complète des frères et sœurs consiste à demander à la personne interrogée l'histoire génésique de sa propre mère en s'en excluant elle-même. Par conséquent, on peut se trouver face à plusieurs réponses relatives à la même personne. Par exemple, si deux filles de la même mère sont interrogées dans le même ménage, on obtiendra deux descriptions des autres membres de leur fratrie. Les EDS déterminent les événements et les durées d'exposition uniquement sur la base des frères et sœurs déclara-

és, sans prendre en compte la durée d'exposition de la femme interrogée elle-même (survivante). De plus, on affecte aux événements et durées d'exposition des frères et sœurs le coefficient de pondération de la personne interrogée, sans tenir compte du nombre de membres survivants de la fratrie susceptibles d'être interrogés. Trussell et Rodriguez (1990) montrent que, s'il n'y a aucune corrélation entre les risques de décès respectifs des frères et sœurs, ce calcul fournit une estimation sans biais de la mortalité générale. Gakidou et King (2006) estiment que la fratrie devrait inclure la personne interrogée et que les données de ses membres devraient en outre être pondérées par leur probabilité d'être déclarés – c'est-à-dire l'inverse du nombre de personnes interrogées potentielles dans la fratrie. Ils pensent également qu'un ajustement serait nécessaire pour prendre en compte les fratries qui ne sont pas déclarées faute de membre survivant. En analysant les FSH de plusieurs EDS, Obermeyer, Rajaratnam, Park *et al.* (2010) estiment que le fait de ne pas procéder à un ajustement en fonction de la probabilité d'être déclaré peut entraîner pour les indices de mortalité générale une sous-estimation de quelque 20 %. Mais Masquelier (2013) réplique que l'analyse de Obermeyer *et al.*, prenant en compte tous les frères et sœurs survivants et pas seulement ceux qui sont susceptibles d'être interrogés, exagère l'ampleur du biais éventuel. Il recommande d'utiliser le mode de calcul des EDS, pour des raisons qui seront exposées plus loin (voir aussi l'analyse de l'histoire des frères et sœurs – chapitre 27).

DONNÉES REQUISES ET HYPOTHÈSES

Hypothèses importantes

- Il n'y a aucune corrélation entre le risque de décès d'une femme et le nombre de ses frères et sœurs.
- Il n'y a pas d'effet de sélection dû à la migration.

Tabulations nécessaires

- Répartition des femmes par groupes quinquennaux d'âge, d'après le questionnaire-ménage.
- Répartition des décès de sœurs par ancienneté avant l'enquête (habituellement de 0 à 6 ans), par groupes quinquennaux d'âge, et selon que ces sœurs sont décédées alors qu'elles étaient enceintes, ou lors d'un accouchement ou dans les 42 jours (ou 2 mois) de la fin d'une grossesse.
- Nombres de personnes-années d'exposition vécues par les sœurs, selon l'ancienneté du décès et le groupe quinquennal d'âge de la femme interrogée.
- Taux de fécondité par âge et taux global de fécondité de la période considérée avant l'enquête.

Nous supposons ici que l'on suit l'approche EDS. La construction de tableaux synthétiques à partir des EDS déborde du cadre de ce manuel. Un programme standard de calcul de ces tabulations à partir des données de base existe pour le logiciel CPro. Il est indispensable de décider, dès le début de l'analyse, de l'horizon temporel à utiliser. La plupart des EDS fournissent les tabulations de base des événements et des durées d'exposition pour les sept années antérieures à l'enquête (de 0 à 6 ans), mais cette période est parfois de cinq ans et parfois de dix ans. Pour choisir une période plutôt qu'une autre, on examine, entre autres, les erreurs d'échantillonnage : avec un petit échantillon et une mortalité adulte assez basse, il vaut mieux opter pour une période rétrospective de dix ans, afin d'éviter une trop grande erreur d'échantillonnage, mais si l'échantillon est grand et la mortalité élevée, on peut se contenter de cinq ans. Cet aspect de la méthode est détaillé dans la section consacrée à l'interprétation des résultats.

Travaux préparatoires et recherches préliminaires

L'évaluation de la qualité des données d'une histoire complète des frères et sœurs et des taux de fécondité par âge récents est décrite par ailleurs. Le seul problème d'évaluation de la qualité des données qui soit propre à l'estimation de la mortalité liée à la grossesse concerne les renseignements sur la part des décès liés à la grossesse parmi les décès de femmes en âge de procréer (*en anglais : proportion of deaths of women of reproductive age that are pregnancy-related, PPRD*) et sur les proportions de décès liés à la grossesse survenus respectivement pendant la grossesse, lors de l'accouchement et dans les 42 jours (ou 2 mois) de la fin d'une grossesse.

Il n'existe pas de méthode spécifique pour procéder à de telles évaluations. Mais la structure par âge des proportions de décès liés à la grossesse devrait être similaire à celle des

taux de fécondité par âge, car l'événement à risque est la naissance. On pense généralement que le risque de décès lié à la grossesse est un peu plus élevé aux extrémités de la tranche d'âge 15-49 ans qu'aux âges intermédiaires, ce qui peut « gonfler » quelque peu les queues de distribution des proportions de décès liés à la grossesse par comparaison avec celles de la distribution des taux de fécondité par âge.

Mise en garde

On pense généralement que l'histoire des frères et sœurs a tendance à sous-estimer la mortalité, en particulier les décès survenus dans un passé éloigné. On ne devrait donc pas essayer d'interpréter les tendances de la mortalité liée à la grossesse sur la base d'une seule série de données. Et ces tentatives menées sur plusieurs séries de données devraient également tenir compte des grands intervalles de confiance qui encadrent les estimations, même pour une période rétrospective de sept ans.

APPLICATION DE LA MÉTHODE

Étape 1 : Élaboration des tableaux de données

Comme nous l'avons signalé plus haut, un logiciel facile d'accès permet de construire les tabulations nécessaires relatives aux décès des sœurs, à leurs durées d'exposition au risque et aux décès liés à la grossesse. On a également besoin d'estimations des taux de fécondité par âge. (Si les données concernant les sœurs proviennent d'une EDS, la méthode d'estimation directe des taux de fécondité à partir de ces données est décrite ailleurs dans ce manuel.) Enfin, une estimation de la structure par âge de la population féminine enquêtée est nécessaire.

Définitions de quelques symboles :

${}_5D_x^s$ – nombre de sœurs déclarées décédées entre les âges x et $x + 5$

${}_5PY_x^s$ – nombre de personnes-années vécues par les sœurs entre les âges x et $x + 5$

${}_5PRD_x^s$ – nombre de décès de sœurs liés à la grossesse survenus entre les âges x et $x + 5$

${}_5f_x$ – taux de fécondité des femmes âgées de x à $x + 5$

${}_5N_x^f$ – nombre de femmes enquêtées âgées de x à $x + 5$.

Des tabulations de chacune des quantités ci-dessus sont nécessaires pour l'application de la méthode.

Étape 2 : Calcul des indices de mortalité

On calcule un taux de mortalité par groupe d'âge en divisant le nombre de décès de sœurs survenus dans ce groupe d'âge

par le nombre de personnes-années vécues par celles-ci dans cette tranche d'âge :

$${}_5M_x = \frac{{}_5D_x^s}{{}_5PY_x^s}. \tag{1}$$

Le taux de mortalité liée à la grossesse (*en anglais : pregnancy-related mortality rate, PRMRate*), par groupe d'âge, est :

$${}_5PRMRate_x = \frac{{}_5PRD_x^s}{{}_5PY_x^s}. \tag{2}$$

La part des décès liés à la grossesse parmi l'ensemble des décès de sœurs, par groupe d'âge, est :

$${}_5PPRD_x = \frac{{}_5PRD_x^s}{{}_5D_x^s} \tag{3}$$

et le rapport de mortalité liée à la grossesse (*en anglais : pregnancy-related mortality ratio : PRMRatio*), par groupe d'âge, est :

$${}_5PRMRatio_x = 100\,000 \cdot \left(\frac{{}_5PRD_x^s}{{}_5PY_x^s \cdot {}_5f_x} \right). \tag{4}$$

Étape 3 : Estimation des indices au niveau de la population

On calcule les estimations du taux de mortalité liée à la grossesse et de la proportion de décès liés à la grossesse parmi les décès de femmes en âge de procréer (15-49 ans) en pondérant les indices calculés ci-dessus par les nombres de femmes âgées de 15 à 49 ans des ménages enquêtés. Ainsi,

$${}_{35}PRMRate_{15} = \frac{\sum_{x=15,5}^{45} {}_5PRMRate_x \cdot {}_5N_x^f}{\sum_{x=15,5}^{45} {}_5N_x^f} \tag{5}$$

$${}_{35}PPRD_{15} = \frac{\sum_{x=15,5}^{45} {}_5PPRD_x \cdot {}_5N_x^f}{\sum_{x=15,5}^{45} {}_5N_x^f} \tag{6}$$

$${}_{35}PRMRatio_{15} = 100\,000 \cdot \frac{\sum_{x=15,5}^{45} {}_5PRMRate_x \cdot {}_5N_x^f}{\sum_{x=15,5}^{45} {}_5f_x \cdot {}_5N_x^f}. \tag{7}$$

EXEMPLE
À titre d'exemple, nous appliquerons la méthode aux données de l'EDS 2004 du Malawi. On a demandé aux femmes interrogées si leurs sœurs étaient en vie, et, pour les sœurs décédées, de situer le décès par rapport à une éventuelle grossesse.

Étape 1 : Élaboration des tableaux de données
Les tabulations de données requises sont présentées dans le tableau 29.1. Les déclarations de décès de sœurs et de durées d'exposition au risque se réfèrent aux sept années précédant l'enquête.

TABLEAU 29.1 Données de base utilisées pour estimer la mortalité liée à la grossesse, EDS Malawi 2004

| Groupe d'âge | Décès de sœurs | Durée d'exposition au risque de décès des sœurs (personnes-années) | Décès liés à la grossesse | Taux de fécondité par groupe d'âge | Effectifs féminins des ménages |
|--------------|----------------|--|---------------------------|------------------------------------|--------------------------------|
| | (i) | (ii) | (iii) | (iv) | (v) |
| 15-19 | 117 | 27 622 | 8 | 0,162 | 2 570 |
| 20-24 | 227 | 29 331 | 52 | 0,293 | 3 036 |
| 25-29 | 299 | 23 763 | 64 | 0,254 | 2 247 |
| 30-34 | 245 | 17 228 | 50 | 0,222 | 1 516 |
| 35-39 | 230 | 12 206 | 33 | 0,163 | 1 122 |
| 40-44 | 177 | 7 892 | 22 | 0,080 | 970 |
| 45-49 | 82 | 4 574 | 12 | 0,035 | 743 |
| TOTAL | 1 376 | 122 616 | 240 | 0,204* | 12 204 |

* Taux global de fécondité : moyenne des taux de fécondité par groupe d'âge pondérée par la structure par âge de la population féminine des ménages.
Source : EDS Malawi 2004, Tableaux 13.2 (p. 245) et 13.3 (p. 247).

Étape 2 : Calcul des indices de mortalité

L'application de la méthode aux données du tableau 29.1 est détaillée dans le tableau 29.2. La première colonne présente les taux de mortalité par groupe d'âge, calculés en divisant le nombre de décès féminins du groupe d'âge considéré (col. i du tableau 29.1) par la durée d'exposition au risque de décès des sœurs (col. ii du tableau 29.1), conformément à l'équation 1. Dans la deuxième colonne figurent les taux de mortalité liée à la grossesse par groupe d'âge, calculés sur le modèle des taux de mortalité générale par groupe d'âge, mais en ne retenant au numérateur que les décès liés à la grossesse (col. iii du tableau 29.1 - équation 2). La troisième colonne présente, pour chaque groupe d'âge, la part des décès déclarés liés à la grossesse dans l'ensemble des décès féminins déclarés (col. iii/col. i

du tableau 29.1 – équation 3). Dans la quatrième colonne, on trouve les rapports de mortalité liée à la grossesse par groupe d'âge, calculés en divisant le nombre de décès liés à la grossesse (col. iii du tableau 29.1) par la durée d'exposition au risque de décès des sœurs (col. ii) en multipliant par les taux de fécondité par groupe d'âge (col. iv), et en multipliant par 100 000 (équation 4).

Il est important de souligner que les résultats de la ligne Total du tableau 29.2 ne sont pas les simples sommes des chiffres des divers groupes d'âge. La raison en est que la structure par âge des sœurs exposées au risque de décès diffère de celle de la population féminine en âge de procréer. Pour obtenir des totaux corrects au niveau de l'ensemble de la population, il faut repondérer les taux et rapports par groupe d'âge

TABLEAU 29.2 Taux de mortalité féminine aux âges adultes et taux de mortalité liée à la grossesse, EDS Malawi 2004

| Groupe d'âge | Taux de mortalité par groupe d'âge | Taux de mortalité liée à la grossesse par groupe d'âge | Part des décès liés à la grossesse | Rapports de mortalité liée à la grossesse par groupe d'âge |
|---------------|------------------------------------|--|------------------------------------|--|
| | = 1000*(i)/(ii) | = 1000*(iii)/(ii) | = (iii)/(i) | = 100000* iii /(ii*iv) |
| 15-19 | 4,24 | 0,29 | 0,0684 | 178,8 |
| 20-24 | 7,74 | 1,77 | 0,2291 | 605,1 |
| 25-29 | 12,58 | 2,69 | 0,2140 | 1 060,3 |
| 30-34 | 14,22 | 2,90 | 0,2041 | 1 307,3 |
| 35-39 | 18,84 | 2,70 | 0,1435 | 1658,6 |
| 40-44 | 22,43 | 2,79 | 0,1243 | 3 484,5 |
| 45-49 | 17,93 | 2,62 | 0,1463 | 7495,8 |
| TOTAL* | 11,51 | 1,99 | 0,1681 | 970,7 |

Note : tous les totaux de ce tableau sont pondérés par la structure par âge de la population féminine des ménages enquêtés (voir le texte).

TABLEAU 29.3 Comparaison de la distribution des naissances avec celle des décès liés à la grossesse par groupe d'âge, EDS Malawi 2004

| Groupe d'âge | Effectifs féminins des ménages enquêtés | Taux de fécondité par groupe d'âge | Décès liés à la grossesse | Naissances | Proportions de : décès liés à la grossesse | naissances |
|--------------|---|------------------------------------|---------------------------|--------------|---|--------------|
| 15-19 | 2 570 | 162 | 8 | 416 | 0,033 | 0,167 |
| 20-24 | 3 036 | 293 | 52 | 890 | 0,217 | 0,356 |
| 25-29 | 2 247 | 254 | 64 | 571 | 0,267 | 0,228 |
| 30-34 | 1 516 | 222 | 52 | 337 | 0,208 | 0,135 |
| 35-39 | 1 122 | 163 | 33 | 183 | 0,138 | 0,073 |
| 40-44 | 970 | 80 | 22 | 78 | 0,092 | 0,031 |
| 45-49 | 743 | 35 | 12 | 26 | 0,050 | 0,010 |
| TOTAL | 12 204 | | 240 | 2 500 | 1,000 | 1,000 |

du tableau 29.2 par la structure par âge de la population féminine (col. v du tableau 29.1) conformément aux équations 5, 6 et 7. De même, le dénominateur de l'équation 7 n'est pas le taux global de fécondité tel qu'on le calcule habituellement (le nombre de naissances divisé par le nombre de femmes âgées de 15 à 49 ans), mais la somme des taux de fécondité par âge pondérée par la structure par âge de la population féminine.

Étape 3 : Estimation des indices au niveau de la population

Le tableau 29.3 compare la distribution des naissances par groupe quinquennal d'âge (obtenue en multipliant les effectifs féminins des ménages par les taux de fécondité par âge correspondants) avec celle des décès liés à la grossesse (calculée en multipliant les effectifs féminins des ménages par les taux de mortalité liée à la grossesse par âge correspondants).

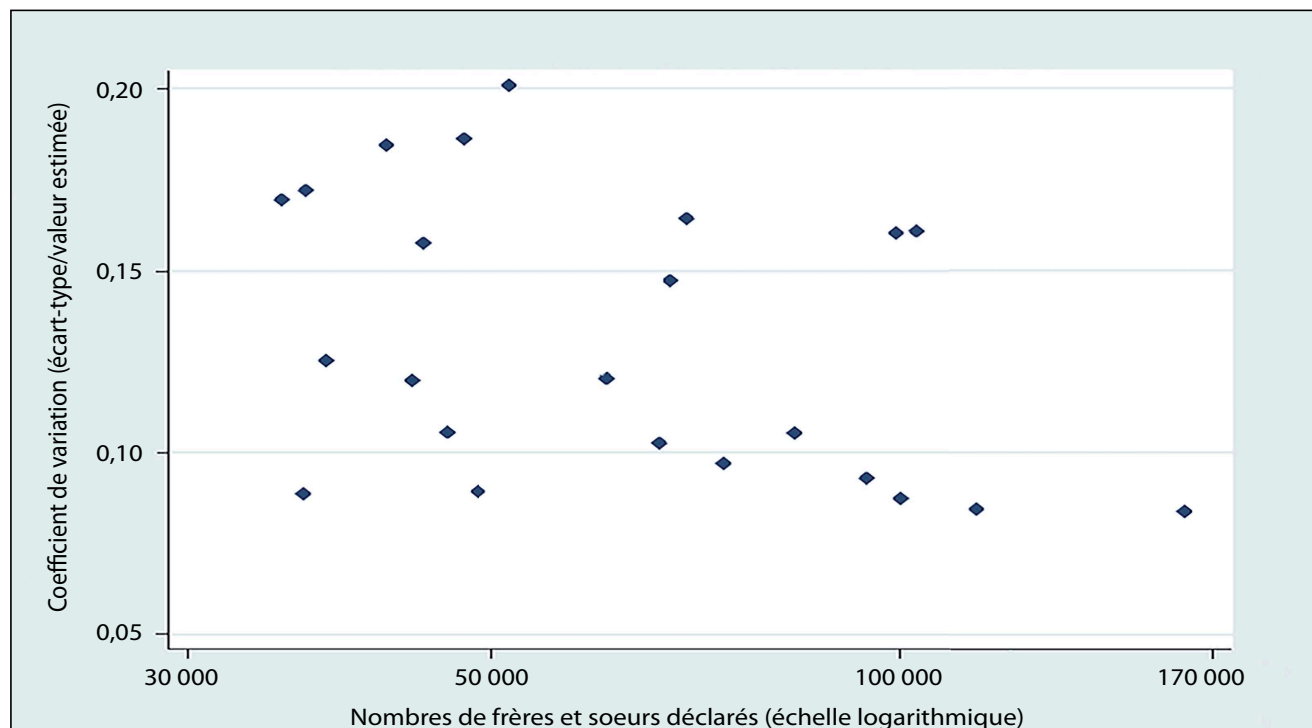
DIAGNOSTICS, ANALYSE ET INTERPRÉTATION

Contrôles et validation

Pour le contrôle et la validation des estimations de la mortalité féminine en général, voir les chapitres de ce manuel consacrés à l'analyse des histoires des frères et sœurs (chapitre 23 et chapitre 27). Le contrôle et la validation des données supplémentaires sur la mortalité liée à la grossesse reposent sur des tests de crédibilité. La proportion globale de décès liés à la grossesse parmi les décès de sœurs en âge de procréer est-elle plausible au regard de l'estimation du rapport de mortalité liée à la grossesse ? La structure par âge des décès liés à la grossesse est-elle crédible par rapport à la structure par âge des naissances (les événements à risque) ?

Il n'y a pas de méthode largement reconnue d'évaluation de la plausibilité de la proportion de décès liés à la grossesse.

FIGURE 29.1 Coefficient de variation du taux de mortalité liée à la grossesse en fonction de la taille de l'échantillon, EDS



Sources : Stanton, Abderrahim and Hill (2000) pour les estimations antérieures à l'année 2000, et rapports nationaux des EDS pour les années 2004 et suivantes.

En général, on observe une association positive entre cette proportion et le taux de mortalité liée à la grossesse, mais elle dépend du niveau de la mortalité non liée à la grossesse et ne fournit aucune base utile pour l'évaluation. On estime la plausibilité de la structure par âge des décès liés à la grossesse par comparaison avec la structure par âge des naissances, comme on le voit dans le tableau 29.3. Dans le cas de l'EDS Malawi 2004, les proportions de décès liés à la grossesse survenus dans les groupes d'âge 15-19 ans et 20-24 ans sont beaucoup plus faibles que les proportions correspondantes de naissances, et c'est l'inverse que l'on observe après 35 ans. Ce dernier phénomène peut s'expliquer par le caractère de plus en plus risqué de la grossesse et de l'accouchement à partir de l'âge de 35 ans, mais aucune explication plausible n'existe pour les importants écarts observés avant l'âge de 25 ans. On a de bonnes raisons de soupçonner que certains décès de sœurs de moins de 25 ans liés à la grossesse n'ont pas été déclarés comme tels.

Interprétation

Pour interpréter les estimations de la mortalité liée à la grossesse à partir des histoires des frères et sœurs, on doit

accorder une grande attention aux erreurs d'échantillonnage et aux erreurs habituelles dans les données. L'erreur d'échantillonnage est très importante, comparée à celle qui affecte les estimations de la mortalité en-dessous de 5 ans dans les EDS. La figure 29.1 met en relation les coefficients de variation (écart-type/valeur estimée) des estimations EDS des taux de mortalité liée à la grossesse avec les nombres de frères et sœurs déclarés. Les coefficients de variation sont supérieurs à 0,08 même pour les très grands échantillons, et supérieurs à 0,10 pour la quasi-totalité des enquêtes.

Étant donné la grande marge d'incertitude inhérente à l'échantillonnage, l'interprétation de différences observées entre régions infranationales ou entre d'autres sous-groupes tels que les classes d'âge est sujette à caution. Les erreurs courantes des données, en particulier l'omission de certains décès survenus longtemps avant l'enquête, rendent suspecte toute interprétation des tendances décelées. Toutes les conclusions en termes de tendances d'évolution devraient être fondées sur des estimations provenant de deux ou plusieurs enquêtes, portant sur des périodes rétrospectives comparables, et tenir compte de l'erreur d'échantillonnage.

RÉFÉRENCES

- Gakidou E and G King. 2006. "Death by survey: estimating adult mortality without selection bias from sibling survival data", *Demography* **43**(3):569-585. doi: <http://dx.doi.org/10.1353/dem.2006.0024>
- Masquelier B. 2013. "Adult mortality from sibling survival data: A reappraisal of selection biases", *Demography* **50**(1):207-228. doi: <http://dx.doi.org/10.1007/s13524-012-0149-1>
- Obermeyer Z, JK Rajaratnam, CH Park, E Gakidou *et al.* 2010. "Measuring adult mortality using sibling survival: a new analytical method and new results for 44 countries, 1974-2006", *PLoS Medicine* **7**(4):e1000260. doi: <http://dx.doi.org/10.1371/journal.pmed.1000260>
- Stanton C, N Abderrahim and K Hill. 2000. "An assessment of DHS maternal mortality indicators", *Studies in Family Planning* **31**(2):111-123. doi: <http://dx.doi.org/10.1111/j.1728-4465.2000.00111.x>
- Trussell J and G Rodriguez. 1990. "A note on the sisterhood estimator of maternal mortality", *Studies in Family Planning* **21**(6):344-346. doi: <http://dx.doi.org/10.2307/1966923>

Chapitre 30 ESTIMATION DE LA MORTALITÉ LIÉE À LA GROSSESSE À PARTIR DES DÉCÈS SURVENUS DANS LES MÉNAGES

Kenneth Hill

DESCRIPTION DE LA MÉTHODE

Si, lors d'un recensement ou d'une grande enquête auprès des ménages, des questions ont été posées sur les décès survenus dans le ménage au cours d'une certaine période de référence, et si, à propos des décès de femmes en âge de procréer, des renseignements complémentaires ont été obtenus sur le moment du décès par rapport à une éventuelle grossesse, on peut en tirer des estimations de mortalité liée à la grossesse. Si on dispose aussi de renseignements sur les causes des décès, par exemple grâce à un questionnaire d'autopsie verbale, on peut évaluer la mortalité maternelle, mais cette situation se présente très rarement, et nous ne la traiterons pas explicitement ici.

L'utilisation des données de recensements ou de grandes enquêtes sur les décès récemment survenus dans les ménages afin d'estimer la mortalité adulte, ainsi que la critique de telles données, sont décrites par ailleurs. Toute erreur dans la déclaration des décès est susceptible d'avoir un impact proportionnel sur les estimations de mortalité liée à la grossesse, l'évaluation de la qualité des données, et leur redressement si besoin est, sont donc des phases capitales de l'analyse.

DONNÉES REQUISES ET HYPOTHÈSES

Tabulations nécessaires

- Répartition des femmes par groupes quinquennaux d'âge, d'après le questionnaire-ménage.
- Nombres de décès de membres du ménage survenus au cours des 12 derniers mois (ou d'une autre période) par sexe et âge.
- Répartition des décès de femmes en âge de procréer (habituellement entre 15 et 49 ans) selon qu'ils sont survenus alors que la femme était enceinte, ou lors d'un accouche-

ment ou dans les 42 jours (ou 2 mois) de la fin d'une grossesse.

- Taux de fécondité par âge et taux global de fécondité de la période de référence.

Si des coefficients de pondération ou d'extrapolation sont associés aux données, il faut les appliquer comme il convient lors de l'élaboration des tableaux.

Hypothèses importantes

- Les dissolutions de ménages résultant du décès d'un de leurs membres sont négligeables.
- (Si une correction des nombres de décès ou de naissances est nécessaire), les erreurs de déclaration de décès et de naissances sont du même ordre de grandeur à tous les âges.
- Le risque d'omission d'un décès est indépendant du fait que ce décès est, ou non, lié à une grossesse.
- Les déclarations de décès en tant que décès liés à la grossesse sont à peu près fiables.

TRAVAUX PRÉPARATOIRES ET RECHERCHES PRÉLIMINAIRES

L'examen de la qualité des données relatives aux décès survenus dans les ménages et aux taux de fécondité par âge récents est décrit par ailleurs. La seule évaluation de qualité des données qui soit propre à l'estimation de la mortalité liée à la grossesse concerne les données sur la part des décès liés à la grossesse parmi les décès de femmes en âge de procréer et les proportions de décès liés à la grossesse survenus respectivement en cours de grossesse, lors d'un accouchement ou dans les 42 jours (ou 2 mois) de la fin d'une grossesse.

La possibilité d'évaluation de la qualité des données pour des problèmes autres que le simple enregistrement des

nombre de naissances et de décès se limite à un test de crédibilité des déclarations de certains décès comme étant liés à la grossesse. La structure par âge de la proportion de décès liés à la grossesse parmi l'ensemble des décès devrait être similaire à la répartition des naissances selon l'âge de la mère.

MISE EN GARDE

On pense généralement que les questions sur les décès de membres du ménage et sur les naissances survenues dans l'année qui précède le recensement ou l'enquête en sous-estiment les nombres réels. Une évaluation minutieuse du taux de couverture de ces deux types d'événements est indispensable. Dans le cas, improbable, où les deux événements sont sous-déclarés dans des proportions à peu près égales, le rapport de mortalité liée à la grossesse ne sera pas gravement biaisé, alors que le taux de mortalité liée à la grossesse restera affecté. De toute façon, si l'examen des données met en évidence des omissions de décès et de naissances, les données devront être corrigées avant toute estimation de la mortalité liée à la grossesse.

APPLICATION DE LA MÉTHODE

L'application de la méthode se fait en plusieurs étapes :

Étape 1 : Élaboration des tableaux de données

Le mode d'élaboration des tableaux de données dépasse le cadre de ce manuel. Il n'est généralement pas difficile de construire des tableaux croisés des nombres de femmes par groupes d'âge, des nombres de décès féminins par groupes d'âge et du moment du décès par rapport à une éventuelle grossesse (en cours de grossesse, lors d'un accouchement ou dans les 42 jours – ou 2 mois – de la fin d'une grossesse). Le cas échéant, ces tabulations doivent être pondérées pour compenser un éventuel sous-enregistrement (évalué par une enquête de contrôle) ou s'il s'agit d'un micro-échantillon. On a également besoin d'estimations des taux de fécondité par âge. Le processus d'estimation de la fécondité est décrit par ailleurs.

Définitions de quelques symboles :

- ${}_5D_x$ – nombre de femmes membres du ménage décédées entre les âges x et $x + 5$
- ${}_5PRD_x$ – nombre de décès liés à la grossesse de femmes membres du ménage survenus entre les âges x et $x + 5$
- ${}_5f_x$ – taux de fécondité des femmes âgées de x à $x + 5$
- ${}_5N_x^f$ – nombre de femmes recensées ou enquêtées âgées de x à $x + 5$

Des tabulations de chacune des variables ci-dessus sont nécessaires.

Étape 2 : Calcul des indices de mortalité

On calcule un taux de mortalité par groupe d'âge en divisant, dans chaque groupe d'âge, le nombre de décès féminins déclarés dans le ménage par le nombre de personnes-années vécues par les femmes du ménage :

$${}_5M_x = \frac{{}_5D_x}{{}_5N_x^f}. \quad (1)$$

Le taux de mortalité liée à la grossesse (*en anglais : pregnancy-related mortality rate, PRMRate*) par groupe d'âge est :

$${}_5PRMRate_x = \frac{{}_5PRD_x}{{}_5N_x^f} \quad (2)$$

et le rapport de mortalité liée à la grossesse (*en anglais : pregnancy-related mortality ratio, PRMR*) par groupe d'âge est :

$${}_5PRMR_x = 100000 \cdot \left(\frac{{}_5PRD_x}{{}_5N_x^f \cdot {}_5f_x} \right). \quad (3)$$

La proportion des décès liés à la grossesse dans l'ensemble des décès féminins (*en anglais : proportion of deaths that are pregnancy-related, PPRD*) par groupe d'âge est :

$${}_5PPRD_x = \frac{{}_5PRD_x}{{}_5D_x^f}. \quad (4)$$

La répartition des décès liés à la grossesse (DPRD) entre 15 ans et 49 ans est :

$${}_5DPRD_x = \frac{{}_5PRD_x}{\sum_{x=15,5} {}_5PRD_x}. \quad (5)$$

EXEMPLE

À titre d'exemple, nous appliquerons la méthode aux données du recensement du Malawi de 2008, en particulier aux données sur les décès survenus dans les 12 mois précédant le recensement.

Étape 1 : Élaboration des tableaux de données

Le tableau 30.1 synthétise les données individuelles d'un échantillon public au dixième (IPUMS) du recensement du Malawi. Il présente la distribution de la population fémi-

nine en âge de procréer par groupes quinquennaux d'âge, les décès féminins déclarés survenus dans les 12 mois précédant le recensement, et, pour les décès féminins survenus entre 15 et 49 ans, le moment du décès par rapport à une éventuelle grossesse : en cours de grossesse, lors de l'accouchement ou dans les 42 jours de la fin de la grossesse.

Étape 2 : Calcul des indices de mortalité

Le tableau 30.2 montre l'application de la méthode aux données telles que les présente le tableau 30.1. Les taux de mortalité générale par âge (1^{re} col.) sont calculés en divisant les nombres de décès (col. vi du tableau 30.1) par les effectifs féminins (col. i du tableau 30.1) conformément à l'équa-

TABLEAU 30.1 Population féminine de 15 à 49 ans et décès féminins situés par rapport à une éventuelle grossesse, par groupes quinquennaux d'âge, Malawi, recensement de 2008

| Groupe d'âge | Population féminine | Décès féminins | | | | |
|--------------------|---------------------|-----------------------|------------------------|-------------|--|--------------------|
| | | En cours de grossesse | Lors de l'accouchement | Post-Partum | Ensemble des décès liés à la grossesse | Ensemble des décès |
| | (i) | (ii) | (iii) | (iv) | (v) | (vi) |
| 15-19 | 67 918 | 43 | 25 | 26 | 94 | 235 |
| 20-24 | 69 069 | 68 | 40 | 36 | 144 | 389 |
| 25-29 | 57 478 | 84 | 31 | 32 | 147 | 442 |
| 30-34 | 41 073 | 92 | 24 | 37 | 153 | 471 |
| 35-39 | 29 993 | 56 | 15 | 23 | 94 | 346 |
| 40-44 | 22 294 | 42 | 4 | 14 | 60 | 238 |
| 45-49 | 17 564 | 38 | 3 | 4 | 45 | 185 |
| Total 15-49 | 310 748 | 423 | 142 | 172 | 737 | 2 306 |

Source : Malawi recensement de 2008 échantillon au dixième.

TABLEAU 30.2 Taux de mortalité féminine adulte et taux de mortalité liée à la grossesse, Malawi, recensement de 2008

| Groupe d'âge | Taux de mortalité féminine par âge | Taux de mortalité liée à la grossesse par âge | Taux de fécondité par âge | Rapport de mortalité liée à la grossesse par âge | Proportion de décès liés à la grossesse | Proportion des décès liés à la grossesse | Proportion des naissances des 12 derniers mois |
|--------------------|------------------------------------|---|---------------------------|--|---|--|--|
| | Tableau 30.1 (vi)/(i) | Tableau 30.1 (v)/(i) | | 100 000* (ii)/(iii) | Tableau 30.1 (v)/(vi) | Tableau 30.1 (v)/Total(v) | |
| 15-19 | 0,00346 | 0,00138 | 0,1108 | 1 249,12 | 0,40000 | 0,127544 | 0,14408 |
| 20-24 | 0,00563 | 0,00208 | 0,2464 | 846,13 | 0,37018 | 0,195387 | 0,32584 |
| 25-29 | 0,00769 | 0,00256 | 0,2296 | 1 113,89 | 0,33258 | 0,199457 | 0,25267 |
| 30-34 | 0,01147 | 0,00373 | 0,1941 | 1 919,15 | 0,32484 | 0,207598 | 0,15264 |
| 35-39 | 0,01154 | 0,00313 | 0,1457 | 2 151,04 | 0,27168 | 0,127544 | 0,08367 |
| 40-44 | 0,01068 | 0,00269 | 0,0718 | 3 748,34 | 0,25210 | 0,081411 | 0,03065 |
| 45-49 | 0,01053 | 0,00256 | 0,0311 | 8 238,13 | 0,24324 | 0,061058 | 0,01046 |
| Total 15-49 | 0,00755 | 0,00241 | 0,1713 | 1 895,63 | 0,33804 | 1 | 1 |

Source : Tableau 30.1 et Malawi, recensement de 2008, échantillon au dixième.

tion 1. Remarquons que, en toute rigueur, la date de référence des décès est en moyenne de 6 mois antérieure à celle de la population recensée, mais l'erreur commise en n'en tenant pas compte est négligeable, et elle sera rectifiée si les décès sont corrigés au moyen d'une méthode appropriée de répartition des décès (voir chapitre 19). Les taux de mortalité liée à la grossesse par âge sont à leur tour calculés en divisant les décès liés à la grossesse (col. v du tableau 30.1) par les effectifs féminins (col. i du tableau 30.1), en application de l'équation 2. Dans la 4^e colonne du tableau 30.2, les rapports de mortalité liée à la grossesse par âge sont obtenus en divisant les taux de mortalité liée à la grossesse par âge (2^e col.) par les taux de fécondité par âge (3^e col., données provenant d'une autre source), suivant l'équation 3. Pour calculer les proportions de décès liés à la grossesse par âge, on divise les décès liés à la grossesse (col. v du tableau 30.1) par l'ensemble des décès (col. vi du tableau 30.1) (équation 4). Enfin, la distribution des décès liés à la grossesse par groupe d'âge se calcule en divisant le nombre de décès liés à la grossesse de chaque groupe d'âge par le nombre total de décès liés à la grossesse (col. v du tableau 30.1) (équation 5).

DIAGNOSTICS, ANALYSE ET INTERPRÉTATION

Contrôles et validation

Avec cette méthode, les principaux contrôles portent sur le taux de couverture des décès de femmes adultes et des naissances (voir Méthodes de distribution des décès – chapitre 19 et Évaluation des données sur la fécondité récente – chapitre 5). Les seuls contrôles propres à cette méthode concernent, l'un, la structure par âge des décès liés à la grossesse, et l'autre – mais il est très peu efficace – la répartition des décès liés à la grossesse selon qu'ils sont survenus en cours de grossesse, lors d'un accouchement ou dans les 6 semaines/2 mois de la fin d'une grossesse. Pour la structure par âge des décès liés à la grossesse, le principal élément de comparaison est la structure par âge des naissances. Ces deux distributions figurent dans les deux dernières colonnes du tableau 30.2. Les proportions de décès liés à la grossesse sont inférieures aux proportions correspondantes de naissances jusqu'au groupe d'âge 25-29 ans, et c'est l'inverse que l'on constate après 30 ans. Ceci reflète probablement la réalité, étant donné que, d'après de très nombreuses observations, le risque de décès lié à la grossesse est pratiquement stable jusqu'à 30 ans, mais augmente fortement après cet âge. Il est éclairant de comparer ces distributions avec celles qui

reposent sur les histoires des frères et sœurs de l'EDS Malawi 2004 (décrites dans le tableau 29.3). Dans le cas de l'histoire des frères et sœurs, les proportions de décès liés à la grossesse survenus avant 25 ans sont très inférieures aux proportions correspondantes de naissances. Ici, l'écart va dans le même sens, mais pas au point de susciter des inquiétudes quant à la qualité des données.

Nous avons noté plus haut que le contrôle de la répartition des décès liés à la grossesse selon qu'ils se sont produits en cours de grossesse, lors d'un accouchement ou dans les 6 semaines/2 mois de la fin d'une grossesse était très peu efficace. Il n'est guère utile parce que les expériences antérieures ne nous indiquent pas clairement quelle allure cette répartition devrait présenter dans différents contextes, et parce que, en pratique, elle varie énormément d'une source à l'autre.

Interprétation

On n'insistera jamais assez sur l'importance de l'évaluation du taux de couverture des décès de femmes adultes et des naissances. Dans certaines situations, la sous-déclaration des décès survenus dans le ménage peut atteindre 50 % ; une telle erreur peut entraîner un biais de même amplitude dans l'estimation du rapport de mortalité liée à la grossesse. Il peut aussi y avoir des omissions de naissances récentes, et cela peut compenser en partie les omissions de décès dans le calcul du rapport de mortalité liée à la grossesse. Les méthodes de distribution des décès indiquent que le sous-enregistrement des décès féminins dans le recensement du Malawi de 2008 était de l'ordre de 40 à 50 %, alors que l'application d'un modèle relationnel de Gompertz aux mêmes données donne un sous-enregistrement des naissances des 12 derniers mois d'environ 18 %. Ces deux biais se compensant partiellement, l'effet net serait une sous-estimation du rapport de mortalité liée à la grossesse de l'ordre de 2/5.

Dans l'interprétation des données sur la mortalité liée à la grossesse au Malawi, il est essentiel de se souvenir que ce pays a été touché par une importante épidémie de VIH à la fin des années 1990 et au début des années 2000. La mortalité associée au VIH affecte les résultats des méthodes de distribution des décès, décrites ailleurs, en sorte que le facteur correcteur présenté plus haut doit alors être appliqué avec prudence.

Les résultats fondés sur des données provenant de recensements ou de grands échantillons censitaires présentent l'avantage de ne pas être soumis à une erreur d'échantillonnage. Donc, s'il est hasardeux de tirer des conclusions

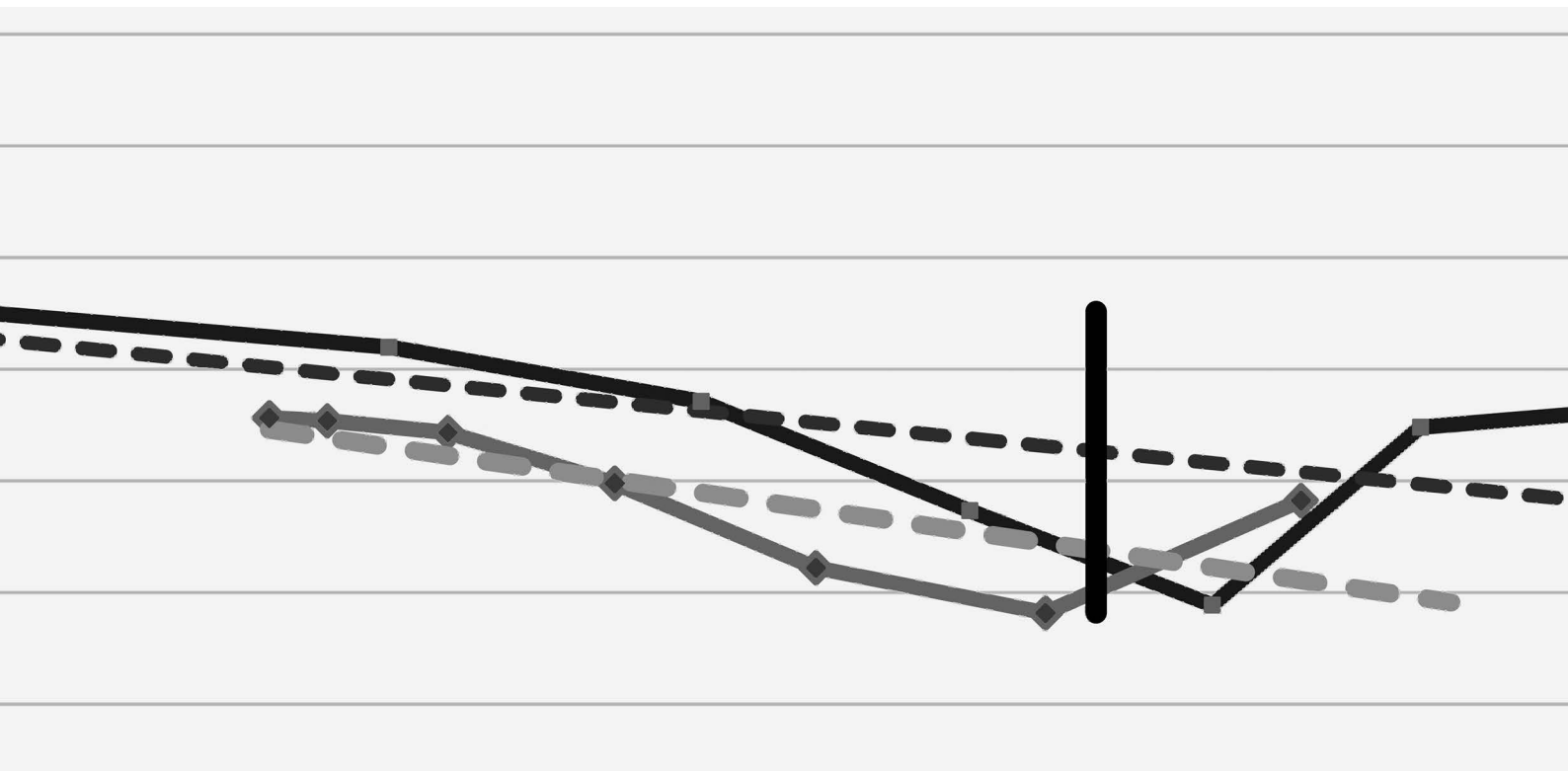
en termes différentiels à partir des données des histoires des frères et sœurs, la même réserve ne s'impose pas autant quand il s'agit d'estimations basées sur des données de recensement (quoique, même dans un recensement, certains effectifs puissent être très faibles pour certains sous-groupes et entraîner une erreur aléatoire). En outre, étant donné qu'on doit évaluer, et souvent corriger, des données

de recensement, certaines estimations restent sujettes à une grande incertitude.

RÉFÉRENCES

Hill, K, C. Stanton, M. Levin *et al.* (2011) *Measuring Maternal Mortality from a Census: Guidelines for Potential Users*. Geneva: World Health Organization.

Utiliser des modèles
pour obtenir des tables de mortalité
à partir de données incomplètes



Chapitre 31 INTRODUCTION AUX TABLES-TYPES DE MORTALITÉ

Tom A Moultrie et Ian M Timæus

UTILISATION DE TABLES-TYPES POUR L'ESTIMATION DE LA MORTALITÉ

On utilise des tables-types de mortalité comme références quand on examine des estimations empiriques de mortalité, pour lisser ou rectifier des valeurs défectueuses et pour compléter une table de mortalité quand on ne dispose d'estimations que pour certains groupes d'âge.

Dans ce contexte, le terme « lisser » renvoie à tout processus d'élimination ou de minimisation des irrégularités observées dans les données recueillies ou dans les premiers indices calculés à partir de ces données. Les techniques de lissage comprennent une grande variété de procédés, qui vont de l'ajustement de modèles au simple calcul de moyennes. Les techniques traditionnelles de lissage ou d'interpolation des structures par âge et des taux de mortalité par âge observés, comme le recours aux splines cubiques, sont décrites dans la littérature actuarielle et démographique, nous ne les abordons pas ici. Nous nous intéresserons plutôt aux procédures fondées sur des modèles, indiquées quand les données de base sont défectueuses ou incomplètes.

En analyse démographique classique, on calcule une table de mortalité en transformant une série complète de taux de mortalité par âge (${}_n m_x$) en quotients de mortalité (${}_n q_x$). On peut alors en déduire les effectifs de survivants ($l(x)$) et les autres fonctions de la table. Mais, quand on analyse des données de recensement ou d'enquête, on n'obtient souvent des estimations de mortalité que pour une partie seulement des groupes d'âge. Par exemple, les estimations de mortalité calculées à partir des données d'une histoire génésique (chapitre 17) ou d'une histoire des frères et sœurs (chapitre 27) ne donnent aucune information sur la mortalité des jeunes sortis de la petite enfance ou sur celle des adultes de plus de 50 ans. Avec des estimations de ce genre, les tables-types de mortalité peuvent être utiles pour lisser les séries de taux de mortalité estimés et pour compléter la table de mortalité en

formulant des hypothèses plausibles sur la valeur des taux aux âges auxquels la mortalité n'a pas pu être mesurée directement.

En outre, si on a évalué indirectement les probabilités de survie à partir d'informations sur les enfants nés vivants et survivants (chapitre 16) ou sur la survie des parents (chapitre 22) ou d'autres membres adultes de la famille, les résultats ne donnent que le niveau de la mortalité dans les grandes catégories d'âge. En particulier, pour les adultes, ces méthodes fournissent des probabilités conditionnelles de survie, c'est-à-dire des probabilités de survie de l'âge A à l'âge B , $l(B)/l(A)$. Dans ce cas, on peut utiliser des tables-types à la fois pour estimer les taux de mortalité par groupes quinquennaux d'âge et pour compléter la table en faisant une hypothèse acceptable sur la mortalité aux grands âges.

Les deux chapitres qui suivent présentent deux manières de construire des tables de mortalité complètes à partir de données sur la mortalité dans certains groupes d'âge seulement. Le Chapitre 32 porte sur les méthodes qui combinent une estimation isolée de la mortalité infanto-juvénile (${}_5 q_0$) et une estimation de survie conditionnelle adulte de la même année ou de la même période, pour en déduire une table de mortalité complète. Plusieurs variantes de cette méthode sont présentées. Le chapitre 33 décrit une méthode qui combine une série d'estimations datées de mortalité infanto-juvénile (comme celles que l'on déduit indirectement des informations sur les enfants nés vivants et les enfants survivants) et une série d'estimations datées de probabilités conditionnelles de survie à l'âge adulte (comme celles fournies par les méthodes indirectes d'exploitation des informations sur la survie des parents ou des frères et sœurs), pour obtenir la table de mortalité relative à une date déterminée.

Dans ce manuel, nous utilisons abondamment les tables-types de mortalité fondées sur des modèles relationnels logit,

premièrement pour évaluer et lisser des séries d'estimations de mortalité des enfants et de mortalité adulte, et deuxièmement pour combiner des estimations indépendantes de mortalité des enfants et de mortalité adulte afin d'obtenir des tables complètes. Le système et ses propriétés font l'objet de la section suivante.

VUE D'ENSEMBLE DU SYSTEME RELATIONNEL LOGIT DE TABLES-TYPES DE MORTALITE

Brass et ses collègues (Brass 1964, 1971 ; Brass et Coale 1968) ont développé un système flexible à deux paramètres de tables-types de mortalité couramment appelé système relationnel logit de tables-types de mortalité. En gros, le premier paramètre, α , représente les écarts de niveau de mortalité entre populations différentes, et le deuxième, β , représente la variabilité d'une population à l'autre du rapport entre mortalité des enfants et celle des adultes.

Ce système de modèles est de type relationnel. C'est-à-dire qu'il est basé sur une transformation mathématique de la fonction de survie par âge, $l(x)$, qui permet de relier entre elles, par une simple équation, deux tables de mortalité différentes. En particulier, Brass a découvert qu'une transformation logit des probabilités de survie jusqu'à l'âge x , $l(x)$, rendait à peu près linéaire la relation entre les probabilités transformées de différentes tables de mortalité.

Donc, si on définit le logit de $l(x)$ comme :

$$Y(x) = \text{logit}(l(x)) = -\frac{1}{2} \ln \left(\frac{l(x)}{1-l(x)} \right), \quad (1)$$

la relation linéaire suivante se vérifie approximativement à tous les âges :

$$Y(x) = \alpha + \beta Y^*(x), \quad (2)$$

$Y(x)$ et $Y^*(x)$ étant les logits des probabilités de survie par âge, $l(x)$ et $l^*(x)$, de deux tables de mortalité différentes, et α et β étant des paramètres constants.

Les habitués de la régression logistique reconnaîtront dans $Y(x)$ la moitié du logarithme du rapport de cotes (*odds*) de la probabilité de décès entre la naissance et l'âge x , puisque :

$$\begin{aligned} \text{logit}(l(x)) \times 2 &= -\ln \left(\frac{l(x)}{1-l(x)} \right) = \ln \left(\frac{1-l(x)}{l(x)} \right) \\ &= \ln \left(\frac{{}_x q_0}{1 - {}_x q_0} \right). \end{aligned}$$

Si l'équation 2 était vérifiée pour toute paire de tables de mortalité, cela signifierait qu'il serait possible de calculer n'importe quelle table de mortalité à partir d'un même modèle de référence, appelé le standard de mortalité, $l^s(x)$, en donnant des valeurs appropriées aux paramètres α et β . En réalité, l'équation n'est qu'approximativement satisfaite par des paires de tables de mortalité réelles, mais l'approximation est assez bonne pour permettre le recours à cette relation pour étudier et modéliser des schémas de mortalité observés.

Avant d'expliquer comment utiliser l'équation 2 pour calculer des tables-types de mortalité, arrêtons-nous un instant sur la signification des paramètres α et β . Considérons l'ensemble des tables de mortalité qui peuvent être calculées à partir d'une certaine table-type, $l^s(x)$, et de différentes valeurs de α et β . Si β est maintenu constant et égal à 1, modifier la valeur de α va soit augmenter soit diminuer les probabilités de survie à chaque âge. Donc, modifier α va engendrer des tables de mortalité dont le profil est essentiellement celui de la table-type $l^s(x)$ utilisée pour les construire, mais dont le niveau général de mortalité varie. Si, en revanche, on donne à α la valeur 0 et si on laisse β varier, les tables obtenues n'auront plus le même profil que $l^s(x)$. Toutes les tables ainsi *calculées* auront une intersection commune en un point unique, situé quelque part vers le milieu de l'éventail des âges, où $l^s(x) = 0,5$ et $Y^s(x) = 0$. Leurs probabilités de survie seront alors soit plus faibles aux âges jeunes et plus fortes aux âges élevés, soit plus fortes aux âges jeunes et plus faibles aux âges élevés, que celles de la table-type qui a servi à les calculer. Ainsi, β affecte la structure de la mortalité calculée plutôt que son niveau. Des variations simultanées de α et β feront varier à la fois le niveau et la structure de la fonction de survie calculée.

On peut déduire de l'équation 1 l'équivalence suivante :

$$l(x) = \frac{1}{1 + \exp(-2Y(x))}$$

et en combinant ceci avec l'équation 2, on obtient :

$$l(x) = \frac{1}{1 + \exp(-2(\alpha + \beta Y^s(x)))}. \quad (3)$$

Ainsi, à partir d'une série quelconque de valeurs de $l^s(x)$ définissant une table-type de mortalité, on peut obtenir une autre série $l(x)$ associée à chaque couple formé d'une valeur de α et d'une valeur de β . (Notons que, aux extrémités de l'éventail des âges, on ne peut plus utiliser l'équation 3 pour

calculer $l(x)$: on doit affecter les valeurs 1 et 0 à $l(0)$ et $l(\omega)$ respectivement.)

L'équation 3 peut servir au calcul de familles de tables-types de mortalité à partir d'une table choisie comme standard, $l^s(x)$. En principe, n'importe quelle table de mortalité peut être retenue comme standard. Par exemple, on peut choisir une table de mortalité fiable, mais relative à une autre date, de la population concernée ou une table de mortalité d'un pays voisin. Mais, quand on n'a pas de telles tables fiables ou appropriées sous la main, on prend souvent comme standard une table-type extraite des tables-types régionales de Princeton (Coale, Demeny et Vaughan 1983) ou des tables-types de mortalité des Nations Unies pour les pays en développement (UN Population Division, UN Model Life Tables for Developing Countries, 1982). Le choix des tables standards employées dans ce manuel est expliqué dans la section suivante.

Du fait de la simplicité mathématique des équations 2 et 3, les tables-types obtenues à l'aide d'un modèle relationnel logit et basées sur un standard quelconque, se calculent facilement dans un tableur, ce qui évite de devoir manipuler de grandes quantités de tables publiées. La forme mathématique simple de l'équation 3 facilite également l'emploi du modèle relationnel logit des tables-types pour procéder à des simulations ou des projections de mortalité. Si on connaît les structures passées et actuelle de la mortalité d'une population, on peut déterminer les tendances des paramètres α et β en appliquant le modèle relationnel logit de tables-types à chacune d'elles, et on peut, moyennant quelques précautions, projeter ces deux paramètres pour obtenir des estimations de la mortalité future.

DESCRIPTION DES TABLES-TYPES DE MORTALITÉ UTILISÉES DANS CE MANUEL

Toutes les tables de mortalité exprimées sous forme logit dans ce manuel sont basées sur une même série de tables standards. Celles-ci sont empruntées aux tables-types régionales de Princeton (Nord, Sud, Est et Ouest) et aux tables-types des Nations Unies pour les pays en développement (modèle général, Amérique Latine, Chili, Asie du Sud, Extrême-Orient), sexes séparés. Toutes ont une espérance de vie à la naissance de 60 ans. Les tables originales ont été modifiées, étendues et améliorées au fil du temps pour pouvoir être appliquées aux âges élevés. Nous utilisons les tables actualisées que la Division de la population des Nations Unies

(UN Population Division 2010) a développées et employées dans ses projections démographiques. Elles fournissent les valeurs de $l(x)$ et de L_x (entre autres) pour les âges 0, 1, 5, 10, ..., 130 ans.

Pour les tables standards basées sur les tables-types régionales de Princeton, les valeurs de $l(2)$, $l(3)$ et $l(4)$ ont été déterminées en appliquant les multiplicateurs proposés par Coale, Demeny et Vaughan (1983, p. 21) à $l(1)$ et $l(5)$. Pour les tables standards basées sur les tables-types des Nations Unies pour les pays en développement, les décès entre 1 et 5 ans ont été répartis par année d'âge dans les mêmes proportions que dans les tables de mortalité par sexe et par région originales.

Certaines méthodes d'estimation de la mortalité des enfants et des adultes nécessitent des tables de mortalité sexes confondus (c'est-à-dire sans distinction de sexe). Comme ces tables (ou leur mise en œuvre) ne sont pas particulièrement sensibles au rapport de masculinité à la naissance, on a fixé celui-ci à 105 garçons pour 100 filles. Les tables de mortalité sexes confondus ont alors été calculées en appliquant les pondérations appropriées aux tables pour les deux sexes séparés :

$$l^c(x) = \frac{(1,05)l^m(x) + l^f(x)}{2,05}$$

où $l^c(x)$ représente le nombre de survivants à l'âge x de la table sexes confondus, et $l^m(x)$ et $l^f(x)$ étant les valeurs correspondantes pour le sexe masculin et le sexe féminin.

Comme ces tables de mortalité sont utilisées presque exclusivement dans un contexte relationnel (comme Brass (1971) l'a prévu initialement), les logits standards des $l^s(x)$ ont été calculés pour tous les âges supérieurs à zéro au moyen de la formule suivante :

$$Y^s(x) = 0,5 \ln \left(\frac{1 - l^s(x)}{l^s(x)} \right).$$

Les valeurs numériques de ces logits peuvent être téléchargées sur le site web associé à cet ouvrage.

LE CHOIX DU STANDARD

Il y a un choix capital à faire lors de la mise en œuvre de méthodes basées sur les tables-types de mortalité ou quand on combine des estimations obtenues par différentes méthodes pour construire une table unique à l'aide du système relationnel de tables-types, c'est le choix de la table standard qui servira de base aux calculs.

Les neuf tables de mortalité standards (quatre tables-types régionales de Princeton et cinq tables-types des Nations Unies pour les pays en développement) employées dans ce manuel représentent des schémas de mortalité très différents. La figure 31.1 montre la relation entre la mortalité des enfants et la mortalité adulte des tables standards sexes confondus – qui ont toutes une espérance de vie à la naissance de 60 ans. L'indice de mortalité des enfants est ${}_5q_0$, la probabilité de mourir entre la naissance et le 5^e anniversaire ; l'indice de mortalité adulte est ${}_{45}q_{15}$, la probabilité de décéder entre 15 et 60 ans (conditionnelle au fait d'avoir survécu à 15 ans).

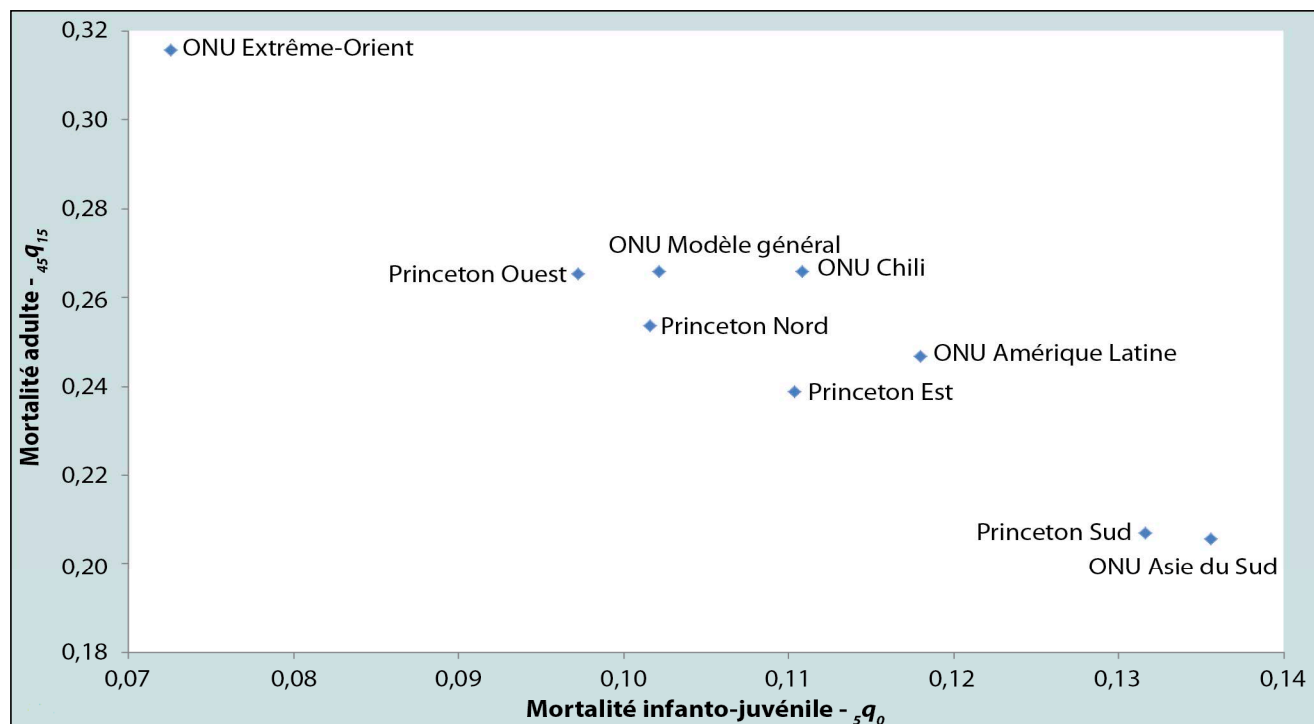
Ainsi, par exemple, la table Extrême-Orient des Nations unies s'avère avoir une mortalité adulte très élevée et une mortalité des enfants très faible, comparativement aux autres tables utilisées, alors que, à l'autre extrême, les tables Sud de Princeton et Asie du Sud des Nations unies ont une mortalité adulte relativement faible, mais une mortalité des enfants très élevée.

Un autre point de divergence entre ces diverses tables est le rapport entre la mortalité infantile (avant le 1^{er} anniversaire) et la mortalité juvénile (entre les 1^{er} et 5^e anniversaires). En comparant ces quotients (figure 31.2), on voit que, si les tables Extrême-Orient et Chili des Nations Unies ont à peu près le même niveau de mortalité juvénile, leurs niveaux de mortalité infantile sont très différents.

Idéalement, le choix d'une table standard devrait refléter au mieux le rapport entre mortalité infantile et mortalité juvénile, d'une part, et le rapport entre mortalité des enfants et mortalité adulte, d'autre part. Ainsi, si on dispose d'une estimation acceptable du schéma de mortalité d'un pays donné, on déterminera la meilleure table standard en comparant le schéma de mortalité observé avec ceux des tables-types. Mais quand l'analyste n'a pas, ou presque pas, d'informations fiables sur la mortalité par âge, il ne peut faire mieux que choisir le schéma qui lui paraît le plus approprié.

Dans les situations où on ne sait rien du schéma de mortalité par âge, il est recommandé d'utiliser soit le modèle Ouest de

FIGURE 31.1 Rapport entre la mortalité adulte et la mortalité des enfants de moins de 5 ans dans différentes table-types de mortalité ayant une espérance de vie à la naissance de 60 ans



Princeton, soit le modèle général des Nations Unies, parce que ces tables-types ont été construites à partir d'un éventail de données relativement large. En outre, le modèle général des Nations unies, en particulier, peut être considéré comme un schéma moyen par rapport aux paramètres des figures 31.1 et 31.2.

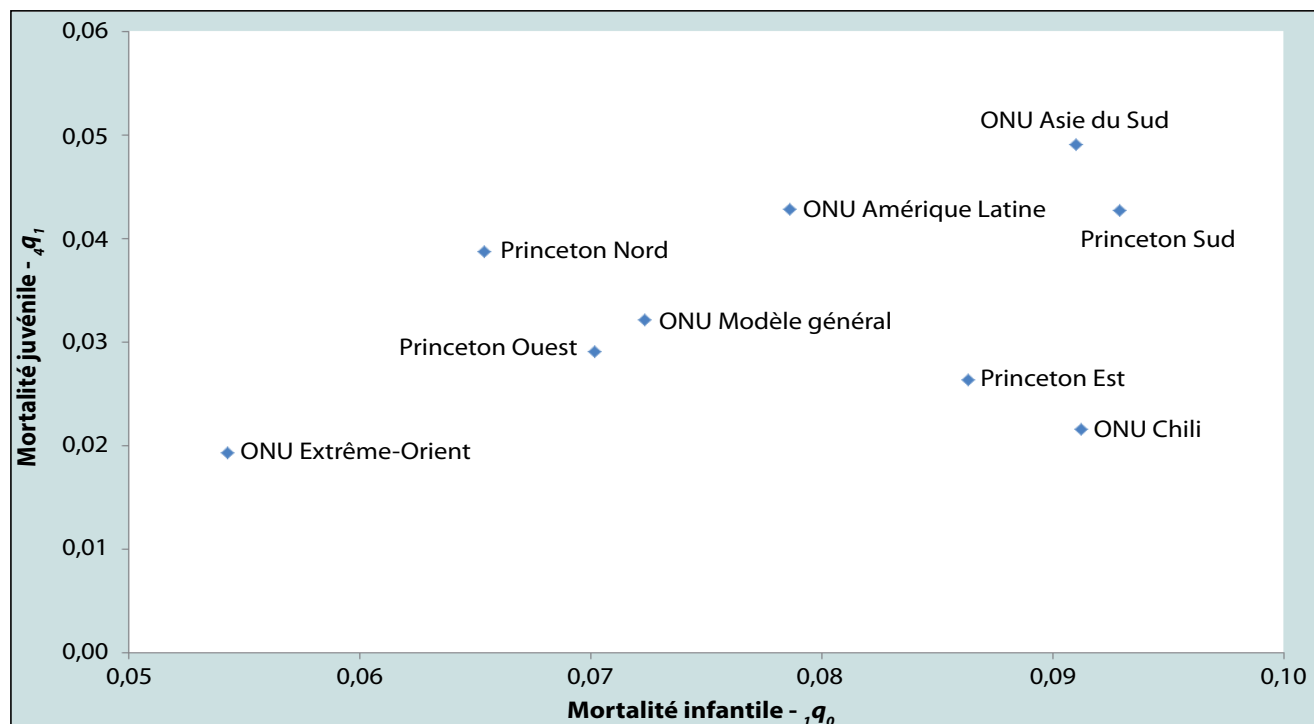
Quant aux « renseignements complémentaires » qui pourraient aider à faire un choix pertinent, ils varient considérablement en genre et en qualité. Ils peuvent aller, par exemple, de taux de mortalité par âge estimés à partir de données d'état civil à la connaissance de quelques traits généraux comme la prévalence et la durée habituelle de l'allaitement maternel dans la population étudiée ou une estimation de la prévalence de la tuberculose.

Quand on dispose d'une série de taux de mortalité par âge observés (de préférence ajustés au moyen d'une méthode de répartition des décès comme celles qui sont décrites aux chapitres 24 et 25), on peut choisir un schéma standard de mortalité en comparant les logits des $l(x)$ observés à ceux des diverses tables-types de mortalité. Cette comparaison peut

être menée à bien en portant sur un graphique la relation entre les valeurs observées de $Y(x)$ et celles qui résultent des diverses tables-types, et en retenant la table-type qui présente la relation la plus linéaire entre les deux séries de valeurs.

À considérer la description donnée plus haut des schémas de mortalité des diverses tables-types, on voit d'emblée qu'ils diffèrent très nettement en termes de valeurs aux premiers âges et de rapport entre mortalité infantile (${}_1q_0$) et mortalité juvénile (${}_4q_1$). Il s'ensuit que des estimations très différentes de la mortalité des enfants peuvent être déduites des mêmes données observées selon la famille de tables-types que l'on aura retenue. De plus, en pareil cas, il peut être difficile de trouver des renseignements extérieurs solides qui aident au choix d'un standard, principalement parce que, bien souvent, les décès infantiles sont largement sous-déclarés. En l'absence de données empiriques sérieuses qui orientent le choix d'une table de mortalité standard adéquate, on peut suggérer quelques lignes de conduite générales qui réduisent l'éventail des possibilités et mènent à un choix raisonnable :

FIGURE 31.2 Rapport entre la mortalité infantile et la mortalité juvénile dans différentes table-types de mortalité ayant une espérance de vie à la naissance de 60 ans



a) Dans une population où l'allaitement maternel est couramment pratiqué et où le sevrage est relativement tardif (au-delà de 12 mois), on peut raisonnablement s'attendre à une mortalité juvénile (${}_4q_1$) relativement élevée comparativement à la mortalité infantile (${}_1q_0$), étant donné que l'allaitement maternel peut s'avérer efficace pour prévenir des décès dus à la malnutrition et aux maladies infectieuses chez les nourrissons. Mais, au moment du sevrage, l'enfant est moins protégé contre ces dangers et donc plus exposé à un risque de décès plus élevé. Dans ce cas, la mortalité des enfants sera probablement bien représentée par le modèle Nord de Princeton ou le modèle général des Nations Unies. Toutefois, on ne peut pas déduire de cette observation que ces tables-types représenteront correctement la mortalité aux autres âges. Seules des informations indépendantes sur la mortalité adulte peuvent étayer cette conclusion.

b) Dans certaines populations d'aujourd'hui, l'allaitement maternel a été abandonné par une grande partie de la population féminine, et, dès leur plus jeune âge, les bébés reçoivent des doses de « lait maternisé » non stériles et souvent inadéquates. Quand cette pratique est adoptée par des mères qui vivent dans un environnement relativement insalubre et que s'y ajoutent de mauvaises conditions d'assistance lors de l'accouchement et dans les heures qui suivent (ce qui peut conduire à une fréquence élevée de tétanos néonatal), la mortalité infantile peut être élevée comparativement à la mortalité des enfants plus âgés. Dans un tel contexte, le modèle Est de Princeton ou le modèle chilien des Nations Unies peuvent représenter correctement la mortalité des enfants. L'avertissement du paragraphe précédent reste valable : il convient de s'assurer que ces tables rendent bien compte du rapport entre mortalité des enfants et mortalité adulte.

c) Le sevrage précoce n'est pas la seule cause de malnutrition susceptible d'entraîner une mortalité infanto-juvénile élevée. Chez certaines populations, l'allaitement maternel est quasi universel, mais, l'hygiène et l'état nutritionnel des enfants étant insuffisants, la mortalité infantile et la mortalité juvénile sont élevées. Pour les pays peu développés de ce genre, le modèle Sud de Princeton ou le modèle sud-asiatique des Nations Unies peuvent être les plus adaptés.

d) En l'absence de données permettant de sélectionner la famille de tables-types qui convient le mieux à un pays donné, on peut choisir la famille qui a été utilisée dans le cas d'un pays voisin dont les caractéristiques culturelles et socio-économiques sont similaires.

e) Pour les raisons exposées plus haut, si on ne connaît pas grand-chose sur la population concernée, on recommande d'utiliser le modèle Ouest de Princeton ou le modèle général des Nations Unies.

Suite à ces remarques, il est clair que nos connaissances sur les schémas de mortalité sont encore assez limitées et que, de toute évidence, de meilleurs renseignements sur la mortalité des populations des pays en développement sont nécessaires pour évaluer la pertinence des modèles disponibles.

AUTRES SYSTÈMES

DE TABLES-TYPES DE MORTALITÉ

Deux des méthodes d'ajustement des tables de mortalité aux données observées qui seront présentées dans le chapitre suivant reposent sur deux autres approches de la modélisation des schémas de mortalité, qui diffèrent quelque peu du travail pionnier de Brass. Nous allons décrire brièvement ces deux autres systèmes de tables-types de mortalité. Pour des renseignements plus complets, nous renvoyons le lecteur intéressé à la littérature correspondante.

Le système logit modifié

Murray et ses collègues (Murray, Ferguson, Lopez *et al.* 2003) ont proposé un système relationnel logit modifié basé sur une table standard générale unique et deux séries de paramètres par âge supplémentaires, $\gamma(x)$ et $\theta(x)$:

$$Y(x) = \alpha + \beta Y^s(x) + \gamma(x) \left(1 - \left(\frac{l(5)}{l^s(5)} \right) \right) + \theta(x) \left(1 - \left(\frac{l(60)}{l^s(60)} \right) \right).$$

Comme précédemment, $Y(x)$ représente la transformation logit de $l(x)$. Les deux premiers paramètres sont ceux du système relationnel logit de Brass. Le premier des deux paramètres additionnels, $\gamma(x)$, relie la probabilité de survie des enfants de moins de 5 ans au standard, tandis que $\theta(x)$ joue le même rôle pour le niveau de survie des adultes.

Bien qu'à première vue on ait affaire à quatre paramètres, cette version modifiée du système relationnel logit reste en fait un modèle à deux paramètres. Puisqu'on donne par définition la valeur zéro à $\gamma(5)$, $\theta(5)$, $\gamma(60)$ et $\theta(60)$, α et β suffisent à déterminer $l(5)$ et $l(60)$ dans le modèle ajusté, et donc aussi les deux séries d'écarts par âge par rapport au schéma standard de mortalité, $\gamma(x)$ et $\theta(x)$. En réalité, ces écarts servent à réduire l'impact qu'a sur la mortalité dans

l'enfance et aux grands âges l'utilisation d'une valeur de β autre que 1 pour modéliser la relation globale entre mortalité des enfants et mortalité adulte.

Les utilisateurs de ce système de modèles doivent être avertis que les valeurs de $\gamma(x)$ et $\theta(x)$ publiées dans l'article de 2003 sont affectées d'un signe erroné. Les paramètres du tableau 3 de cet article doivent donc être multipliés par -1 avant toute utilisation.

Le système log-quadratique

Un autre système de tables-types de mortalité à deux paramètres a été publié récemment par Wilmoth, Zureick, Canudas-Romo *et al.* (2012). Il met en relation quatre constantes scalaires par âge, $a(x)$, $b(x)$, $c(x)$ et $v(x)$, et deux paramètres, h et k , de la manière suivante :

$$\ln({}_n m_x) = a(x) + b(x)h + c(x)h^2 + v(x)k.$$

Les valeurs de $a(x)$, $b(x)$, $c(x)$ et $v(x)$ ont été calculées à partir des données de mortalité de la *Human Mortality Database* ; avec les deux paramètres restants (h et k), on ajuste le modèle aux estimations empiriques de la mortalité.

Le premier paramètre, h , mesure le niveau général de la mortalité ; il est défini comme le logarithme du ${}_5q_0$ observé. Le second, k (combiné avec $v(x)$), représente l'écart entre la structure par âge de la mortalité observée et celle d'une population standard. En pratique, il est choisi pour s'ajuster à un indice observé, ou une série d'indices observés, de la mortalité adulte (par exemple ${}_{45}q_{15}$ ou ${}_{30}q_{30}$).

RÉFÉRENCES

- Brass W. 1964. *Uses of census or survey data for the estimation of vital rates*. Paper prepared for the African Seminar on Vital Statistics, Addis Ababa 14-19 December 1964. Document No. E/CN.14/CAS.4/V57. New York: United Nations.
- Brass W. 1971. "On the scale of mortality," in Brass, W (ed). *Biological Aspects of Demography*. London: Taylor and Francis, pp. 69-110.
- Brass W and AJ Coale. 1968. "Methods of analysis and estimation," in Brass, W, AJ Coale, P Demeny, DF Heisel, *et al.* (eds). *The Demography of Tropical Africa*. Princeton NJ: Princeton University Press, pp. 88-139.
- Coale AJ, P Demeny and B Vaughan. 1983. *Regional Model Life Tables and Stable Populations*. New York: Academic Press.
- Murray CJ, BD Ferguson, AD Lopez, M Guillot, J Salomon and O Ahmad. 2003. "Modified logit life table system: principles, empirical validation, and application", *Population Studies* **57**(2):165-182. doi: 10.1080/0032472032000097083
- UN Population Division. 1982. *Model Life Tables for Developing Countries*. New York: United Nations, Department of Economic and Social Affairs, ST/ESA/SER.A/77. http://www.un.org/esa/population/techcoop/DemMod/model_lifetabs/model_life-tabs.html
- UN Population Division. 2010. *World Population Prospects 2012: Extended Model Life Tables*. New York: United Nations, Department of Economic and Social Affairs. http://esa.un.org/wpp/Model-Life-Tables/data/MLT_UN2010-130_1y.xls
- Wilmoth JR, S Zureick, V Canudas-Romo, M Inoue and C Sawyer. 2012. "A flexible two-dimensional mortality model for use in indirect estimation", *Population Studies* **66**(1):1-28. doi: <http://dx.doi.org/10.1080/00324728.2011.611411>

Chapitre 32 AJUSTER UNE TABLE-TYPE DE MORTALITÉ À UNE ESTIMATION DE LA MORTALITÉ DES ENFANTS ET UNE ESTIMATION DE LA MORTALITÉ DES ADULTES

Ian M Timæus et Tom A Moultrie

DESCRIPTION DE LA MÉTHODE

Dans des pays où les statistiques sont médiocres et défectueuses, il est souvent impossible d'analyser la mortalité en calculant des séries complètes de taux de mortalité par âge à partir des données disponibles. Cependant, la plupart des pays ont recueilli des données qui peuvent servir à l'estimation de la mortalité des enfants. En particulier, il est généralement possible d'évaluer le quotient de mortalité de la naissance au 5^e anniversaire, ${}_5q_0$. De plus, dans de nombreux pays, on peut utiliser soit les statistiques de l'état civil, soit des données de recensement ou d'enquête pour estimer la mortalité adulte. Habituellement, ces indices de mortalité adulte mesurent la probabilité conditionnelle de survie sur un grand intervalle d'âge (par exemple, ${}_{45}p_{15}$, la probabilité de survie du 15^e au 60^e anniversaire). Dans ce chapitre, nous expliquons comment ajuster un système relationnel logit (associé à une table-type de mortalité) à un couple d'estimations de la mortalité des enfants et de la mortalité adulte relatives à la même année ou à la même période.

Quand on peut estimer à la fois la mortalité des enfants et la mortalité adulte d'une population, il est possible d'ajuster à ce couple d'estimations une table-type de mortalité à deux paramètres qui en intègre les valeurs observées. Ainsi, des modèles à deux paramètres exploitent pleinement les données disponibles dans une telle situation, et, du fait qu'ils reproduisent la relation observée entre les niveaux respectifs de la mortalité des enfants et de la mortalité adulte, ils sont beaucoup mieux à même de décrire la structure par âge de la mortalité de la population étudiée qu'un modèle à un paramètre ajusté sur la seule mortalité des enfants.

Les approches présentées ci-dessous peuvent être appliquées à n'importe quel couple d'indices de mortalité adulte et de mortalité des enfants, pourvu que ce dernier soit exprimé en termes de probabilité de survie à partir de la naissance. Dans la plupart des cas, l'indice observé de mortalité des enfants sera le quotient de mortalité infanto-juvénile, ${}_5q_0$. Par contre, l'indice de mortalité adulte est généralement une probabilité conditionnelle, ${}_np_x$, la probabilité de survivre jusqu'à l'âge $x + n$ sous la condition que l'on a déjà survécu jusqu'à l'âge x . Généralement, la limite inférieure de cet intervalle d'âge adulte est plus élevée que la limite supérieure de l'intervalle d'âge sur lequel a été mesurée la mortalité des enfants. Cela rend impossible la conversion directe des probabilités conditionnelles de survie à l'âge adulte en probabilités non conditionnelles. On ne peut donc pas appliquer les méthodes d'ajustement de systèmes relationnels associés à des tables-types de mortalité présentées dans de nombreux manuels proposant une introduction à la modélisation, et on doit recourir à des techniques d'ajustement plus complexes.

Ce chapitre décrit les méthodes qui peuvent être utilisées pour ajuster à un couple d'estimations de mortalité des enfants et de mortalité adulte deux catégories différentes de tables-types de mortalité à deux paramètres. La première série de méthodes se base sur le système relationnel logit associé à des tables-types de mortalité, présenté dans le chapitre 31. Nous exposerons trois variantes de cette approche. La « méthode de la greffe » (*en anglais : splicing method*) utilise différentes tables-types à un paramètre exprimées sous forme logit pour représenter respectivement la mortalité des enfants et celle des adultes (en

posant l'hypothèse que β , le paramètre qui caractérise le schéma par âge de la mortalité, est égal à 1 dans chaque modèle) et greffe la table-type de mortalité adulte sur la table-type de mortalité des enfants au niveau du 15^e anniversaire. Deuxièmement, la « méthode logit de Brass » s'appuie sur les deux paramètres de niveau (α) et de structure (β) du système pour déterminer une table de mortalité correspondant à un standard déterminé. Troisièmement, la « méthode logit modifiée », proposée par Murray, Ferguson, Lopez *et al.* (2003), ajuste, elle aussi, un système relationnel logit à l'aide des deux paramètres α et β , mais utilise sa propre table standard et l'ajuste aux indices estimés de manière à réduire l'effet des variations de β sur la mortalité infantile et la mortalité aux grands âges.

La deuxième série de tables-types de mortalité à deux paramètres est le système « log-quadratique » proposé récemment par Wilmoth, Zureick, Canudas-Romo *et al.* (2012). Basé sur la régression, c'est un système de modèles de $\ln(m_x)$. Il n'est pas explicitement relationnel dans sa formulation, mais sa paramétrisation s'appuie sur le vaste corpus de données de mortalité constitué par la *Human Mortality Database*. On pourrait facilement le reformuler sous forme relationnelle en prenant l'un des modèles comme standard.

Étant donné qu'on a besoin de procédures itératives pour ajuster une table-type de mortalité aux probabilités conditionnelles de survie aux âges adultes, nous n'en donnerons pas ici d'exemple détaillé. Nous renvoyons le lecteur intéressé à la feuille Excel qui est associée au manuel dans sa version en ligne. Cependant, la dernière section de ce chapitre présentera brièvement les résultats de l'application des quatre méthodes à une série d'estimations de la mortalité du Kenya au milieu des années 1980.

DONNÉES REQUISES ET HYPOTHÈSES

Pour la mise en œuvre des méthodes exposées dans ce chapitre, on a besoin d'une estimation de la mortalité infanto-juvénile (${}_5q_0$) et d'une estimation de la mortalité adulte sous condition de survie jusqu'au seuil de l'âge adulte (par exemple ${}_{45}q_{15}$ ou ${}_{35}q_{15}$). Ces deux indices doivent se rapporter à la même année ou à la même période. En principe, on pourrait aussi utiliser des probabilités conditionnelles de survie prenant comme seuil de l'âge adulte un autre anniversaire que le 15^e.

Toutes les méthodes décrites dans ce chapitre font appel, explicitement ou implicitement, à une table standard de

mortalité, que l'on modifie ensuite pour l'adapter aux valeurs observées. Deux des trois méthodes basées sur le système relationnel logit imposent à l'analyste de choisir une table standard à partir de laquelle il calculera la table-type ajustée. Ces méthodes font donc une hypothèse importante en présumant que ce choix est pertinent. La section consacrée à l'introduction aux tables-types de mortalité (chapitre 31) donne quelques règles à suivre pour choisir une table standard adéquate.

MISE EN GARDE

Les quatre approches fournissent généralement des estimations assez similaires des indices résumés tels que l'espérance de vie à la naissance. Mais il y a de grandes différences entre les estimations des indices de mortalité par âge auxquelles elles aboutissent. Ces écarts sont souvent les plus grands pour les populations dans lesquelles le rapport entre la mortalité adulte et la mortalité des enfants diffère beaucoup de la moyenne. Pour l'instant, rien ne démontre que l'une de ces approches soit, généralement ou absolument, supérieure aux autres. De nouvelles recherches sont nécessaires à cet égard. Cependant, Murray, Ferguson, Lopez *et al.* (2003) montrent que le système logit modifié tend à donner de meilleurs résultats que le système original de tables-types de mortalité à deux paramètres, et Wilmoth, Zureick, Canudas-Romo *et al.* (2012) concluent que les modèles log-quadratique et logit modifié sont aussi performants l'un que l'autre.

Étant donné que toutes ces méthodes s'appuient, explicitement ou implicitement, sur la mortalité d'une population standard, elles ne doivent pas être utilisées pour modéliser la mortalité de populations dont le schéma de mortalité par âge s'écarte nettement de celui du standard choisi. Cet avertissement est surtout important dans le cas de populations fortement touchées par le VIH/sida, car, dans les pays atteints par une épidémie généralisée de VIH, la structure par âge de la mortalité est complètement différente de celles des tables-types de Princeton et des Nations Unies, ainsi que de celles qui sont à la base des systèmes logit modifié et log-quadratique à deux paramètres.

Utilisation de la fonction Solver de Microsoft Excel

Les techniques décrites dans ce chapitre sont conçues pour être utilisées quand les seules mesures de la mortalité adulte disponibles sont soumises à une condition de survie jusqu'à un certain âge pris comme seuil de l'âge adulte. À

partir de ce genre d'indices, on ne peut pas déterminer analytiquement quelles estimations des paramètres du modèle produiront les tables-types les plus performantes. Il faut recourir à des méthodes itératives par tâtonnements ou à des fonctions d'optimisation (telles que la fonction *Solver* de Microsoft Excel) pour mettre en œuvre les méthodes décrites ici, à l'exception de la « méthode la greffe » (*splining method*).

Solver n'est pas fourni d'office avec les installations classiques de Microsoft Excel. Pour l'activer, cliquez sur « *File* → *Options* → *Add-ins* → *Manage Excel Add-ins* → *Go...* », puis assurez-vous que la case *Solver Add-in* est cochée.

Les spécifications de la fonction *Solver* ont été préparées dans la feuille Excel disponible en ligne et associée à ce chapitre, en respectant les conditions et contraintes auxquelles il faut se plier. Pour appliquer cette fonction à la feuille de calcul, cliquez sur « *Data* → *Solver* → *Solve* ». De légers changements de caractéristiques (dans les fenêtres « *Set Objective* », « *By Changing Variable Cells* » et « *Subject to the Constraints* ») sont nécessaires pour calculer d'abord le modèle masculin puis le modèle féminin. On identifie facilement les cellules réservées au calcul de la table de mortalité féminine en examinant celles qui sont réservées au sexe masculin.

MÉTHODE 1 : LA « MÉTHODE DE LA GREFFE »

La « méthode de la greffe » est la plus simple des quatre. Elle ajuste une table de mortalité unique aux estimations de mortalité adulte et de mortalité des enfants fournies comme inputs, en s'appuyant sur une table standard judicieusement choisie. Elle combine deux tables-types de mortalité à un paramètre, l'une pour les enfants, qui s'ajuste parfaitement à l'indice observé de mortalité infanto-juvénile (${}_5q_0$), et l'autre pour les adultes, qui s'ajuste parfaitement à la probabilité conditionnelle observée de survie à l'âge adulte (par exemple ${}_{45}q_{15}$ ou ${}_{35}q_{15}$).

Ces deux tables-types de mortalité, pour les enfants et pour les adultes, sont ensuite greffées l'une à l'autre au niveau du 15^e anniversaire. Cet âge est proche de l'âge auquel la mortalité atteint son niveau minimum dans la plupart des tables de mortalité. Ainsi, chacun de ces deux modèles représente une branche du schéma de mortalité par âge, et, par convention, le 15^e anniversaire est choisi comme ligne de démarcation entre la mortalité des enfants et celle des adultes.

Étape 1.1 : Calcul d'une estimation de α pour la branche de la table de mortalité relative aux enfants

À partir de l'estimation de ${}_5q_0$, on calcule la valeur de α pour la branche relative aux enfants de la table de mortalité (jusqu'à 15 ans exacts) par la formule suivante :

$$\alpha^{child} = Y(5) - Y^s(5),$$

où $Y(5)$ est le logit du ${}_5q_0$ observé,

$$Y(5) = \frac{1}{2} \ln \left(\frac{{}_5q_0}{1 - {}_5q_0} \right),$$

et $Y^s(5)$ est son homologue dans la table standard choisie (identifiée par l'exposant s).

Étape 1.2 : Calcul d'une estimation de α pour la branche de la table de mortalité relative aux adultes

On estime la valeur de α pour la mortalité adulte au moyen de la formule suivante :

$$\begin{aligned} \alpha^{adult} &= \text{logit}({}_n p_{15}) - \text{logit}({}_n p_{15}^s) \\ &= \frac{1}{2} \ln \left(\frac{{}_n q_{15}}{1 - {}_n q_{15}} \right) - \frac{1}{2} \ln \left(\frac{1 - {}_n p_{15}^s}{{}_n p_{15}^s} \right), \end{aligned}$$

où le premier terme se calcule à partir de l'indice observé de mortalité adulte sous condition de survie au 15^e anniversaire, ${}_n q_{15}$, et le second se calcule au moyen de l'identité suivante :

$$\begin{aligned} {}_n p_{15}^s &= \frac{l_{15+n}^s}{l_{15}^s} = \frac{(1 + \exp(2Y^s(15 + n)))^{-1}}{(1 + \exp(2Y^s(15)))^{-1}} \\ &= \frac{1 + \exp(2Y^s(15))}{1 + \exp(2Y^s(15 + n))}. \end{aligned}$$

Notons que, le paramètre α pour les adultes se rapportant à une table de mortalité qui commence à 15 ans, il n'est pas possible de le comparer directement avec le paramètre α pour les enfants.

Étape 1.3 : Greffe des deux branches de la table de mortalité

Pour greffer les deux branches de la table au niveau du 15^e anniversaire, on doit calculer les logits des probabilités conditionnelles de survie de la table standard à partir de 15 ans, ${}_a p_{15}^s$, pour $a = 5, 10, 15 \dots$. En utilisant la formule

de ${}_n p_{15}^s$ présentée dans la section précédente pour calculer ${}_a p_{15}^s$, on obtient :

$$\begin{aligned} \text{logit}({}_a p_{15}^s) &= -\frac{1}{2} \ln \left(\frac{{}_a p_{15}^s}{1 - {}_a p_{15}^s} \right) \\ &= -\frac{1}{2} \ln \left(\frac{(1 + \exp(2Y^s(15))) / (1 + \exp(2Y^s(15 + a)))}{1 - (1 + \exp(2Y^s(15))) / (1 + \exp(2Y^s(15 + a)))} \right). \end{aligned}$$

On calcule enfin la table définitive, dont la racine est 1, de la manière suivante :

Pour tout âge inférieur ou égal à 15 :

$$l(x) = \frac{1}{1 + \exp(2(\alpha^{child} + Y^s(x)))}.$$

À partir de 15 ans (c'est-à-dire aux âges $15 + a$ avec $a = 5, 10, 15, \dots$), on multiplie le nombre estimé de survivants à 15 ans de la table ($l(15)$, calculé à partir des valeurs de α pour les enfants) par les probabilités conditionnelles de survie à partir de 15 ans pour obtenir les probabilités non conditionnelles de survie de la manière suivante :

$$l(15 + a) = l(15) \cdot {}_a p_{15}^s = \frac{l(15)}{1 + \exp(2(\alpha^{adult} + \text{logit}({}_a p_{15}^s)))}.$$

Une fois les $l(x)$ calculés, les autres fonctions de la table (entre autres ${}_n m_x$, ${}_n L_x$ et e_x) peuvent être calculées avec les coefficients ${}_n a_x$ appropriés.

MÉTHODE 2 : LA MÉTHODE LOGIT DE BRASS

Dans la section consacrée à l'introduction aux tables-types de mortalité (chapitre 31), nous avons présenté le système logit de Brass, à deux paramètres α et β , dans lequel

$$l(x) = \frac{1}{1 + \exp(2(\alpha + \beta Y^s(x)))}.$$

La méthode des modèles logit de Brass estime les paramètres α et β dans un système relationnel associé à un standard judicieusement choisi, qui reproduit exactement les valeurs observées des indices de mortalité des enfants (${}_5 q_0$) et des adultes (par exemple ${}_{45} q_{15}$ ou ${}_{35} q_{15}$). Ces deux observations sont censées se rapporter à la même date.

Étape 2.1 : Estimation de α et β

Du fait qu'il s'agit d'un modèle à deux paramètres, on peut exprimer chaque paramètre en fonction de l'autre et de l'un des deux indices de mortalité observés. Pour simplifier le

processus d'ajustement du modèle, il est indiqué d'exprimer α en fonction de β et de ${}_5 q_0$. Dans le système de modèles logit relationnels :

$$Y(5) = 0,5 \ln \left(\frac{{}_5 q_0}{1 - {}_5 q_0} \right),$$

$$Y(5) = \alpha + \beta Y^s(5)$$

et donc

$$\alpha = Y(5) - \beta Y^s(5). \quad (1)$$

Sachant que l'indice de mortalité adulte est la probabilité conditionnelle qu'une personne de 15 ans décède avant l'âge exact $15 + n$, ${}_n q_{15}$, alors

$$\begin{aligned} {}_n q_{15} &= 1 - {}_n p_{15} = 1 - \frac{l_{15+n}}{l_{15}} \\ &= 1 - \frac{1 + \exp(2(\alpha + \beta Y^s(15)))}{1 + \exp(2(\alpha + \beta Y^s(15 + n)))}. \end{aligned}$$

En remplaçant α par l'équation 1, on obtient :

$${}_n q_{15} = 1 - \frac{1 + \exp(2(Y(5) + \beta(Y^s(15) - Y^s(5))))}{1 + \exp(2(Y(5) + \beta(Y^s(15 + n) - Y^s(5))))}.$$

$Y(5)$ étant connu, on peut, avec une table standard, résoudre cette équation, par tâtonnements ou en utilisant la fonction *Solve* d'Excel, pour obtenir la valeur de β qui reproduit la valeur observée de ${}_n q_{15}$. Idéalement, cette valeur de β devrait rester raisonnablement proche de sa valeur centrale, 1. Si β est très différent de 1 (par exemple, s'il est inférieur à 0,8 ou supérieur à 1,25), il est recommandé de refaire les calculs avec une autre table standard dont la structure par âge de la mortalité est plus proche de celle de la population étudiée.

Étape 2.2 : Calcul d'une table de mortalité complète

La valeur de β obtenue détermine celle de α (en l'introduisant dans l'équation 1). En utilisant ces deux paramètres et la table standard choisie, on peut développer une table de mortalité complète dont la racine est 1, au moyen de la relation logit habituelle :

$$l(x) = \frac{1}{1 + \exp(2(\alpha + \beta Y^s(x)))}.$$

Une fois les $l(x)$ calculés, les autres fonctions de la table (entre autres ${}_n m_x$, ${}_n L_x$ et e_x) peuvent être calculées avec les coefficients ${}_n a_x$ appropriés.

MÉTHODE 3 : LE SYSTÈME LOGIT MODIFIÉ

Le système logit modifié, décrit par Murray, Ferguson, Lopez *et al.* (2003), est une extension relativement simple du système logit de Brass. Ces auteurs ont proposé un moyen de calculer une table de mortalité en utilisant une table standard unique et des paramètres supplémentaires dépendant de l'âge, $\gamma(x)$ et $\theta(x)$, formant un système logit modifié :

$$Y(x) = \alpha + \beta Y^s(x) + \gamma(x) \left(1 - \frac{Y(5)}{Y^s(5)} \right) + \theta(x) \left(1 - \frac{Y(60)}{Y^s(60)} \right). \quad (2)$$

Comme précédemment, $Y(x)$ représente la transformation logit de $l(x)$. Les deux premiers paramètres sont ceux du système relationnel logit de Brass. La première des deux séries de coefficients additionnels, $\gamma(x)$, est paramétrée par la survie des enfants de moins de 5 ans rapportée à sa valeur dans la table standard, tandis que la deuxième série, les $\theta(x)$, est paramétrée par la survie des adultes jusqu'à 60 ans rapportée à sa valeur dans la table standard. Les valeurs de $\gamma(x)$ et $\theta(x)$,

ainsi que les tables standards pour chaque sexe, sont présentées au tableau 32.1.

Les utilisateurs de ce système doivent être avertis que les valeurs de $\gamma(x)$ et $\theta(x)$ publiées dans l'article de 2003 sont affectées d'un signe erroné. Les paramètres du tableau 3 de cet article doivent donc être multipliés par -1 avant toute utilisation. Cette correction a été faite dans le tableau 32.1.

Bien qu'à première vue on ait affaire à quatre paramètres, cette version modifiée du système relationnel logit reste en fait un modèle à deux paramètres. Puisqu'on donne par définition la valeur zéro à $\gamma(5)$, $\theta(5)$, $\gamma(60)$ et $\theta(60)$, α et β suffisent à déterminer $l(5)$ et $l(60)$ dans le modèle ajusté, et donc aussi les deux séries d'écarts par âge par rapport au schéma standard de mortalité, $\gamma(x)$ et $\theta(x)$. Le modèle peut donc être ajusté, par une procédure itérative, à n'importe quel couple d'estimations de la mortalité des enfants et de la mortalité adulte. En particulier, la valeur de $Y(60)$ qui définit $Y(x)$ pour $x \neq 5, 60$ est celle de la table ajustée finale. Il n'est donc pas nécessaire de connaître $l(60)$ avant d'ajuster le modèle, et un système logit modifié peut être calculé à partir de n'importe quel indice pertinent de mortalité adulte.

| Sexe masculin | | | | Sexe féminin | | |
|---------------|-------------|-------------|----------|--------------|-------------|----------|
| Âge (x) | $\gamma(x)$ | $\theta(x)$ | $l^s(x)$ | $\gamma(x)$ | $\theta(x)$ | $l^s(x)$ |
| 0 | 0,0000 | 0,0000 | 100 000 | 0,0000 | - 0,0000 | 100 000 |
| 1 | - 0,1607 | 0,0097 | 96 870 | - 0,0855 | - 0,0734 | 97 455 |
| 5 | 0 | 0 | 96 010 | 0 | 0 | 96 651 |
| 10 | 0,0325 | - 0,0025 | 95 666 | 0,0026 | 0,0229 | 96 370 |
| 15 | 0,0297 | - 0,0047 | 95 385 | - 0,0291 | 0,0485 | 96 153 |
| 20 | - 0,0427 | - 0,0018 | 94 782 | - 0,1199 | 0,1090 | 95 795 |
| 25 | - 0,1262 | 0,0210 | 93 915 | - 0,1931 | 0,1702 | 95 340 |
| 30 | - 0,1877 | 0,0518 | 93 007 | - 0,2352 | 0,2117 | 94 824 |
| 35 | - 0,2430 | 0,0883 | 91 949 | - 0,2686 | 0,2408 | 94 197 |
| 40 | - 0,2899 | 0,1248 | 90 575 | - 0,3003 | 0,2601 | 93 370 |
| 45 | - 0,3148 | 0,1482 | 88 645 | - 0,3203 | 0,2594 | 92 220 |
| 50 | - 0,2888 | 0,1402 | 85 834 | - 0,2935 | 0,2183 | 90 569 |
| 55 | - 0,1915 | 0,0910 | 81 713 | - 0,1967 | 0,1338 | 88 159 |
| 60 | 0 | 0 | 75 792 | 0 | 0 | 84 679 |
| 65 | 0,2304 | - 0,1170 | 67 493 | 0,2794 | - 0,1859 | 79 481 |
| 70 | 0,5523 | - 0,2579 | 56 546 | 0,7066 | - 0,4377 | 71 763 |
| 75 | 0,9669 | - 0,4150 | 42 989 | 1,2835 | - 0,7534 | 60 358 |
| 80 | 1,5013 | - 0,5936 | 28 117 | 2,0296 | - 1,1360 | 44 958 |
| 85 | 2,2126 | - 0,8051 | 14 364 | 2,9576 | - 1,5774 | 27 123 |

Source : Murray, Ferguson, Lopez *et al.* (2003), tableau 3 avec inversion des signes des coefficients γ et θ (voir l'explication dans le texte).

TABLEAU 32.1 Coefficients $\gamma(x)$ et $\theta(x)$ et table standard du système logit modifié

Étape 3.1 : Estimation de α et β

Le paramètre α produit par cette méthode peut être exprimé en fonction de β et de l'indice observé de mortalité des enfants, ${}_5q_0$:

$$\alpha = Y(5) - \beta Y^s(5),$$

où $Y(5)$ est le logit du ${}_5q_0$ observé et $Y^s(5)$ est le logit du $l(5)$ de la table standard présentée au tableau 32.1. En outre, dans la table ajustée :

$$Y(60) = \alpha + \beta Y^s(60) = Y(5) + \beta(Y^s(60) - Y^s(5)).$$

Le modèle peut alors être récrit en fonction de β , de $Y(5)$ et des valeurs empruntées à la table standard :

$$Y(x) = Y(5) + \beta(Y^s(x) - Y^s(5)) + \gamma(x) \left(1 - \frac{Y(5)}{Y^s(5)}\right) + \theta(x) \left(1 - \frac{Y(5) + \beta(Y^s(60) - Y^s(5))}{Y^s(60)}\right).$$

Au moyen de cette équation, on peut estimer β par itérations, en vue d'obtenir la valeur de ce paramètre qui engendre une table de mortalité dont l'indice de mortalité adulte (par exemple ${}_{45}q_{15}$ ou ${}_{35}q_{15}$) est égal à celui qui a été observé.

Étape 3.2 : Calcul d'une table de mortalité complète

On peut utiliser les valeurs calculées de β et de $Y(5)$ pour déduire α de l'équation 1. Comme $\gamma(60) = \theta(60) = 0$, on peut ensuite calculer le logit de la probabilité de survie à 60 ans :

$$Y(60) = \alpha + \beta Y^s(60).$$

Avec ces quatre éléments (α , β , $Y(5)$ et $Y(60)$), il est possible de calculer toute la série des valeurs de $Y(x)$ au moyen de l'équation 2 et des coefficients du tableau 32.1.

Une fois obtenus les $Y(x)$, la table de mortalité associée se calcule par la relation logit habituelle :

$$l(x) = \frac{1}{1 + \exp(2Y(x))}.$$

Après les $l(x)$, on peut calculer les autres fonctions de la table (par exemple ${}_n m_x$, ${}_n L_x$ et e_x) avec les coefficients ${}_n a_x$ appropriés.

MÉTHODE 4 : LE SYSTÈME LOG-QUADRATIQUE

Une autre manière de construire des tables de mortalité à partir de données partielles a été publiée récemment par Wilmoth, Zureick, Canudas-Romo *et al.* (2012). Il s'agit de modéliser les taux de mortalité par âge (${}_n m_x$) d'une population en fonction de constantes scalaires dépendant de l'âge, $a(x)$, $b(x)$, $c(x)$ et $v(x)$, et des paramètres h et k selon la relation suivante :

$$\ln({}_n m_x) = a(x) + b(x)h + c(x)h^2 + v(x)k.$$

Les valeurs des quatre constantes ont été calculées à partir des données de la *Human Mortality Database*, et les deux paramètres restants (h et k) servent à l'ajustement du modèle. Les valeurs de $a(x)$, $b(x)$, $c(x)$ et $v(x)$, pour chaque sexe, sont présentées dans le tableau 32.2.

Le paramètre h est défini comme le logarithme du quotient observé ${}_5q_0$. Le second, k (combiné avec $v(x)$), représente l'écart entre la structure par âge de la mortalité observée et celle d'une population standard. En pratique, il est choisi pour reproduire l'indice observé de mortalité adulte conditionné par le ${}_5q_0$ observé.

Étape 4.1 : Estimation de h et k

Dans cette méthode, l'estimation de h découle de celle de l'indice observé de mortalité des enfants, ${}_5q_0$:

$$h = \ln({}_5q_0).$$

La valeur de k se calcule par itérations, en la faisant varier de manière à augmenter ou diminuer les taux de mortalité par âge jusqu'à ce que l'indice conditionnel de survie de la table ajustée (par exemple ${}_{45}q_{15}$ ou ${}_{35}q_{15}$) coïncide avec sa valeur observée.

Étape 4.2 : Calcul d'une table de mortalité complète

Une fois que h et k ont été évalués, on calcule directement une série complète de ${}_5 m_x$ avec les coefficients du tableau 32.2. Notons que, pour s'assurer que la valeur estimée de ${}_5q_0$ coïncide avec sa valeur observée, la mortalité entre 1 et 4 ans se calcule comme la différence entre l'estimation initiale de ${}_5q_0$ et le quotient estimé de mortalité infantile, ${}_1q_0$:

$${}_4q_1 = 1 - \frac{1 - {}_5q_0}{1 - {}_1q_0}.$$

Comme les résultats de l'application des modèles log-quadratiques sont les taux de mortalité entre les âges x et $x + 5$, ${}_n m_x$, on a besoin des multiplicateurs ${}_n a_x$ qui mesurent le

nombre moyen d'années vécues dans un groupe d'âge donné par les individus qui meurent dans ce groupe d'âge, pour convertir les taux de mortalité en quotients de mortalité et calculer des indices tels que ${}_{45}q_{15}$ ou ${}_{35}q_{15}$. Une simplification qui donne généralement des résultats raisonnables entre 5 et 59 ans consiste à supposer que les personnes décédées entre les âges x et $x + 5$ sont mortes en moyenne à l'âge exact $x + 2,7$. (C'est cette hypothèse que l'on a faite dans les feuilles Excel associées à ce chapitre, et on a utilisé des multiplicateurs progressivement plus faibles pour les grands âges.)

Il est plus délicat de déterminer le multiplicateur à employer pour évaluer ${}_1q_0$ à partir de ${}_1m_0$, car l'hypothèse adoptée peut avoir un effet important sur le résultat obtenu.

En l'absence de renseignements empiriques sur la valeur de ${}_1a_0$ qui convient pour la population étudiée, on peut estimer une valeur plausible en recourant aux équations proposées par Preston, Heuveline et Guillot (2001, p. 48), qui se basent sur la famille Ouest de tables-types de mortalité de Princeton. (C'est ainsi qu'a été calculé ${}_1a_0$ dans les feuilles Excel associées à ce chapitre.)

EXEMPLE

Comme nous l'avons dit dans l'introduction, il serait laborieux de présenter des exemples détaillés de mise en œuvre de ces méthodes, étant donné que des procédures itératives sont nécessaires pour calculer les tables ajustées. Les feuilles Excel associées et disponibles en ligne fournissent un exemple

TABLEAU 32.2 Coefficients $a(x)$, $b(x)$, $c(x)$ et $v(x)$ du système log-quadratique de tables de mortalité

| Sexe masculin | | | | | | Sexe féminin | | | |
|---------------|-----|----------|----------|----------|--------|--------------|----------|----------|--------|
| x | n | $a(x)$ | $b(x)$ | $c(x)$ | $v(x)$ | $a(x)$ | $b(x)$ | $c(x)$ | $v(x)$ |
| 0 | 1 | − 0,5101 | 0,8164 | − 0,0245 | 0 | − 0,6619 | 0,7684 | − 0,0277 | 0 |
| 1 | 4 | | | | | | | | |
| 5 | 5 | − 3,0435 | 1,5270 | 0,0817 | 0,1720 | − 2,5608 | 1,7937 | 0,1082 | 0,2788 |
| 10 | 5 | − 3,9554 | 1,2390 | 0,0638 | 0,1683 | − 3,2435 | 1,6653 | 0,1088 | 0,3423 |
| 15 | 5 | − 3,9374 | 1,0425 | 0,0750 | 0,2161 | − 3,1099 | 1,5797 | 0,1147 | 0,4007 |
| 20 | 5 | − 3,4165 | 1,1651 | 0,0945 | 0,3022 | − 2,9789 | 1,5053 | 0,1011 | 0,4133 |
| 25 | 5 | − 3,4237 | 1,1444 | 0,0905 | 0,3624 | − 3,0185 | 1,3729 | 0,0815 | 0,3884 |
| 30 | 5 | − 3,4438 | 1,0682 | 0,0814 | 0,3848 | − 3,0201 | 1,2879 | 0,0778 | 0,3391 |
| 35 | 5 | − 3,4198 | 0,9620 | 0,0714 | 0,3779 | − 3,1487 | 1,1071 | 0,0637 | 0,2829 |
| 40 | 5 | − 3,3829 | 0,8337 | 0,0609 | 0,3530 | − 3,2690 | 0,9339 | 0,0533 | 0,2246 |
| 45 | 5 | − 3,4456 | 0,6039 | 0,0362 | 0,3060 | − 3,5202 | 0,6642 | 0,0289 | 0,1774 |
| 50 | 5 | − 3,4217 | 0,4001 | 0,0138 | 0,2564 | − 3,4076 | 0,5556 | 0,0208 | 0,1429 |
| 55 | 5 | − 3,4144 | 0,1760 | − 0,0128 | 0,2017 | − 3,2587 | 0,4461 | 0,0101 | 0,1190 |
| 60 | 5 | − 3,1402 | 0,0921 | − 0,0216 | 0,1616 | − 2,8907 | 0,3988 | 0,0042 | 0,0807 |
| 65 | 5 | − 2,8565 | 0,0217 | − 0,0283 | 0,1216 | − 2,6608 | 0,2591 | − 0,0135 | 0,0571 |
| 70 | 5 | − 2,4114 | 0,0388 | − 0,0235 | 0,0864 | − 2,2949 | 0,1759 | − 0,0229 | 0,0295 |
| 75 | 5 | − 2,0411 | 0,0093 | − 0,0252 | 0,0537 | − 2,0414 | 0,0481 | − 0,0354 | 0,0114 |
| 80 | 5 | − 1,6456 | 0,0085 | − 0,0221 | 0,0316 | − 1,7308 | − 0,0064 | − 0,0347 | 0,0033 |
| 85 | 5 | − 1,3203 | − 0,0183 | − 0,0219 | 0,0061 | − 1,4473 | − 0,0531 | − 0,0327 | 0,0040 |
| 90 | 5 | − 1,0368 | − 0,0314 | − 0,0184 | 0 | − 1,1582 | − 0,0617 | − 0,0259 | 0 |
| 95 | 5 | − 0,7310 | − 0,0170 | − 0,0133 | 0 | − 0,8655 | − 0,0598 | − 0,0198 | 0 |
| 100 | 5 | − 0,5024 | − 0,0081 | − 0,0086 | 0 | − 0,6294 | − 0,0513 | − 0,0134 | 0 |
| 105 | 5 | − 0,3275 | − 0,0001 | − 0,0048 | 0 | − 0,4282 | − 0,0341 | − 0,0075 | 0 |
| 110 | | − 0,2212 | − 0,0028 | − 0,0027 | 0 | − 0,2966 | − 0,0229 | − 0,0041 | 0 |

Source : Wilmoth, Zureick, Canudas-Romo *et al.* (2012, tableau 3).

d'application de chaque méthode aux estimations de mortalité du Kenya au milieu des années 1980. (Notons que ces estimations sont antérieures aux premières conséquences démographiques significatives de l'épidémie de sida au Kenya, et que l'on peut donc calculer des tables de mortalité pour cette population en utilisant des modèles qui ne tiennent pas compte de l'impact qu'une grave épidémie de VIH a sur la structure par âge de la mortalité.)

Les couples d'indices de mortalité des enfants et de mortalité adulte auxquels les quatre modèles ont été ajustés sont présentés dans le tableau 32.3.

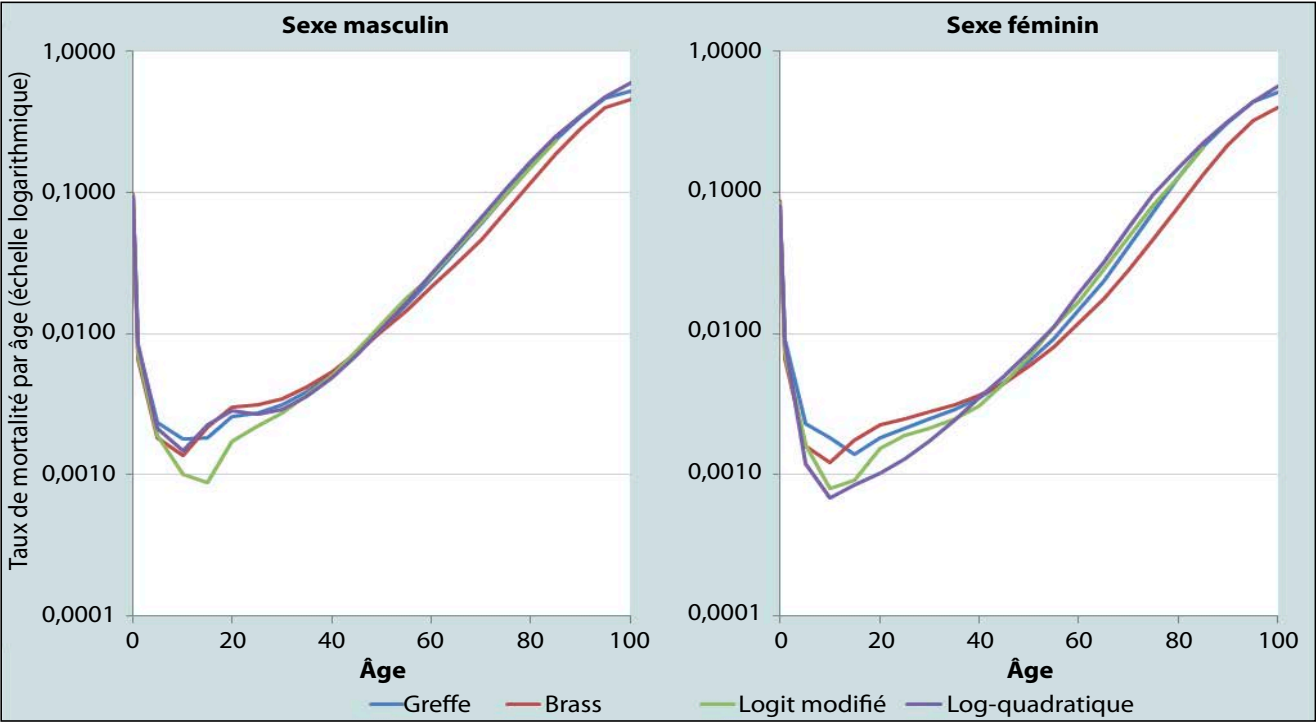
TABLEAU 32.3 Estimations de la mortalité infanto-juvénile et de la mortalité adulte, par sexe, au Kenya au milieu des années 1980

| | Sexe masculin | Sexe féminin |
|-----------------|---------------|--------------|
| ${}_5q_0$ | 0,1180 | 0,1080 |
| ${}_{45}q_{15}$ | 0,2352 | 0,1581 |

La figure 32.1 présente les ${}_nm_x$ (sur une échelle logarithmique) des tables de mortalité ajustées, basées, dans le cas des deux premières méthodes, sur les tables-types de la famille Ouest de Princeton. D'autres paramètres qui mettent en lumière les différences entre les modèles sont présentés dans le tableau 32.4.

On voit sur la figure 32.1 que les taux de mortalité masculins donnés par la « méthode de la greffe » et par la méthode log-quadratique sont presque identiques. Ceux de la méthode logit modifiée sont également très similaires à ceux-ci aux âges adultes (pour les hommes), mais ils sont beaucoup plus faibles entre 5 et 24 ans. Par ailleurs, le modèle logit de Brass concorde avec la méthode de la greffe et la méthode log-quadratique aux âges jeunes, mais donne des taux nettement inférieurs au-dessus de 60 ans, à cause de la faible valeur de β . Les caractéristiques des quatre courbes féminines sont identiques à celles des hommes, sauf pour la méthode log-quadratique qui donne une mortalité plus faible que les autres méthodes entre 10 et 34 ans.

FIGURE 32.1 Valeurs estimées des ${}_nm_x$ (en échelle logarithmique) dans le cas de quatre tables de mortalité calculées sur la base des estimations du tableau 32.3 (famille Ouest de tables-types de Princeton)



TABEAU 32.4 Indices de mortalité selon quatre tables de mortalité calculées sur la base des estimations du tableau 32.3 (famille Ouest de tables-types de Princeton)

| Sexe masculin | | | | | Sexe féminin | | | |
|-----------------|--------|-------------|----------------|-----------------|--------------|-------------|----------------|-----------------|
| Indice | Grefte | Brass logit | Logit modifiée | Log-quadratique | Grefte | Brass logit | Logit modifiée | Log-quadratique |
| e_0 | 59,7 | 60,9 | 60,3 | 59,3 | 64,3 | 66,6 | 64,4 | 63,9 |
| ${}_1q_0$ | 0,0891 | 0,0944 | 0,0923 | 0,0880 | 0,0815 | 0,0842 | 0,0823 | 0,0768 |
| $l(15)$ | 0,8640 | 0,8679 | 0,8693 | 0,8662 | 0,8737 | 0,8796 | 0,8812 | 0,8836 |
| ${}_{20}q_{60}$ | 0,6706 | 0,5793 | 0,6771 | 0,7054 | 0,5346 | 0,4050 | 0,5840 | 0,6416 |

Le tableau 32.4 illustre les effets de ces différences de taux de mortalité par âge. Les modèles logit de Brass donnent une mortalité aux grands âges nettement plus faible que les autres modèles, et des espérances de vie à la naissance sensiblement plus élevées. D'autre part, les deux modèles log-quadratiques présentent la plus forte mortalité aux grands âges et les plus faibles espérances de vie. Tous les modèles donnent des estimations assez semblables de la survie dans l'enfance (c'est-à-dire jusqu'à 15 ans), mais des valeurs diverses de la mortalité infantile. Les résultats des modèles de Brass sont les plus élevés, ils reflètent les mêmes valeurs faibles de β qui ont donné une mortalité basse aux grands âges. Ces différences entre les modèles impliquent que l'on ne peut pas espérer évaluer correctement les taux de mortalité infantile ou de mortalité aux grands âges sinon par les seules méthodes directes.

Il faut garder à l'esprit que, si les deux premiers modèles s'appuyaient sur un standard différent, cela altérerait les caractéristiques des tables ajustées par le système logit modifié et log-quadratique. Par exemple, adopter comme standard le modèle Sud de Princeton ou le modèle asiatique des

Nations Unies donnerait des tables ajustées par la « méthode de la greffe » ou celle de Brass plus proches de celles obtenues par les deux autres méthodes.

Bien qu'il n'y ait guère d'argument décisif qui incite à choisir une méthode plutôt qu'une autre, les approches logit modifiée et log-quadratique sont, en théorie, meilleures que les approches plus anciennes pour ajuster une table-type de mortalité. En outre, comme nous l'avons souligné plus haut dans la section Mise en garde, certaines constatations empiriques suggèrent qu'elles sont en moyenne plus performantes que la « méthode de la greffe » et celle de Brass. Enfin, dans des contextes où le choix d'une table standard paraît difficile, on peut contourner ce problème en adoptant l'une de ces deux approches. Donc, en règle générale, on préférera le système logit modifié et le système log-quadratique aux méthodes plus anciennes, sauf s'il existe des raisons concrètes de choisir une méthode basée sur un standard appartenant à une famille particulière des tables-types de Princeton ou des Nations Unies.

RÉFÉRENCES

Murray CJ, BD Ferguson, AD Lopez, M Guillot, J Salomon and O Ahmad. 2003. "Modified logit life table system: principles, empirical validation, and application", *Population Studies* 57(2):165–182. doi: <http://dx.doi.org/10.1080/0032472032000097083>
Preston SH, P Heuveline and M Guillot. 2001. *Demography: Measuring and Modelling Population Processes*. Oxford: Blackwell.

Wilmoth JR, S Zureick, V Canudas-Romo, M Inoue and C Sawyer. 2012. "A flexible two-dimensional mortality model for use in indirect estimation", *Population Studies* 66(1):1–28. doi: <http://dx.doi.org/10.1080/00324728.2011.611411>

Chapitre 33 COMBINER DES ESTIMATIONS INDIRECTES DE LA MORTALITÉ DES ENFANTS ET DE LA MORTALITÉ ADULTE POUR PRODUIRE UNE TABLE DE MORTALITÉ

Ian M Timæus et Tom A Moultrie

DESCRIPTION DE LA MÉTHODE

Les méthodes indirectes d'évaluation de la mortalité des enfants et des adultes décrites dans ce manuel fournissent des séries d'estimations qui – en utilisant les approches de la datation dues à Feeney (1980, 1991) pour les enfants, et à Brass et Bamgboye (1981) et Brass (1985) pour les adultes – se rapportent à des dates diverses. Cependant, dans beaucoup d'applications démographiques, il est utile de calculer une table de mortalité abrégée qui reflète la mortalité de l'ensemble des groupes d'âge à une date précise pour la période couverte par ces séries d'estimations. C'est le cas pour la préparation de projections de population ou pour l'évaluation des variations de l'espérance de vie à la naissance ou de la mortalité à travers le temps.

Le type et la période de référence des estimations fournies par les méthodes indirectes les plus importantes sont sommairement présentés dans le tableau 33.1.

Une caractéristique importante des estimations de mortalité adulte obtenues par ces méthodes est qu'il s'agit toujours d'indices conditionnels de survie, qui mesurent la probabilité de survivre d'un âge (par exemple 25 ans dans le cas de la méthode de la survie de la mère) à un autre (par exemple 35 ans). On ne peut pas convertir directement ces indices conditionnels de survie à l'âge adulte en indices non conditionnels. Dès lors, les méthodes d'ajustement d'un modèle relationnel logit exposées en introduction de bien des manuels consacrés à la modélisation ne sont donc pas applicables, et on doit recourir à des techniques d'ajustement plus complexes.

Pour combiner des estimations de mortalité des enfants et de mortalité adulte en une table de mortalité unique qui se rapporte à une date déterminée, on a besoin d'une méthode qui règle les problèmes suivants :

| Méthode | Indices et tranche d'âge habituellement couverte | Période de référence habituelle |
|---|--|---------------------------------|
| Mortalité des enfants : estimation indirecte | $l(1) \dots l(20)$ | de 1 à 15 ans avant l'enquête |
| Mortalité adulte : survie de la mère | $10p_{25} \dots 40p_{25}$ | de 3 à 15 ans avant l'enquête |
| Mortalité adulte : survie du père | $15p_{35} \dots 35p_{35}$ | de 5 à 15 ans avant l'enquête |
| Mortalité adulte : survie des frères et sœurs | $10p_{15} \dots 35p_{15}$ | de 3 à 15 ans avant l'enquête |

TABEAU 33.1 Périodes de référence des indices de mortalité estimés par quelques méthodes indirectes d'évaluation de la mortalité des enfants et de la mortalité adulte

- Les estimations de mortalité adulte, sous forme de probabilités conditionnelles de survie, doivent être converties en probabilités de survie depuis la naissance ;
- Les estimations de mortalité des enfants et de mortalité adulte peuvent impliquer différents niveaux ou schémas par âge de la mortalité, ou des tendances différentes dans le temps, ou les deux à la fois ;
- Certaines données peuvent être défectueuses ou entachées de fluctuations aléatoires qui faussent la tendance générale, ce qui implique que la courbe doit être lissée ou corrigée ;
- Les estimations de mortalité des enfants et de mortalité adulte se rapportent généralement à des dates différentes et peuvent couvrir des périodes de référence différentes ;
- Ni les méthodes d'estimation de la mortalité des enfants ni les méthodes d'estimation de la mortalité adulte ne renseignent sur la mortalité de tous les âges, ce qui implique que l'on ne peut construire une table de mortalité complète qu'en recourant à des modèles.

La méthode présentée ici vise à déterminer les paramètres α et β d'un système relationnel logit associé à une table type (décrite dans le chapitre 31) applicable à une date précise et qui s'ajuste le mieux possible aux données observées utilisées comme inputs. Le calcul d'un modèle à deux paramètres n'est possible que si on dispose, à la date en question, d'estimations indépendantes de la mortalité des enfants et de la mortalité adulte. Si ces données sont disponibles, il est recommandé de calculer un modèle à deux paramètres car rien, a priori, ne justifie généralement l'hypothèse que la structure par âge de la mortalité dans la population étudiée correspond à celle d'une famille particulière de tables-types de mortalité à un paramètre.

En partant des valeurs observées des indices de mortalité des enfants et de mortalité adulte, la méthode permet d'abord de calculer et disposer sur un graphique les valeurs de α (le paramètre de niveau d'un système relationnel logit de mortalité) en fonction de la date de chacun de ces indices, en posant l'hypothèse que β (le paramètre de structure) est égal à 1. Ce graphique des valeurs de α (« *alpha plot* ») sert à identifier les points qui dessinent une tendance cohérente et stable des valeurs de α à travers le temps. Les points retenus sont ensuite utilisés pour calculer par itération les estimations définitives de α et β à la date à laquelle on souhaite rattacher la table de mortalité. On peut alors employer ces valeurs α et β pour calculer une table de mortalité ajustée à partir de la table qui sert de standard. La méthode autorise les deux paramètres α et β à varier en fonction du temps, mais leur impose de varier linéairement (Timæus 1990).

On peut utiliser cette méthode pour calculer des tables de mortalité abrégées à partir d'indices de mortalité des enfants par sexe obtenus par la méthode indirecte, en analysant des données sur les enfants nés vivants et les enfants survivants (chapitre 16), et d'indices de mortalité adulte par sexe obtenus en appliquant soit la méthode des proportions d'orphelins à un seul recensement (chapitre 22), soit la méthode indirecte basée sur la survie des frères et sœurs (chapitre 23). Normalement, les estimations de mortalité des enfants et de mortalité adulte obtenues par une méthode directe ou par l'analyse de deux recensements se rapportent à une année ou une période précise. On peut ajuster des tables-types de mortalité à des couples d'estimations de la mortalité des enfants et de la mortalité adulte qui se rapportent à la même date ou période, en appliquant les méthodes décrites dans la section sur l'ajustement d'une table-type de mortalité à un couple d'estimations de la mortalité des enfants et de la mortalité adulte (chapitre 32).

DONNÉES REQUISES ET HYPOTHÈSES

Tabulations nécessaires

- Une série d'estimations indirectes de mortalité des enfants par sexe, avec leurs dates, provenant de l'analyse de données sur les enfants nés vivants et survivants ;
- Une série d'estimations de mortalité adulte par sexe, avec leurs dates, calculées soit par une méthode indirecte basée sur la survie des frères et sœurs, soit par la méthode des proportions d'orphelins déclarés à un seul recensement.

En principe, l'approche adoptée pour traiter ce type de données pourrait être étendue au calcul de tables de mortalité pour des populations disposant de séries multiples et partiellement chevauchantes d'estimations indirectes de mortalité des enfants et de mortalité adulte. Mais le tableur en ligne associé à ce thème ne peut traiter qu'une seule série de données pour la mortalité des enfants et une seule pour les adultes.

Hypothèses

Dans la méthode que nous présentons ici, la table de mortalité ajustée est basée sur une table-type de mortalité qui sert de standard. Cette table standard est supposée avoir une structure par âge de la mortalité similaire à celle de la population étudiée. En particulier, le rapport entre mortalité des enfants et mortalité adulte dans la table standard devrait être le même que celui qui existe entre les estimations indirectes auxquelles le modèle doit être ajusté. Des conseils pour le choix d'une

table standard adéquate sont donnés dans le chapitre 31, qui explique aussi les mécanismes de base du système relationnel logit associé à des tables-types de mortalité. Il n'est pas obligatoire de choisir la table standard dans la famille de tables-types qui est à la base des coefficients utilisés pour calculer les estimations indirectes de mortalité des enfants : la famille de tables-types qui s'adapte le mieux à la structure par âge de la mortalité *dans l'enfance* peut ne pas être la plus indiquée pour représenter le rapport entre mortalité des enfants et mortalité des adultes dans la même population.

MISE EN GARDE

La plausibilité des tables ajustées construites par cette méthode dépend de la pertinence de la table standard choisie pour la population étudiée. Par exemple, dans une population touchée par une épidémie de VIH/sida, le rapport entre mortalité des enfants et mortalité adulte, et la structure par âge détaillée de la mortalité, s'écartent fortement de ceux qui caractérisent les tables-types les plus couramment utilisées. Par conséquent, il est déconseillé d'utiliser régulièrement cette méthode en pareil cas, ainsi que quand on n'a pu identifier aucune table standard reproduisant correctement le rapport entre mortalité des enfants et mortalité adulte.

APPLICATION DE LA MÉTHODE

La mise en œuvre de la méthode suit plusieurs étapes.

Étape 1 : Déterminer la date à laquelle doit se rapporter la table de mortalité souhaitée

Pour éviter les risques associés aux extrapolations abusives lors de la détermination de α et β , la table de mortalité devrait se rapporter à une date située à l'intérieur de la période couverte par les estimations de mortalité des enfants et de mortalité adulte qui entrent dans l'analyse. Nous appellerons cette date D .

La date exacte à laquelle doit être située la table peut être déterminée par l'usage auquel elle est destinée. Mais, idéalement, il faudrait choisir une date à laquelle la mortalité des enfants et celle des adultes semblent correctement estimées. Par exemple, si les estimations de la mortalité des enfants remontant le plus loin dans le passé paraissent biaisées à la baisse par la sous-déclaration des décès d'enfants, ou si les estimations de la mortalité adulte du passé le plus récent basées sur la survie des parents semblent biaisées à la baisse par l'effet d'adoption, on devrait s'abstenir de calculer une table de mortalité se rapportant aux dates couvertes par ces estimations défectueuses. Malheureusement, ces considé-

rations conduisent parfois à conclure que les données disponibles ne sont pas assez fiables pour servir de base à la construction d'une table de mortalité !

Si on a besoin d'une table de mortalité pour une date postérieure, ou éventuellement antérieure, à la période couverte par les indices disponibles, on peut envisager une légère extrapolation en dehors des limites de cette période. Cette extrapolation ne devrait pas dépasser trois années avant la plus ancienne estimation de l'un ou l'autre des deux indices de mortalité, ou trois années après le plus ancien des deux indices les plus récents.

Étape 2 : Choisir une table standard pour servir de base au calcul de la table de mortalité ajustée

Le tableur associé à ce chapitre permet de choisir entre neuf tables standards par sexe : les cinq tables-types des Nations Unies (modèle général, Asie du Sud, Extrême-Orient, Amérique Latine, Chili) et les quatre tables-types régionales de Princeton (Nord, Sud, Est, Ouest). Toutes les tables standards ont une espérance de vie à la naissance de 60 ans. Le calcul de ces logits est décrit dans le chapitre 31, et un tableur contenant leurs valeurs numériques peut être téléchargé sur le site web associé à cet ouvrage.

Lors de la recherche d'une table standard, le premier objectif devrait être d'en choisir une dans laquelle le rapport entre mortalité des enfants et mortalité adulte est à peu près le même que celui qui existe entre leurs valeurs observées. Pratiquement, si le paramètre β de la table finale calculée est inférieur à 0,75 ou supérieur à 1,25, on devrait au moins envisager le choix d'une autre table standard. Dans des situations plus problématiques, les tables-types dans lesquelles β est inférieur à 0,6 ou supérieur à 1,4 n'ont guère de chances de représenter correctement la mortalité observée. Un autre objectif de la recherche d'une table standard devrait être d'en choisir une qui possède d'autres caractéristiques de la population étudiée, telles que le rapport entre mortalité infantile et mortalité juvénile. Les caractéristiques des diverses familles de tables-types de Princeton et des Nations Unies sont brièvement décrites dans le chapitre 31.

Étape 3 : Porter sur un graphique les valeurs de α (β étant supposé égal à 1) calculées à partir des indices de mortalité observés, en fonction de leurs dates de référence

Quand β (le paramètre qui caractérise la structure dans un système relationnel de tables-types de mortalité) est égal à 1, le système peut s'écrire :

$$Y(x) = \alpha + Y^s(x)$$

où $Y(x)$ est la transformation logit de $l(x)$:

$$Y(x) = \text{logit}(l(x)) = \frac{1}{2} \ln \left(\frac{1 - l(x)}{l(x)} \right) = -\frac{1}{2} \ln \left(\frac{1 - q(x)}{q(x)} \right).$$

Pour la mortalité des enfants, le calcul de la série des valeurs de α à partir des indices estimés est très simple. On soustrait les logits de la table standard choisie pour les âges 1, 2, 3, 5, 10, 15 et 20 ans de ceux des quotients estimés ($q(1)$, $q(2)$, $q(3)$, $q(5)$, $q(10)$, $q(15)$ et $q(20)$) :

$$\alpha^{child} = Y(x) - Y^s(x).$$

Pour la mortalité adulte, le calcul de α est plus compliqué, car les probabilités de survie fournies par les méthodes d'estimation sont des probabilités conditionnelles, la condition étant d'être en vie à un âge initial fixé. La formule de calcul de α est la suivante :

$$\alpha^{adults} = \frac{1}{2} \left\{ \ln(1 - {}_n p_x) - \ln[{}_n p_x \cdot \exp(2Y^s(x + n)) - \exp(2Y^s(x))] \right\},$$

où x est l'âge initial de la probabilité conditionnelle de survie (25 ans pour la méthode de la survie des mères) et n est l'intervalle de temps sur lequel est mesurée la probabilité de survie, qui dépend du groupe d'âge auquel appartient la personne interrogée. (On trouvera le calcul de cette expression à la fin de ce chapitre.)

Les estimations de α (séparément pour les enfants et pour les adultes) sont ensuite portées sur un même graphique en fonction de leurs dates respectives.

Étape 4 : Éliminer du graphique des valeurs de α les points qui s'écartent trop de la tendance générale

Pour estimer les valeurs des paramètres α et β à une date déterminée, la méthode exige qu'ils suivent une tendance linéaire. Pour satisfaire cette condition, il faut d'abord que chacune des deux séries de points représentant les valeurs de α (l'une pour les enfants et l'autre pour les adultes) en fonction du temps, calculées à l'étape 3, soit alignée.

Certains points α de l'une ou l'autre série, calculés au moyen des deux formules ci-dessus, peuvent s'écarter de l'alignement avec les autres pour diverses raisons. Premièrement, les tendances sous-jacentes de la mortalité peuvent

se révéler nettement non linéaire. Mais c'est assez peu probable, car les données observées couvrent des périodes relativement courtes et les méthodes d'estimation indirectes ont tendance à gommer les fluctuations à court terme de la mortalité. Cependant, même si le graphique de diagnostic dont les points ont été calculés plus haut montre que l'on est dans ce cas, il reste possible d'obtenir une table de mortalité ajustée correcte en la calculant à une date à laquelle les tendances linéaires imposées aux paramètres par la méthode coupent la courbe dessinée par les points du graphique. Deuxièmement, les séries peuvent être assez irrégulières à cause des erreurs d'échantillonnage et de déclaration, telles que les erreurs sur l'âge. Si c'est le seul défaut des données, on les inclura normalement toutes dans l'analyse, et on fera confiance à la procédure d'ajustement linéaire pour niveler ces fluctuations.

Troisièmement, les estimations indirectes sont sujettes aux biais résultant des mauvaises réponses données par les personnes interrogées aux questions clés, ou au non-respect des hypothèses associées aux méthodes concernées. Les erreurs possibles sont détaillées dans les pages consacrées aux diverses méthodes, et nous y renvoyons le lecteur qui voudrait des conseils pour identifier les signes pouvant faire penser que certains points sont biaisés et doivent être écartés de la procédure d'ajustement. Il est relativement fréquent que le point qui correspond au groupe d'âge 15-19 ans dans la méthode d'estimation de la mortalité des enfants soit biaisé vers le haut, et que le point qui correspond au groupe d'âge 5-14 ans dans la méthode de la survie des parents avec un seul recensement soit biaisé vers le bas. Il sera donc souvent nécessaire d'exclure ces points de la procédure d'ajustement.

Une quatrième explication possible du non-alignement des valeurs calculées de α est que la table standard choisie pour le calcul des estimations originales peut ne pas être la plus appropriée. Si c'est le cas, il peut être nécessaire de recalculer les indices ou la table ajustée sur la base d'un autre standard (comme expliqué à l'étape 2).

Une fois choisies les valeurs de α et β , pour les enfants et pour les adultes, à retenir pour la procédure d'ajustement, la suite des calculs se déroule automatiquement.

Étape 5 : Déterminer la tendance de β par itération

Les valeurs de β n'étant pas faciles à calculer manuellement par itération, le manuel d'exercices a été programmé pour faire ces calculs. Pour activer la procédure d'itération, assurez-vous que Microsoft Excel est configuré comme il faut.

Pour ce faire, cliquez sur « *File* → *Options* → *Formulas* » puis contrôlez la fenêtre « *Enable iterative calculation* ». Fixer le nombre maximum d'itérations à 1 000 et le maximum de variation à 0,00001 est plus que suffisant pour parvenir à une solution.

La procédure itérative qui ajuste α et β est décrite dans la section consacrée à la présentation mathématique de la méthode. Les principales contraintes imposées à cette procédure sont les suivantes :

- Quelles que soient les valeurs originales de x et n qui entrent dans l'estimation de $q(x)$ et de ${}_np_b$ pour les enfants et pour les adultes à la date choisie, β se calcule toujours sur la base de la probabilité de survie de 15 à 60 ans de la table standard.
- α et β peuvent tous les deux varier au cours du temps, mais on suppose qu'ils varient linéairement.

L'ensemble de ces deux hypothèses réduit l'effet de distorsion que les erreurs sur les indices observés et les écarts mineurs entre les structures par âge de la mortalité de la population et de la table standard peuvent avoir sur la table de mortalité ajustée (Timæus 1990). Par contre, si on utilise la méthode, décrite dans le chapitre 32, d'ajustement d'une table-type logit à deux paramètres à un couple d'estimations indirectes récentes de la mortalité des enfants et de la mortalité adulte qui se rapportent à peu près à la même date, mais ne mesurent la mortalité que sur une partie de l'éventail des âges (Brass 1975, 1985), par exemple $q(2)$ et ${}_{10}p_{25}$, on obtient souvent des valeurs extrêmes de β qui aboutissent à des tables ajustées peu vraisemblables.

Étape 6 : Examiner les valeurs ajustées de α

L'avant-dernière étape consiste à examiner le graphique des valeurs de α qui découlent de la procédure itérative d'ajustement, c'est-à-dire le deuxième graphique de la feuille des « *alpha plots* » de la feuille Excel associée. C'est ce second graphique, présentant les estimations de α réajustées en fonction du niveau et de la tendance de β , qui permet de vérifier l'hypothèse que α a suivi une évolution linéaire. En outre, si la table standard à laquelle les données ont été adaptées est adéquate, les séries d'estimations de la mortalité des enfants et de la mortalité adulte devraient être proches l'une de l'autre sur ce graphique.

Étape 7 : Construction d'une table de mortalité ajustée

Une fois identifiées, au moyen de la procédure itérative d'ajustement, les meilleures tendances linéaires d'évolution temporelle de α et β , les valeurs ajustées de ces deux para-

mètres à la date choisie pour la table de mortalité ajustée, D , se calculent comme suit :

$$\alpha^* = Z(\alpha) + D.S(\alpha)$$

$$\beta^* = Z(\beta) + D.S(\beta).$$

On calcule la table de mortalité abrégée ajustée à partir de ces valeurs de α^* et β^* et de la table standard, au moyen de la formule :

$$l^*(x) = \frac{1}{1 + \exp(2(\alpha^* + \beta^*.Y^s(x)))}.$$

EXEMPLE

L'exemple détaillé que nous présentons ici utilise des données sur la population féminine de la République Dominicaine. Les estimations indirectes de mortalité des filles découlent des données sur les enfants nés et les enfants survivants recueillies par une EDS réalisée en 2002. Les estimations indirectes de mortalité des femmes adultes se basent sur les déclarations de survie des mères enregistrées lors du recensement effectué pendant la même année. Les données de base sont présentées dans le tableau 33.2.

TABEAU 33.2 Données de base utilisées pour combiner les estimations respectives de la mortalité des enfants et de la mortalité adulte. République Dominicaine, 2002

| Mortalité des enfants (EDS 2002) | | | Mortalité adulte (recensement de 2002) | | |
|-------------------------------------|--------|---------|---|--------------|---------|
| x | $q(x)$ | Date | n | ${}_np_{25}$ | Date |
| 1 | 0,0338 | 2001,71 | 10 | 0,9858 | 1999,23 |
| 2 | 0,0429 | 2000,24 | 15 | 0,9801 | 1997,07 |
| 3 | 0,0355 | 1998,48 | 20 | 0,9680 | 1995,13 |
| 5 | 0,0467 | 1996,43 | 25 | 0,9479 | 1993,43 |
| 10 | 0,0619 | 1994,16 | 30 | 0,9214 | 1992,02 |
| 15 | 0,0710 | 1991,52 | 35 | 0,8872 | 1991,00 |
| 20 | 0,0799 | 1987,99 | 40 | 0,8373 | 1990,51 |

Étape 1 : Déterminer la date à laquelle doit se rapporter la table de mortalité souhaitée

Dans le cas des données de la République Dominicaine, le tableur en ligne permet de calculer une table de morta-

lité pour des dates situées entre la plus ancienne des deux dates 1987,99 – 3 et 1990,51 – 3 et la plus ancienne des deux dates 2001,71 + 3 et 1999,23 + 3, c’est-à-dire entre 1984,99 et 2002,23.

Dans cet exemple, nous calculerons une table de mortalité pour le milieu de 1997, soit 1997,5.

Étape 2 : Choisir une table standard pour servir de base au calcul de la table de mortalité ajustée

Étant donné l’origine géographique des données, il est raisonnable de faire l’hypothèse (au moins pour commencer) que la mortalité des femmes dominicaines a une structure par âge du type de celle de la table standard féminine Amérique Latine des Nations Unies. Les logits de la table de mortalité choisie comme standard sont présentés dans le tableau 33.3.

TABLEAU 33.3 Logits de la table de mortalité féminine Amérique Latine des Nations Unies ayant une espérance de vie de 60 ans

| Âge (x) | $Y^s(x) = \text{logit}(I(x))$ |
|---------|-------------------------------|
| 0 | |
| 1 | – 1,2375 |
| 2 | – 1,1006 |
| 3 | – 1,0398 |
| 4 | – 1,0046 |
| 5 | – 0,9815 |
| 10 | – 0,9304 |
| 15 | – 0,9054 |
| 20 | – 0,8735 |
| 25 | – 0,8313 |
| 30 | – 0,7828 |
| 35 | – 0,7285 |
| 40 | – 0,6670 |
| 45 | – 0,6005 |
| 50 | – 0,5248 |
| 55 | – 0,4356 |
| 60 | – 0,3230 |
| 65 | – 0,1795 |

Notons que, si, dans la feuille Excel associée (disponible en ligne), on passe des tables des Nations Unies à celles de Princeton, ou vice versa, on doit forcer le tableur à refaire les calculs en faisant passer la case « *Recalculate* » (**B3** sur la feuille **Method**) de « *True* » à « *False* » et inversement. Omettre cette manipulation provoquerait une erreur.

Étape 3 : Porter sur un graphique les valeurs de α (β étant supposé égal à 1) calculées à partir des indices de mortalité observés, en fonction de leurs dates de référence

En prenant les données dominicaines du tableau 33.2 et une table de mortalité de la famille Amérique Latine des Nations Unies comme table standard, on calcule comme suit la valeur de α pour la mortalité des enfants quand $x = 3$:

$$\begin{aligned}\alpha^{child} &= -\frac{1}{2} \ln \left(\frac{1 - q(3)}{q(3)} \right) - \text{logit}(I^s(3)) \\ &= -\frac{1}{2} \ln \left(\frac{1 - 0,0355}{0,0355} \right) + 1,0398 \\ &= -0,6112.\end{aligned}$$

Cette valeur de α se rapporte à la date 1998,48, comme l’indique le tableau 33.1. On calcule de la même manière les valeurs de α pour les autres indices de mortalité des enfants avec leurs dates respectives.

En prenant les données de mortalité adulte du tableau 33.2 et la même table standard, le calcul de α quand n vaut 25 est le suivant :

$$\begin{aligned}\alpha^{adults} &= \frac{1}{2} \left\{ \ln(1 - {}_{25}p_{25}) - \ln[{}_{25}p_{25} \cdot \exp(2Y^s(50)) - \exp(2Y^s(25))] \right\} \\ &= \frac{1}{2} \left\{ \ln(1 - 0,9479) - \ln[0,9479 \exp(2(-0,5248)) - \exp(2(-0,8313))] \right\} \\ &= -0,5021.\end{aligned}$$

Cette valeur de α se rapporte à la date 1993,43. On calcule de la même manière les valeurs de α pour les autres indices de mortalité adulte avec leurs dates respectives.

Le tableau 33.4 récapitule ces estimations de α et leurs dates.

Avec toutes les valeurs estimées de α pour la mortalité des enfants et celle des adultes et leurs dates de référence respectives, on construit le graphique de la figure 33.1.

Étape 4 : Éliminer du graphique des valeurs de α les points qui s’écartent trop de la tendance générale

Le chapitre 16 explique que la mesure indirecte la plus récente de la mortalité des enfants, basée sur les déclarations des femmes âgées de 15 à 19 ans, a tendance à être biaisée vers le haut, parce que les mères très jeunes constituent un groupe spécifique dont les enfants font face à des risques de

mortalité élevées car, entre autres raisons, ces mères viennent généralement de milieux socialement défavorisés. Cette première estimation est presque toujours ignorée quand on détermine la tendance de la mortalité des enfants sur la base d'estimations indirectes, et c'est ce que nous avons fait dans cet exemple.

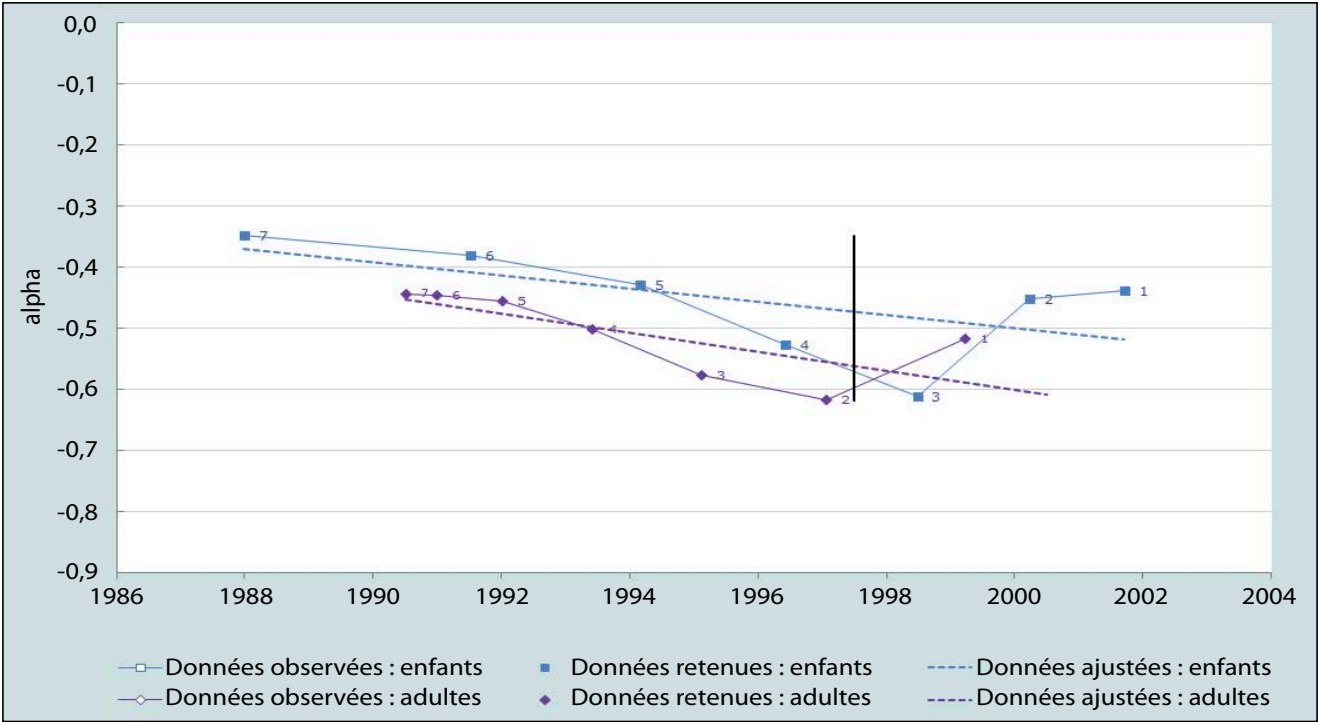
L'estimation la plus récente de la mortalité adulte, basée sur les déclarations des enfants de 5 à 9 ans quant à la survie de leur mère, dans le cadre de la méthode des pro-

portions d'orphelins déclarées dans un seul recensement, sous-estime aussi la mortalité dans bien des cas. Mais en République Dominicaine, cette donnée s'avère beaucoup plus élevée que ce que l'on pouvait attendre au vu de la tendance indiquée par les autres estimations de mortalité adulte féminine. Cela pourrait résulter d'un fort sous-enregistrement des âges des enfants, ou signifier que les modèles impliqués dans le processus d'estimation ne conviennent pas pour cette population. Quoi qu'il en soit, nous avons

TABLEAU 33.4 Valeurs de α , avec leurs dates de référence. République Dominicaine, 2002

| Enfants | | | Adultes | | |
|-----------------|----------|---------|-----------------|----------|---------|
| Indice original | α | Date | Indice original | α | Date |
| $q(1)$ | - 0,4389 | 2001,71 | $10p_{25}$ | - 0,5176 | 1999,23 |
| $q(2)$ | - 0,4519 | 2000,24 | $15p_{25}$ | - 0,6183 | 1997,07 |
| $q(3)$ | - 0,6112 | 1998,48 | $20p_{25}$ | - 0,5779 | 1995,13 |
| $q(5)$ | - 0,5266 | 1996,43 | $25p_{25}$ | - 0,5021 | 1993,43 |
| $q(10)$ | - 0,4288 | 1994,16 | $30p_{25}$ | - 0,4567 | 1992,02 |
| $q(15)$ | - 0,3803 | 1991,52 | $35p_{25}$ | - 0,4463 | 1991,00 |
| $q(20)$ | - 0,3483 | 1987,99 | $40p_{25}$ | - 0,4436 | 1990,51 |

FIGURE 33.1 Premier graphique des valeurs de α et de leurs dates de référence pour la mortalité des enfants et celle des adultes. République Dominicaine



décidé d'ignorer cette estimation anormale. Ainsi, le point le plus récent de chacune des deux séries d'estimations a été écarté du processus d'ajustement linéaire des α en vidant les cellules correspondantes de la feuille « *Alpha plots* » de la feuille Excel disponible en ligne.

Les autres estimations obtenues par la méthode des proportions d'orphelins sont cohérentes entre elles et indiquent que la mortalité féminine a rapidement chuté en République Dominicaine pendant les années 1990. Les indices de mortalité des enfants indiquent aussi une baisse de mortalité, mais les estimations les plus récentes sont quelque peu contradictoires. Selon les 3^e et 4^e points, basés sur les déclarations des mères de 25 à 34 ans, le rythme de la baisse de la mortalité des enfants se serait accéléré dans la seconde moitié des années 1990. Cependant, le 2^e point, fondé sur les déclarations des femmes de 20 à 24 ans, suggère plutôt un ralentissement. En l'absence de certitudes sur la nature des erreurs qui ont conduit à ces incohérences, nous avons décidé d'écarter ces trois points de notre analyse.

La sélection finale aboutit au graphique de la figure 33.2. Ce graphique met en évidence la cohérence des points 2 à 7 de la mortalité adulte et montre qu'une droite de régression calculée sur les points 2 à 7 de la mortalité des enfants non seulement passe au milieu des estimations les plus récentes, mais s'ajuste aussi bien aux trois estimations plus anciennes.

Notons que, sur la figure 33.2, les valeurs de α basées sur les estimations de la mortalité adulte sont inférieures à celles qui proviennent des estimations de la mortalité des enfants et s'en écartent avec le temps. Cela signifie que, comparativement au standard Amérique Latine des Nations Unies, la mortalité adulte des années 1990 en République Dominicaine était basse et diminuait plus rapidement que la mortalité des enfants. Le paramètre β des tables ajustées pour cette population sera alors inférieur à 1 et diminuera au fil du temps.

Étape 5 : Déterminer la tendance de β par itération

Le tableur calcule par itération les valeurs ajustées de α et β à la date voulue (1997,5). Les estimations de α^* et β^* sont respectivement $-0,658$ et $0,849$. Elles signifient que le niveau de la mortalité en République Dominicaine est un peu plus faible que dans la table standard Amérique Latine ($\alpha < 0$), et que la mortalité est légèrement plus forte aux jeunes âges et légèrement plus faible aux âges élevés ($\beta < 1$) que dans ce

standard. L'estimation de β^* est suffisamment proche de 1 pour ne pas susciter d'inquiétude quant au choix du standard qui a été fait à l'étape 2.

Étape 6 : Examiner les valeurs ajustées de α

L'avant-dernière étape consiste à examiner le graphique des valeurs de α obtenu après la procédure itérative d'ajustement, c'est-à-dire le deuxième graphique de la feuille des « *alpha plots* » du manuel d'exercices en ligne.

La figure 33.3 montre qu'il y a maintenant une étroite proximité entre les α des enfants et ceux des adultes pendant la plus grande partie des années 1990. La mortalité était en baisse à tous les âges, bien que β ait diminué d'environ 0,95 à 0,85 entre le début des années 1990 et le milieu de 1997. C'est bien ce à quoi il fallait s'attendre, puisque nous avons déjà vu que la mortalité adulte en République Dominicaine, à cette époque, diminuait plus rapidement que dans les valeurs logits basées sur le standard Amérique Latine des tables-types des Nations Unies.

Les courbes des enfants et des adultes restent très proches l'une de l'autre en 1997, ce qui reflète la proximité entre la valeur de β à cette époque (0,85) et sa valeur centrale, 1. Les deux estimations de α de 1997 ne s'écarteraient franchement l'une de l'autre que si β , à la même époque, était très différent de 1. En ce cas, il serait conseillé de chercher une table standard plus appropriée à la situation de la population étudiée.

Si les deux séries d'estimations suivaient des tendances très différentes et ne se coupaient pas, ou divergeaient rapidement après leur intersection, ou si le profil de l'une d'elles ou des deux n'était pas du tout linéaire, cela signifierait de nouveau que la table standard utilisée ne convenait pas ou, plus probablement, que l'une ou l'autre série était sérieusement biaisée par des erreurs dans les données, ce qui rendrait impossible toute tentative de les concilier l'une avec l'autre.

Étape 7 : Construction d'une table de mortalité ajustée

On calcule la table de mortalité abrégée à partir des valeurs ajustées $\alpha^* = -0,658$ et $\beta^* = 0,849$ à la date choisie et de la table standard (présentée dans le tableau 33.3) au moyen de la formule suivante :

$$l^*(x) = \frac{1}{1 + \exp(2(\alpha^* + \beta^* \cdot Y^s(x)))}.$$

FIGURE 33.2 Graphique définitif des valeurs de α et de leurs dates de référence, pour la mortalité des enfants et celle des adultes. République Dominicaine

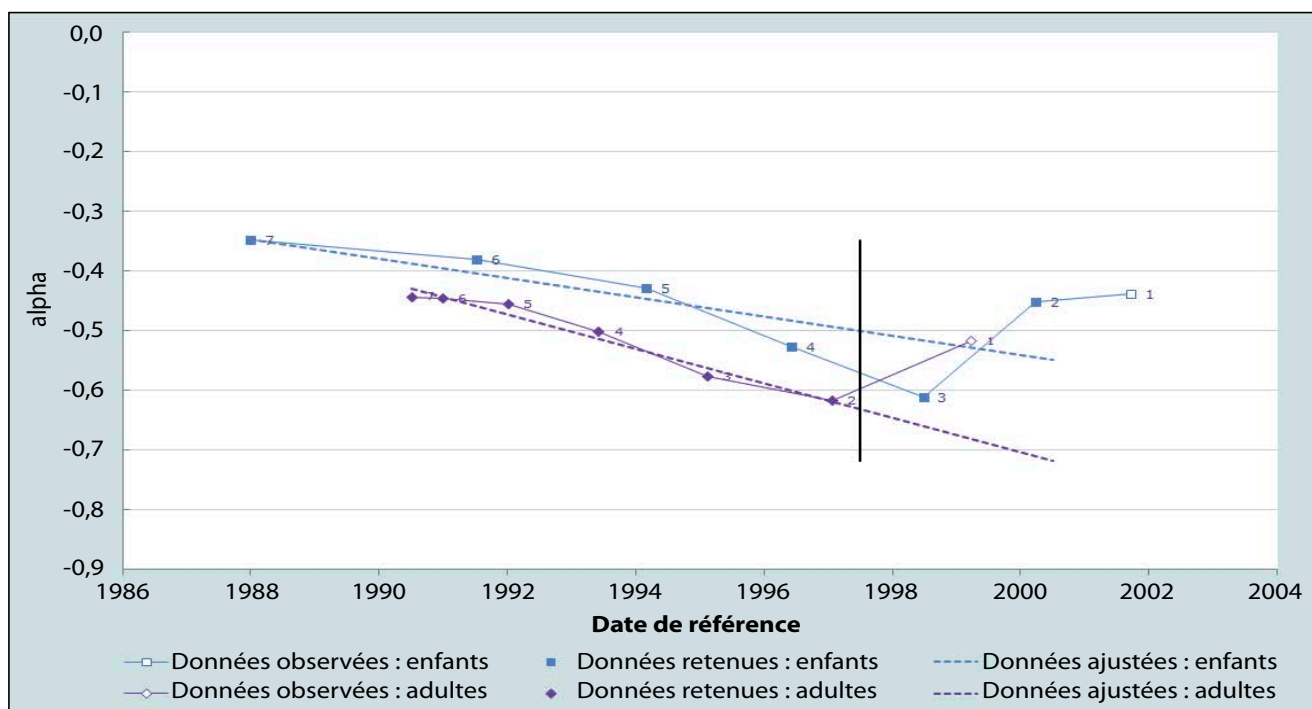
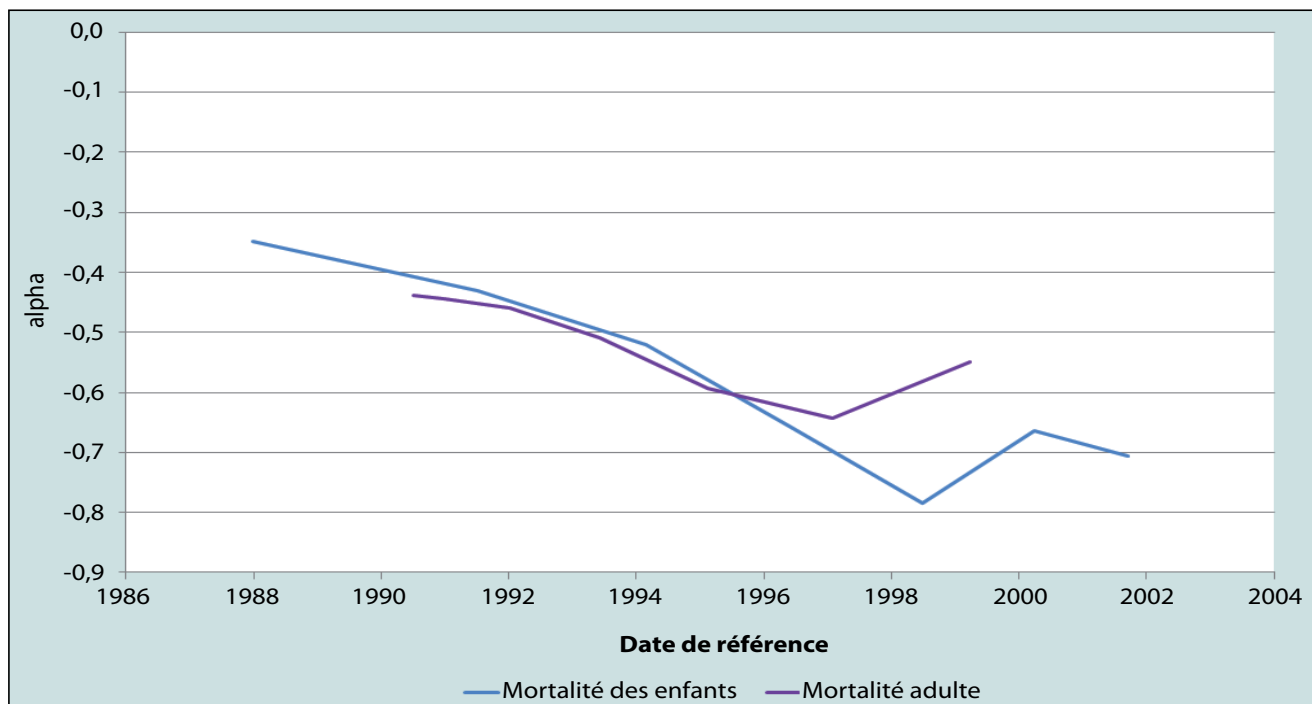


FIGURE 33.3 Graphique des valeurs de α après l'ajustement de β par itération. République Dominicaine



La table ajustée finale est détaillée dans le tableau 33.5. L'espérance de vie à la naissance est de 76,6 ans, à comparer avec l'estimation des Nations Unies pour la même période de cinq ans, 73,1 ans (UN Population Division 2013).

TABLEAU 33.5 Table de mortalité féminine ajustée. République Dominicaine, milieu de 1997

| Âge (x) | <i>l</i> (x) |
|---------|--------------|
| 0 | 1,0000 |
| 1 | 0,9683 |
| 2 | 0,9603 |
| 3 | 0,9561 |
| 4 | 0,9536 |
| 5 | 0,9518 |
| 10 | 0,9476 |
| 15 | 0,9455 |
| 20 | 0,9426 |
| 25 | 0,9386 |
| 30 | 0,9337 |
| 35 | 0,9278 |
| 40 | 0,9204 |
| 45 | 0,9118 |
| 50 | 0,9008 |
| 55 | 0,8865 |
| 60 | 0,8657 |
| 65 | 0,8348 |
| 70 | 0,7863 |
| 75 | 0,7120 |
| 80 | 0,6029 |
| 85 | 0,4452 |
| 90 | 0,2587 |
| 95 | 0,1024 |
| 100 | 0,0242 |

DESCRIPTION DÉTAILLÉE DE LA MÉTHODE

Le tableur en ligne applique la méthode en suivant les étapes décrites ci-dessus. Cette section expose en détail la procédure itérative employée pour calculer les valeurs finales de α et β .

Le premier principe qui sous-tend cette procédure est que la table de mortalité obtenue devra recouper parfaitement les données observées à 15 ans et à 60 ans. La première de ces deux contraintes garantit que la mortalité des enfants et des adolescents est correctement représentée ; la

combinaison des deux conditions garantit que la mortalité adulte entre 15 et 60 ans est proche de celle qu'impliquent les estimations de mortalité adulte utilisées pour ajuster la table.

Procédure d'ajustement

Une fois désignés les points à retenir pour la suite des calculs (voir l'étape 4), la méthode vise à déterminer le modèle de régression linéaire qui représente le mieux l'évolution dans le temps de α pour les enfants, conditionné par la tendance de β , et le modèle de régression linéaire qui représente le mieux l'évolution de β pour les adultes, conditionné par la tendance de α pour les enfants.

En partant de l'hypothèse que $\beta = 1$, on peut calculer un α^{child} correspondant à chaque estimation de mortalité des enfants au moyen de l'équation présentée à l'étape 3 de l'exemple. Comme chaque estimation de α^{child} est associée à sa date de référence (T), on peut ajuster une droite de régression linéaire sur les estimations incluses dans la procédure d'ajustement par rapport au temps pour obtenir la pente $S(\alpha)$ et l'ordonnée à l'origine $Z(\alpha)$.

Les coefficients de la régression obtenus peuvent ensuite servir à calculer une valeur ajustée de α (α^*) pour les dates auxquelles se rapportent les estimations de mortalité adulte :

$$\alpha^* = Z(\alpha) + T.S(\alpha).$$

Avec ces valeurs ajustées de α^{child} , on peut calculer $Y(15)$ aux mêmes dates :

$$Y(15) = \alpha^* + \beta^*Y^s(15),$$

c'est la première itération avec $\beta^* = 1$.

Toujours en faisant l'hypothèse que $\beta = 1$, on peut aussi calculer α^{adult} à partir des estimations conditionnelles de survie des adultes qui ont été introduites dans la procédure d'ajustement, en utilisant l'équation de l'étape 3, et employer ces valeurs de α^{adult} pour calculer les estimations correspondantes de $_{45}q_{15}$. En multipliant la valeur de $l(15)$, déduite de l'estimation de la mortalité des enfants, par une estimation de $_{45}q_{15}$ à la même date, basée sur les données de mortalité adulte, on obtient une estimation non conditionnelle de $l(60)$ et donc de $Y(60)$:

$$Y(60) = -\frac{1}{2}\ln\left(\frac{l(60)}{1-l(60)}\right) = -\frac{1}{2}\ln\left(\frac{l(15)_{\cdot 45}p_{15}}{1-l(15)_{\cdot 45}p_{15}}\right).$$

L'estimation de $l(15)$ se déduit de $Y(15)$:

$$l(15) = \frac{1}{1 + \exp(2Y(15))},$$

tandis que celle de ${}_{45}p_{15}$ se calcule à partir des valeurs de α et β ajustées aux estimations de mortalité adulte :

$${}_{45}p_{15} = \frac{l(60)}{l(15)} = \frac{(1 + \exp(2(\alpha^{adult} + \beta^*(Y^s(15))))}{(1 + \exp(2(\alpha^{adult} + \beta^*(Y^s(60))))}$$

où $Y^s(x)$ représente le logit de $l(x)$ dans la table de mortalité standard (c'est-à-dire avec $\beta = 1$ et $\alpha = 0$) et où β^* , pour cette première itération, est égal à 1.

Ayant estimé une série de valeurs de $Y(60)$ pour les dates auxquelles se rapportent les estimations de mortalité adulte, on peut maintenant calculer des valeurs corrigées de β pour ces dates :

$$\beta = \frac{Y(60) - Y(15)}{Y^s(60) - Y^s(15)}.$$

Comme chacune de ces valeurs corrigées de β se rapporte à une date particulière, on peut procéder à une régression de β par rapport au temps (T) pour obtenir la pente $S(\beta)$ et l'ordonnée à l'origine $Z(\beta)$ d'une droite de régression linéaire, qui servira ensuite au calcul d'une valeur ajustée de β (β^*) pour chaque point, qu'il s'agisse de mortalité des enfants ou de mortalité adulte, et chaque date correspondante (T) :

$$\beta^* = Z(\beta) + T.S(\beta).$$

RÉFÉRENCES

- Brass W. 1975. *Methods for Estimating Fertility and Mortality from Limited and Defective Data*. Chapel Hill: International Program of Laboratories for Population Statistics.
- Brass W. 1985. *Advances in Methods for Estimating Fertility and Mortality from Limited and Defective Data*. London: London School of Hygiene & Tropical Medicine.
- Brass W and EA Bamgboye. 1981. *The Time Location of Reports of Survivorship: Estimates for Maternal and Paternal Orphanhood and the Ever-widowed*. London: London School of Hygiene & Tropical Medicine.
- Feeney G. 1980. "Estimating infant mortality trends from child survivorship data", *Population Studies* **34**(1):109-128. doi: <http://dx.doi.org/10.1080/00324728.1980.10412839>
- Feeney G. 1991. "Child survivorship estimation: Methods and data analysis", *Asian and Pacific Population Forum* **5**(2-3):51-55, 76-87. <http://hdl.handle.net/10125/3600>.
- Timæus IM. 1990. "Advances in the Measurement of Adult Mortality from Data on Orphanhood." Unpublished PhD thesis, London: University of London.
- UN Population Division. 2013. *World Population Prospects: The 2012 Revision*. New York: United Nations, Department of Economic and Social Affairs. <http://esa.un.org/wpp/index.htm>

Le premier cycle d'itération est ainsi achevé. On peut alors calculer des estimations corrigées de α^{child} , qui tiennent compte du fait que β peut maintenant s'écarter de 1, avec la formule suivante :

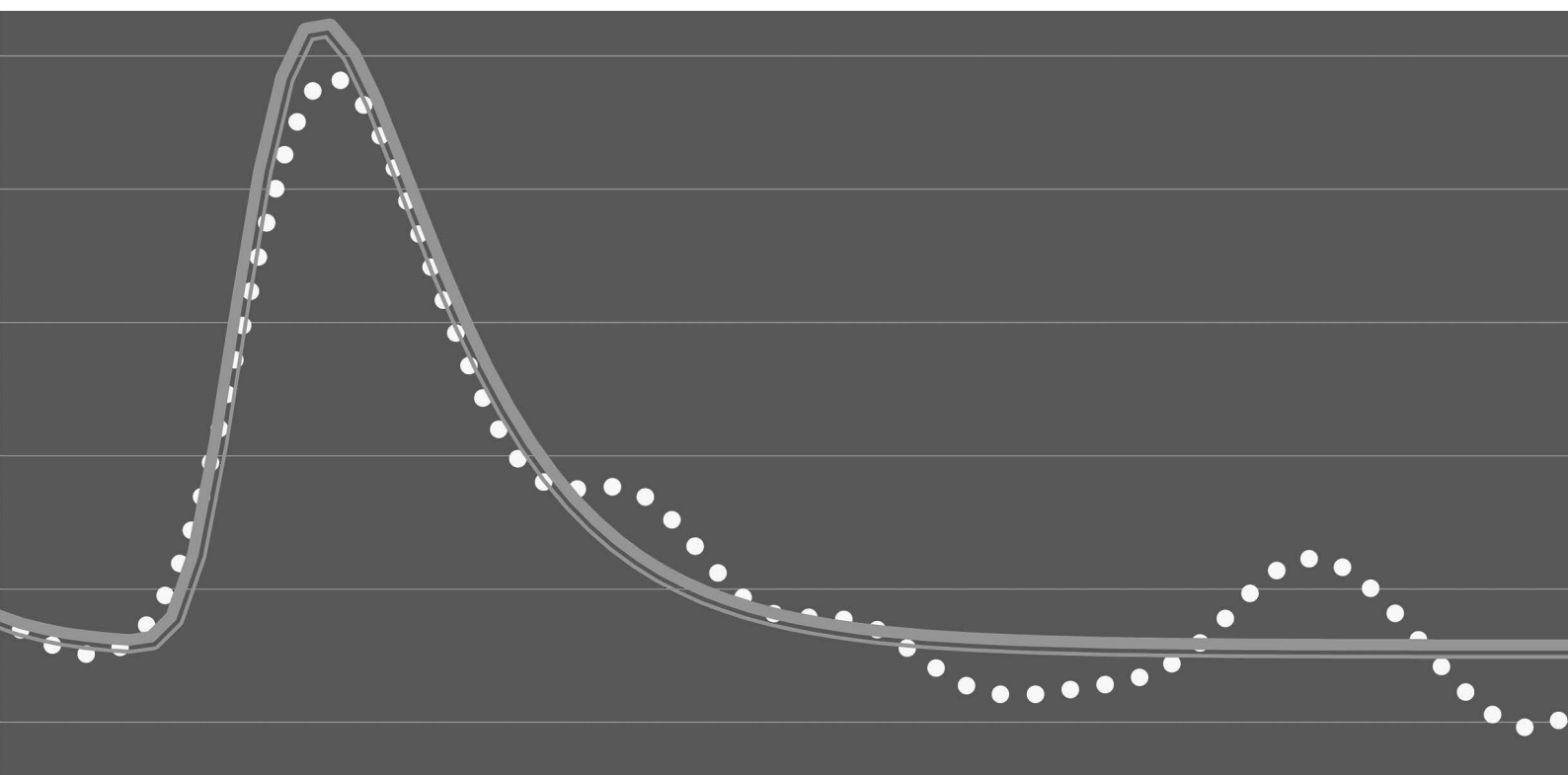
$$\alpha^{child} = \text{logit}(q_x) - \beta^*.Y^s(x).$$

On procède ensuite à une régression des estimations corrigées de α^{child} par rapport au temps, et on combine celles-ci avec celles de β aux dates auxquelles se rapportent les estimations de mortalité adulte, afin de calculer des valeurs corrigées de $Y(15)$, ${}_{45}q_{15}$ et $Y(60)$ et de lancer un deuxième round de calcul de valeurs corrigées de β . Nous avons ainsi mis en place un mécanisme itératif qui déterminera les meilleures régressions de α et β par rapport au temps, chacune tenant compte de l'autre.

Développement de la formule de calcul de α pour la mortalité adulte

$$\begin{aligned} {}_n p_x &= \frac{l(x+n)}{l(x)} = \left[\frac{1 + e^{2(\alpha + \beta Y^s(x))}}{1 + e^{2(\alpha + \beta Y^s(x+n))}} \right] \\ {}_n p_x + {}_n p_x \cdot e^{2(\alpha + \beta Y^s(x+n))} &= 1 + e^{2(\alpha + \beta Y^s(x))} \\ e^{2\alpha} \cdot \left({}_n p_x \cdot e^{2\beta Y^s(x+n)} - e^{2\beta Y^s(x)} \right) &= 1 - {}_n p_x \\ e^{2\alpha} &= \frac{1 - {}_n p_x}{{}_n p_x \cdot e^{2\beta Y^s(x+n)} - e^{2\beta Y^s(x)}} \\ \alpha &= \frac{1}{2} \ln \left(\frac{1 - {}_n p_x}{{}_n p_x \cdot e^{2\beta Y^s(x+n)} - e^{2\beta Y^s(x)}} \right) \\ \alpha &= \frac{1}{2} \left\{ \ln(1 - {}_n p_x) - \ln \left({}_n p_x \cdot e^{2\beta Y^s(x+n)} - e^{2\beta Y^s(x)} \right) \right\}. \end{aligned}$$

Migration



Chapitre 34 INTRODUCTION À L'ANALYSE DES MIGRATIONS

Rob Dorrington et Kenneth Hill

VUE D'ENSEMBLE

La migration est le troisième phénomène (avec la fécondité et la mortalité) qui détermine l'évolution d'une population. Pour la plupart des populations nationales, sa contribution à l'évolution démographique est modeste, comparée à celles des naissances et des décès, mais, en général, plus la subdivision administrative considérée est petite, plus elle prend de l'importance. La migration diffère de la fécondité et de la mortalité, non seulement par son ampleur, mais plus fondamentalement par sa nature. Elle suppose que l'on se déplace en franchissant une frontière géographique, dans le but, atteint ou en projet, de changer de lieu de résidence habituelle. Donc si une naissance et un décès se définissent généralement sans ambiguïté, il y a ou non migration selon les unités géographiques concernées (les subdivisions administratives) et selon l'intention de l'individu ou son comportement ultérieur. Un même individu peut être un migrant aux yeux du chercheur qui étudie l'évolution de la population d'une province, et ne pas être un migrant pour un autre observateur qui analyse le mouvement démographique au niveau national. Toute étude de la migration doit donc, avant toute chose, se fixer un niveau géographique d'analyse. Il faut ensuite définir, parmi les déplacements en tout genre, ceux qui seront comptés comme migrations. Le problème est encore rendu plus compliqué par l'existence de plusieurs types de migrations. En plus du changement « classique » de résidence habituelle, il y a les flux migratoires circulaires, les navettes quotidiennes ou hebdomadaires, les migrations saisonnières et les flux de réfugiés, qui ont tous leurs caractères particuliers. Étant donné ces difficultés de définition, et le fait que la migration (contrairement à la naissance et au décès) est un phénomène véritablement réversible en termes d'effectifs de population, il n'est pas étonnant que sa mesure soit également complexe.

De plus, recueillir des données sur la migration est une entreprise particulièrement problématique. Si, souvent, les pays en développement n'ont pas de système d'enregistrement complet des naissances et des décès, il y a des progrès à cet égard, et on a mis au point des méthodes pour exploiter des données lacunaires. Cependant, les données enregistrées sur les migrants et/ou les migrations, dans la plupart des pays, ne peuvent pas servir de base à des évaluations fiables de nombres d'immigrants internationaux, et encore moins de migrants ou migrations internes. En outre, pour différentes raisons (situation irrégulière, résidence temporaire des migrants récents, crainte de la xénophobie, etc.), les migrants (et particulièrement les immigrants internationaux) sont habituellement sous-représentés dans les recensements et les enquêtes.

Les méthodes de mesure de la migration sont largement les mêmes pour la migration interne que pour la migration internationale (qu'il s'agisse d'immigration ou d'émigration), sauf sur un point particulier très important. Un recensement ou une enquête peut mesurer l'immigration internationale en identifiant les personnes nées à l'étranger, mais il est beaucoup plus difficile de saisir les émigrants, car on ne peut pas procéder au recensement ou à l'enquête dans tous les pays d'accueil. Pour évaluer l'émigration, on dispose de plusieurs approches : (i) identifier systématiquement les nationaux dans les recensements d'autres pays (UN Population Division 2011) ; (ii) interroger chaque ménage sur les membres habituels du ménage qui vivent à l'étranger (exemple : les recensements du Swaziland de 1986 et 1996) ; (iii) demander à la personne interrogée si elle a des parents proches qui vivent à l'étranger, en particulier des frères ou sœurs, ou demander à une femme si elle a des enfants à l'étranger (Zaba 1985) ; et (iv) recourir à des méthodes de comparaison intercensitaire pour estimer les nombres de résidents manquants lors d'un recensement par rapport au

recensement précédent. La première approche dépend de la possibilité et de la volonté des pays d'accueil de fournir les données voulues, et elle ne permet de saisir que les migrations de citoyens natifs du pays de départ ; la deuxième repose sur le concept, plus ou moins vague, de membre d'un ménage, et elle ne permet pas d'observer l'émigration d'un ménage entier ; la troisième ne peut pas non plus constater l'absence d'une famille entière, elle ne fournit pas d'estimations de l'émigration récente et, lors de petites enquêtes expérimentales, elle ne s'est pas montrée convaincante. Seule la quatrième approche semble capable de fournir des estimations plausibles des flux récents, à condition que les deux recensements dénombrent la population avec une précision suffisante, mais elle ne donnera aucune information exploitable sur la destination des absents.

Étant donné ces limites et les difficultés qui grèvent la collecte des données, l'analyse de la migration s'est développée en grande partie indépendamment de l'analyse démographique classique, et elle s'est alors concentrée essentiellement sur les pays développés, parce que la qualité des données existantes sur la migration y est généralement bien meilleure que dans les pays en développement, et peut-être aussi parce que la migration y est souvent un sujet plus sensible sur le plan politique. Une autre conséquence de ces difficultés est le développement d'une terminologie et de techniques qui sont spécifiques à ce domaine, et souvent très éloignées de la démographie dont il est question dans le reste de ce manuel.

DÉFINITIONS

Comme on l'a dit plus haut, on définit la migration comme un déplacement qui franchit une frontière géographique (généralement administrative) à laquelle s'intéresse le chercheur, et a pour effet le changement du lieu de résidence habituelle de l'individu migrant. La frontière étant supposée clairement définie, ceci soulève immédiatement deux questions : comment définit-on le lieu de résidence habituelle, et comment détermine-t-on si ce lieu de résidence a changé ? Malheureusement, aucune réponse précise ne peut être apportée à ces deux questions, ce qui entraîne une inévitable incertitude en ce qui concerne la mesure du phénomène. La définition la plus souvent retenue de la résidence habituelle invoque la durée de résidence : si un individu a l'intention de vivre, ou a vécu, dans un lieu pendant un temps déterminé (par exemple un an), ce lieu est sa résidence habituelle. Notons que la résidence habi-

tuelle n'est pas la même chose que la résidence légale. Les *Principes et recommandations concernant les recensements de la population et des logements* (Division de Statistique des NU, 2009, p. 111, para. 1.463) définissent comme suit la résidence habituelle :

« Il est recommandé que les pays tiennent compte d'un délai de 12 mois pour déterminer le domicile habituel selon l'un des deux critères suivants :

a) Le lieu où la personne recensée a vécu continuellement la plupart du temps au cours des 12 derniers mois (autrement dit, au moins six mois et un jour), sans compter les absences temporaires pendant des vacances ou des missions de travail, ou bien où cette personne a l'intention de vivre pendant au moins six mois ;

b) Le lieu où la personne recensée a vécu continuellement au moins au cours des 12 derniers mois, sans compter les absences temporaires pendant les vacances ou des missions de travail, ou bien où cette personne a l'intention de vivre pendant au moins 12 mois. »

Mais cette définition ne règle pas le cas d'une personne qui a deux logements et passe régulièrement environ six mois dans chacun. En général, on doit accepter les déclarations des individus qui se définissent eux-mêmes comme résidents ou non, quoique certains tests soient possibles (comme demander à la personne où est immatriculée sa voiture, où elle paie ses impôts, où elle vote, où elle passe régulièrement la nuit, etc.). Dans la plupart des cas, une personne peut distinguer si elle est résident habituel ou visiteur, et cette simple distinction est suffisante.

SOURCES DE DONNÉES

La migration peut être considérée comme la « Cendrillon de la démographie », reléguée à l'arrière-plan autant que possible. Les enquêtes consacrées à la migration sont rares, très espacées et spécialisées (un excellent exemple est la description du Projet Migration au Mexique par Massey, Alarcon, Durand *et al.* 1987). Dans une enquête sur la migration, on trouve généralement des histoires migratoires complètes, qui, bien que soulevant des problèmes d'analyse complexes, ne sont pas axées sur l'estimation des nombres de migrants ou de migrations. Dans cette section, nous ne traiterons pas de l'analyse de ces histoires migratoires complètes (il y a très peu de principes généraux qui seraient applicables à beaucoup d'entre elles), nous nous intéresserons plutôt aux différentes sortes de données recueillies par les recensements et les enquêtes générales sur les ménages,

et parfois aussi, dans certains pays développés, par l'un ou l'autre type d'enregistrement.

Le lieu de naissance

L'information utile à l'étude de la migration la plus largement recueillie est le lieu de naissance. Quand on le compare avec le lieu de résidence à la date de l'enquête, on observe la migration sur la vie entière (parfois appelée en français « migration durée de vie »). Cette donnée ne fournit qu'une information limitée sur le calendrier de la migration, et il s'agit d'une migration « nette », en ce sens qu'elle ignore complètement les émigrations suivies d'un retour (au lieu de naissance), ainsi que toutes les migrations intermédiaires. Au moment de la collecte des données, il faut décider du niveau de détail des informations à recueillir, c'est-à-dire, pour les personnes nées à l'étranger, combien de pays différents seront distingués, et pour les personnes nées dans le pays, quel niveau de précision géographique on adoptera. Bien sûr, pour le chercheur qui analyse les données, ces décisions auront été prises lors de la conception du questionnaire, mais il peut avoir besoin d'un niveau d'agrégation plus élevé. L'analyse des données sur le lieu de naissance est présentée plus loin, mais il est utile de noter ici deux points. Premièrement, si on dispose de données sur le lieu de naissance, par sexe et âge, à deux dates, on peut estimer la migration nette (par sexe et âge) sur la période délimitée par ces deux dates. Deuxièmement, bien que le lieu de naissance ne révèle que les « migrations durée de vie », la longueur de cette « durée de vie » varie avec l'âge, et, pour peu que les données du recensement relatives aux enfants soient fiables (ce qui n'est souvent pas le cas dans bien des pays en développement), la migration des enfants de 0-4 ans peut servir d'indicateur de la migration récente de leurs parents (Raymer et Rogers 2007).

Le lieu de résidence à une date précise dans le passé

Cette information est très souvent recueillie en même temps que le lieu de naissance, dans le but explicite d'obtenir des données sur la migration récente. La date en question remonte généralement cinq ans en arrière, parfois un an seulement. Les résultats ont tendance à être meilleurs si cette date est associée à un événement marquant, tel que le recensement précédent, à condition que celui-ci ait été suffisamment complet (en sorte que les gens se souviennent d'avoir été recensés). Une période plus longue enregistre

plus de migrants, mais ignore les déplacements intermédiaires, alors qu'une période plus courte est davantage exposée aux erreurs d'évaluation des durées (« j'ai déménagé il y a à peu près un an »).

Le lieu de résidence antérieure

Cette information est presque toujours envisagée comme une alternative à celle du lieu de résidence à une date précise dans le passé, et on la combine généralement avec une question complémentaire sur la durée de la résidence actuelle (ou la date du dernier déménagement). Le but, là encore, est d'obtenir des données sur la migration récente.

La durée de la résidence actuelle

Cette question porte sur la durée de résidence dans la zone géo-administrative (comme la ville ou la province) et non dans le logement individuel. Prise isolément, elle n'a pas beaucoup d'intérêt, mais, associée avec la question précédente, elle fournit un cadre temporel pour les indices estimés.

Le mouvement intercensitaire de la population

Bien qu'elle n'implique aucune question directe sur la migration, l'évaluation de la variation intercensitaire de la population, par sexe et âge, peut, si les deux dénombrements sont suffisamment précis, fournir, par différence, des estimations de la migration intercensitaire nette (Hill 1987 ; Hill et Wong 2005 ; UN Population Division 1967). On déduit du mouvement intercensitaire par sexe et âge (par cohortes ou par groupes d'âge) la part due à la fécondité et à la mortalité pour obtenir une estimation de la migration intercensitaire nette (c'est-à-dire que l'on traite la migration comme le terme résiduel de l'équation fondamentale de la dynamique démographique). La migration est généralement concentrée dans la tranche d'âge de 20 à 40 ans, âges auxquels les taux de mortalité sont relativement faibles, du moins en l'absence d'épidémie de VIH/sida, et où la natalité n'entre pas en ligne de compte. Les estimations de la migration par cette méthode ne sont donc pas affectées par les hypothèses de mortalité et de fécondité (sauf dans les populations gravement touchées par le VIH/sida, où il est déconseillé d'employer ces données pour évaluer la migration). Mais elles sont extrêmement sensibles aux différences, même légères, de couverture des recensements ; les erreurs que cela entraîne peuvent se manifester sous la forme de taux élevés de migration par âge au-delà

de 50 ans, alors que la migration est habituellement faible dans cette catégorie d'âge.

LES MESURES DE LA MIGRATION

Il n'entre pas dans les objectifs de cette introduction de fournir un récapitulatif complet des définitions et mesures de la migration – nous renvoyons le lecteur intéressé au manuel des Nations Unies consacré aux migrations internes (UN Population Division 1970) – mais deux d'entre elles présentent un intérêt particulier pour les chapitres qui suivent.

La population immigrée (en anglais : migration stocks)

On appelle population immigrée le nombre de personnes (décomposé par sexe et groupe d'âge) qui ne sont pas nées dans la subdivision administrative recensée. La proportion de personnes nées ailleurs (dans le pays ou à l'étranger) donne une bonne idée générale de l'ampleur de l'immigration (interne et internationale), mais aucune indication sur l'évolution récente du phénomène. Néanmoins, la variation de la population immigrée peut servir à l'estimation de l'immigration (nette des migrations intermédiaires ou de retour des personnes nées à l'étranger).

Les taux de migration

En supposant que l'observation des migrations puisse être exhaustive et exacte, on peut calculer des taux d'émigration (interne ou internationale) exactement de la même manière que pour la mortalité, en divisant les nombres d'événements survenus au cours d'une période déterminée par la durée d'exposition au risque : des taux globaux (tous sexes et âges confondus) ou des taux par sexe et âge. Il n'en va pas de même (du moins utilement) pour l'immigration, car la population exposée au risque d'immigrer dans une subdivision administrative est l'ensemble de la population mondiale vivant hors de cette région. Les taux d'immigration se calculent toujours en divisant les nombres d'événements par la durée d'exposition de la seule population qui n'est pas exposée à ce risque, les résidents actuels de la région ; ce peut être des taux globaux (tous sexes et âges confondus) ou des taux par sexe et âge. Définir ces taux de cette manière présente l'avantage de satisfaire l'équation fondamentale de la dynamique démographique, puisque les taux d'accroissement et de diminution se rapportent à la même population. Cela entraîne un autre avantage : les taux de migration nette peuvent être déduits de l'équation fondamentale

comme la variation de la population entre deux dates (par exemple deux recensements) moins les naissances et plus les décès survenus dans le même intervalle de temps. Mais cette approche présente aussi l'inconvénient de repousser les limites des taux « normaux » : par exemple, dans un cas extrême, l'immigration d'un individu dans une région précédemment inhabitée donne lieu à un taux d'immigration infini.

DESCRIPTION DES MÉTHODES PRÉSENTÉES

Les chapitres de cette section sont axés sur l'évaluation et la description quantitative de l'immigration internationale et des migrations internes. Il ne s'agit pas de présenter un panorama exhaustif de toutes les mesures de la migration ; en particulier, nous ne parlerons pas du problème, important mais délicat, de la mesure de l'émigration internationale, sauf pour mentionner que la méthode d'estimation de l'immigration internationale des étrangers (nette des migrations intermédiaires et des retours) peut être appliquée aux données des principaux pays de destination des émigrants pour avoir une certaine idée de l'ampleur et de la structure par âge de l'émigration.

Le chapitre 35 porte sur les méthodes de base d'analyse des données de recensement pour estimer les nombres d'immigrants internationaux (nets des migrations intermédiaires et des retours) à partir de la variation de l'effectif des étrangers, et les nombres de migrants internes à partir de la variation d'effectif de la population en fonction du lieu de naissance et du lieu de résidence à une certaine date antérieure au recensement.

Le chapitre 36 décrit le choix et l'ajustement d'un modèle multi-exponentiel de Rogers-Castro aux probabilités (ou taux) de migration basées sur des estimations de nombres de migrants/migrations, en utilisant des procédures d'optimisation non-linéaires.

Le chapitre 37 présente les modèles multiplicatifs et log-linéaires capables de saisir, comparer et analyser la masse des flux migratoires inter-régionaux par lieu de départ et lieu d'arrivée. On y trouvera également une introduction à la « méthode des offsets » pour étendre l'utilisation de ces modèles à l'estimation des flux inter-régionaux à partir des flux marginaux (c'est-à-dire les flux totaux sortant des régions ou y entrant). Notre intention est de développer la « méthode des offsets » dans un chapitre complémentaire qui sera publié plus tard sur le site web associé à cet ouvrage.

AUTRES LECTURES ET RÉFÉRENCES

Comme nous l'avons déjà signalé, le *Manuel VI* des Nations unies (UN Population Division 1970) fournit une introduction complète, quoiqu'ancienne, à la description et à la mesure de la migration interne. Nous renvoyons le chercheur intéressé par une vue d'ensemble des méthodes indirectes d'estimation de la migration à l'inventaire, toujours utile même s'il date aussi un peu, de Zaba (1987). Sur un plan plus spécifique, Hill (1987) a essayé d'appliquer la logique qui sous-tend la méthode d'estimation de la mortalité adulte dite « méthode généralisée de la balance de l'accroissement démographique » (décrite dans le chapitre 24) à l'évaluation de la migration clandestine, et Hill et Queiroz (2010) ont cherché à estimer la migration nette parallèlement à la mortalité, dans le but d'améliorer les estimations de mortalité. Malheureusement, aucune de ces méthodes ne s'est révélée particulièrement efficace.

Les lecteurs qui souhaitent se documenter davantage sur les modèles de migration (multi-exponentiel, multiplicatif et log-linéaire) ou sur la « méthode des offsets » peuvent parcourir les travaux de Rogers, Willekens et leurs collègues (entre autres, Little et Rogers 2007 ; Raymer et Rogers 2007 ; Rogers (1980, 1986) ; Rogers, Little et Raymer 2010 ; et Willekens 1999).

- Hill K. 1987. "New approaches to the estimation of migration flows from census and administrative data sources", *International Migration Review* 21(4):1279-1303. <http://dx.doi.org/10.2307/2546515>
- Hill K and B Queiroz. 2010. "Adjusting the general growth balance method for migration", *Revista Brasileira de Estudos de População* 27(1):7-20. doi: <http://dx.doi.org/10.1590/S0102-30982010000100002>
- Hill K and R Wong. 2005. "Mexico-US migration: Views from both sides of the border", *Population and Development Review* 31(1):1-18. doi: <http://dx.doi.org/10.1111/j.1728-4457.2005.00050.x>
- Little JS and A Rogers. 2007. "What can the age composition of a population tell us about the age composition of its out-migrants?", *Population, Space and Place* 13(1):23-19. doi: <http://dx.doi.org/10.1002/psp.440>
- Massey DS, R Alarcon, J Durand and H Gonzalez. 1987. *Return to Aztlan: The Social Process of International Migration from Western Mexico*. Berkeley and Los Angeles: University of California Press.
- Raymer J and A Rogers. 2007. "Using age and spacial flow structures in the indirect estimation of migration streams", *Demography* 44(2):199-223. doi: <http://dx.doi.org/10.1353/dem.2007.0016>
- Rogers A. 1980. "Introduction to multistate mathematical demography", *Environment and Planning A* 12:489-498. doi: <http://dx.doi.org/10.1068/a120489>
- Rogers A. 1986. "Parameterized multistate population dynamics and projections", *Journal of the American Statistical Association* 81(393):48-61. doi: <http://dx.doi.org/10.1080/01621459.1986.10478237>
- Rogers A, JS Little and J Raymer. 2010. *The Indirect Estimation of Migration: Methods for Dealing with Irregular, Inadequate, and Missing Data*. Dordrecht: Springer.
- UN Population Division. 1967. *Manual IV: Methods for Estimating Basic Demographic Measures from Incomplete Data*. New York: United Nations, Department of Economic and Social Affairs, ST/SOA/Series A/42. <http://www.un.org/esa/population/techcoop/DemEst/manual4/manual4.html>
- UN Population Division. 1970. *Manual VI: Methods of Measuring Internal Migration*. New York: United Nations, Department of Economic and Social Affairs, ST/SOA/Series A/47. <http://www.un.org/esa/population/techcoop/IntMig/manual6/manual6.html>
- UN Population Division. 2011. *International Migration Report 2009: A Global Assessment*. New York: United Nations, Department of Economic and Social Affairs, ST/ESA/Series A/316. <http://www.un.org/esa/population/publications/migration/WorldMigrationReport2009.pdf>
- Division de statistique des NU. 2009. *Principes et recommandations concernant les recensements de la population et des logements, Deuxième révision*. New York : Nations Unies, Département des affaires économiques et sociales, ST/ESA/STAT/SER.M/67/Rev.2. http://unstats.un.org/unsd/publication/seriesM/seriesm_67Rev2f.pdf
- Willekens FJ. 1999. «Modeling approaches to the indirect estimation of migration flows: From entropy to EM», *Mathematical Population Studies* 7:239-278. doi: <http://dx.doi.org/10.1080/08898489909525459>
- Zaba B. 1985. *Measurement of Emigration Using Indirect Techniques: Manual for the Collection and Analysis of Data on Residence of Relatives*. Liège: Belgium: Ordina Editions.
- Zaba B. 1987. "The indirect estimation of migration: A critical review", *International Migration Review* 21(4):1395-1445. doi: <http://dx.doi.org/10.2307/2546519>

Chapitre 35 ESTIMATION DE LA MIGRATION À PARTIR DE DONNÉES DE RECENSEMENT

Rob Dorrington

DESCRIPTION DES MÉTHODES

Mesurer la migration à partir de données de recensement n'est pas techniquement compliqué. Pourvu que le(s) recensement(s) recueille(nt) les informations nécessaires et soi(en)t suffisamment fiables, on peut parvenir à des estimations de l'immigration internationale nette (c'est-à-dire les immigrations moins les émigrations) de la population née à l'étranger (hors d'un pays déterminé) et des migrations internes (entrées et sorties) entre les régions d'un pays au cours d'une période intercensitaire.

Pour mesurer l'immigration internationale nette de personnes nées à l'étranger, il suffit de retrancher du nombre de personnes recensées nées à l'étranger le nombre des étrangers qui étaient déjà présents lors du recensement précédent et sont supposés avoir survécu depuis lors.

De même, si les recensements enregistrent la région de naissance des individus, on peut mesurer l'immigration interne nette (c'est-à-dire l'immigration interne nette des personnes nées hors de la région considérée, moins l'émigration interne nette des personnes natives de la région) entre les régions du pays. Mais, si on demande aux personnes recensées où elles vivaient à une certaine date antérieure, par exemple au moment du recensement précédent, on peut évaluer directement le nombre de migrants survivants (c'est-à-dire les migrants encore vivants au moment du recensement le plus récent) en provenance et à destination de chaque région du pays depuis cette date antérieure.

Pour mesurer le nombre de migrants à partir du nombre de migrants survivants au moment du second recensement, on doit ajouter à ce dernier chiffre une estimation du nombre de migrants que l'on suppose décédés entre leur migration et le second recensement.

Si le dernier recensement a enregistré d'autres données, comme l'année d'arrivée du migrant à l'endroit où il a été recensé, on peut également évaluer la tendance de la migration au cours du temps.

La migration se différencie de la fécondité et de la mortalité, à la fois parce qu'elle n'est pas un aboutissement au sens où le sont une naissance ou un décès, et par le fait que nous devons nous intéresser non seulement à la population d'origine, que le migrant a quittée (qui correspond à la population soumise au risque, sur la base de laquelle peuvent être calculés des taux analogues aux taux de fécondité et de mortalité), mais aussi à la population de destination, celle que le migrant a rejointe. De plus, pour analyser la migration, on a souvent besoin de distinguer divers types de déplacements (temporaires ou permanents, circulaires ou non, etc.). C'est pourquoi l'éventail des définitions et des outils de mesure associés à la migration est beaucoup plus large que dans le cas de la fécondité ou de la mortalité. Il n'entre pas dans les objectifs de ce chapitre de traiter cette problématique, nous renvoyons donc le lecteur intéressé aux textes classiques sur le sujet, tels que le *Manuel VI* des Nations Unies (UN Population Division (1970), Shryock et Siegel (1976), Siegel et Swanson (2004)).

DONNÉES REQUISES ET HYPOTHÈSES

Tabulations nécessaires

- Pour mesurer l'immigration nette de personnes nées à l'étranger :
- Les nombres d'hommes et de femmes nés à l'étranger, par groupes quinquennaux d'âge et pour un groupe d'âge ouvert $A+$, à deux moments du temps, habituellement deux recensements.

- Pour les décès : soit une table-type de mortalité adéquate, soit les nombres d'hommes et de femmes nés dans le pays étudié, par groupes quinquennaux d'âge et pour un groupe d'âge ouvert $A+$, à deux moments du temps, habituellement deux recensements. À défaut, le taux brut de mortalité de cette population.
- Pour mesurer l'immigration interne nette (entre régions d'un même pays) à partir des données sur le lieu de naissance :
 - Les nombres d'hommes et de femmes par région de résidence et région de naissance, par groupes quinquennaux d'âge et pour un groupe d'âge ouvert $A+$, à deux moments du temps, habituellement deux recensements.
 - Pour les décès : soit une table-type de mortalité adéquate, soit les nombres d'hommes et de femmes nés dans la région considérée, par groupes quinquennaux d'âge et pour un groupe d'âge ouvert $A+$, à deux moments du temps, habituellement deux recensements, soit les nombres de décès par région d'après les registres d'état civil. À défaut, le taux brut de mortalité de cette population.
- Pour mesurer les migrations internes entre régions à partir des données sur le lieu de résidence au recensement précédent :
 - Les nombres d'hommes et de femmes par région de résidence actuelle et région de résidence à une date antérieure, habituellement lors du recensement précédent, par groupes quinquennaux d'âge et pour un groupe d'âge ouvert $A+$.
- Si on ne dispose pas d'effectifs par groupes d'âge, les données agrégées restent utiles pour estimer la migration totale, tous âges confondus.
- Les recensements dénombrent avec exactitude la population des régions et enregistrent avec précision la région de naissance des individus.
- On peut évaluer correctement la mortalité des personnes qui migrent d'une région à une autre (soit que l'on dispose d'une table de mortalité adéquate, soit que cette population ait la même mortalité que celle de la population autochtone du pays telle qu'estimée à partir des recensements).
- Pour la mesure de la migration interne entre régions à partir de données sur le lieu de résidence au recensement précédent :
- Le dernier recensement identifie bien toutes les personnes qui ont migré d'une région à une autre depuis la date de référence antérieure (par exemple depuis le recensement précédent).
- On peut évaluer correctement la mortalité des personnes qui migrent d'une région à une autre (soit que l'on dispose d'une table de mortalité adéquate, soit que cette population ait la même mortalité que celle de la population autochtone du pays telle qu'estimée à partir des recensements). Puisque l'on mesure l'immigration et l'émigration internes séparément (et pas seulement la migration interne nette), cette hypothèse est de moindre importance.

TRAVAUX PRÉPARATOIRES ET RECHERCHES PRÉLIMINAIRES

Avant toute utilisation de la méthode, on doit s'assurer de la qualité des données, au moins sur deux aspects :

- la structure par âge de la population (le cas échéant par région) ; et
- le taux de couverture des recensements (le cas échéant au niveau régional).

Hypothèses importantes

Pour la mesure de l'immigration nette d'étrangers :

- Les recensements dénombrent avec exactitude toutes les personnes nées à l'étranger.
- On peut évaluer avec précision la mortalité de la population née à l'étranger (soit que l'on dispose d'une table de mortalité adéquate, soit que cette population ait la même mortalité que celle de la population autochtone du pays telle qu'estimée à partir des recensements).
- Il n'y a pas de migration de retour des émigrés natifs du pays recensé.
- Pour la mesure de l'immigration interne nette à partir de données sur le lieu de naissance :

MISE EN GARDE

Mesurer la migration à partir des données de deux recensements sur le lieu de naissance nécessite, dans les deux opérations, non seulement que le dénombrement de la population soit suffisamment complet, mais aussi que le lieu de naissance soit enregistré avec exactitude. Or ce n'est pas toujours le cas, en particulier quand on veut mesurer l'immigration internationale dans un contexte où les immigrants préfèrent cacher leur statut d'étranger, et également, pour la mesure des migrations internes, en cas d'éventuelles modifications de frontières ou quand la personne qui répond ignore le lieu de naissance de la personne dont il s'agit.

Mesurer les migrations en interrogeant les migrants dépend de la capacité du recensement à identifier correctement toutes les personnes qui ont immigré, ainsi que le lieu exact d'où elles proviennent. Dans la mesure où, au moment du recensement, les immigrés récents ne sont pas encore définitivement installés comme résidents dans la région d'arrivée, ils peuvent échapper au dénombrement.

La migration nette, par définition, sous-estime les flux de migrants à l'entrée comme à la sortie d'une région ou d'un pays. Ainsi, par exemple, des personnes qui, au cours de la période considérée, ont immigré dans une région et en sont reparties ne sont pas comptabilisées dans l'immigration nette, alors qu'elles ont effectué deux migrations.

APPLICATION DE LA MÉTHODE

A) Mesure de l'immigration internationale nette de personnes nées à l'étranger à partir des données sur le lieu de naissance

Cette méthode permet de mesurer l'immigration internationale nette d'étrangers à partir des données sur le lieu de naissance. Il est important de souligner qu'elle ignore et ne mesure donc pas l'immigration de personnes autochtones qui ont quitté le pays avant le premier recensement et y sont revenues avant le second. Elle est donc déconseillée en cas de nombreux retours de personnes autochtones (par exemple après un temps d'exil ou une migration forcée de réfugiés).

Étape 1 : Choisir les coefficients de survie

Si les effectifs de personnes nées à l'étranger par groupes d'âge sont disponibles pour chaque recensement, on doit évaluer les coefficients de survie à appliquer à ceux du premier recensement pour estimer l'effectif des survivants au moment du second recensement. On peut choisir soit des coefficients basés sur les nombres d'années vécues dans chaque groupe quinquennal d'âge (${}_5L_x$) d'après le modèle général des tables-types de mortalité des Nations Unies ou l'une des quatre familles de tables-types de Princeton, soit des coefficients provenant d'une table-type de mortalité propre à une population touchée par une épidémie de sida (Timæus 2004), que l'on trouvera dans le tableur « Models » du document Excel associé disponible en ligne. Ce tableur permet également d'utiliser les nombres d'années vécues dans les groupes quinquennaux d'âge d'une autre table de mortalité si on a des raisons de penser que celle-ci correspond à un schéma de mortalité similaire à celui de la population étudiée ; à défaut, on peut déterminer les coefficients de survie à partir

des proportions de survivants de chaque groupe quinquennal d'âge entre un recensement et le suivant (supposés espacés de n années, n étant multiple de 5) parmi la population autochtone. Ainsi, ${}_5S_{x,n}$, ${}_{\infty}S_{A-n,n}$ et $S_{B,n}$ respectivement le coefficient de survie sur n années des personnes âgées de x à $x + 5$ lors du premier recensement, celui des personnes âgées de $A - n$ et plus lors du premier recensement, et celui des enfants nés entre les deux recensements, se calculent comme suit :

$$\begin{aligned} {}_5S_{x,n} &= \frac{{}_5L_{x+n}}{{}_5L_x} \text{ ou } \frac{{}_5N_x^{nb}(t+n)}{{}_5N_x^{nb}(t)}, \\ {}_{\infty}S_{A-n,n} &= \frac{T_A}{T_{A-n}} \text{ ou } \frac{{}_{\infty}N_A^{nb}(t+n)}{{}_{\infty}N_{A-n}^{nb}(t)}, \\ \text{et } S_{B,n} &= \frac{nL_0}{nL_0} \text{ ou } \frac{nN_0^{nb}(t+n)}{B^{nb}}, \end{aligned}$$

l'exposant nb signifiant « autochtone » (*en anglais : native-born*), ${}_5N_x^{nb}(t)$ représentant la population autochtone recensée au temps t , et B^{nb} représentant le nombre de naissances autochtones entre t et $t + n$.

Si les données ne sont pas disponibles par groupes quinquennaux d'âge, on peut toutefois évaluer le nombre total net d'immigrants, si on connaît le taux brut de mortalité de la population (qui, sauf indices contraires, peut être supposé identique à celui de la population autochtone).

Étape 2 : Évaluer le nombre de décès d'immigrés

Si on dispose des effectifs de personnes nées à l'étranger par groupes d'âge dans deux recensements (espacés de n années), on doit évaluer les nombres de décès de ces personnes (signalées par l'exposant F) âgées de x à $x + 5$, ${}_5D_x^F$, et de $A - n$ ou plus, ${}_{\infty}D_{A-n}^F$, au temps t (premier recensement), ainsi que des enfants nés entre les deux recensements, D_B^F :

$$\begin{aligned} {}_5D_x^F &= \frac{1}{2} \left({}_5N_x^F(t) \cdot {}_5S_{x,n} + {}_5N_{x+n}^F(t+n) \right) \left(\frac{1}{{}_5S_{x,n}} - 1 \right), \\ {}_{\infty}D_{A-n}^F &= \frac{1}{2} \left({}_{\infty}N_{A-n}^F(t) \cdot {}_{\infty}S_{A-n,n} + {}_{\infty}N_A^F(t+n) \right) \left(\frac{1}{{}_{\infty}S_{A-n,n}} - 1 \right), \\ \text{et } D_B^F &= \frac{1}{2} \left(nN_0^F(t+n) \right) \left(\frac{1}{S_{B,n}} - 1 \right), \end{aligned}$$

et ${}_5N_x^F(t)$ représentant l'effectif des personnes nées à l'étranger âgées de x à $x + 5$ au temps t (premier recensement).

Si les effectifs et/ou les coefficients de survie ne sont pas disponibles par groupes d'âge, on peut évaluer comme suit le nombre total des décès de personnes nées à l'étranger :

$${}_{\infty}D_0^F = \frac{n}{2} \left({}_{\infty}N_0^F(t) + {}_{\infty}N_0^F(t+n) \right) {}_{\infty}m_0$$

${}_{\infty}m_0$ étant une estimation du taux brut de mortalité de la population du pays recensé. Mais, si la structure par âge de la population née à l'étranger est nettement différente de celle de la population totale du pays, on risque d'obtenir une mauvaise approximation du véritable nombre de décès.

Étape 3 : Évaluer le nombre net d'immigrants étrangers

Si on dispose de données par groupe d'âge pour chaque recensement, on peut évaluer l'immigration nette par âge :

$$\text{Net } {}_5M_x^F = {}_5N_{x+n}^F(t+n) - {}_{\infty}N_x^F(t) + {}_5D_x^F$$

pour $x = 0, 5, \dots, A-5-n$

$\text{Net } {}_5M_x^F$ représentant le nombre net d'immigrants entre t et $t+n$ qui étaient âgés de x à $x+5$ au temps t . Pour $x > A-5-n$:

$$\text{Net } {}_{\infty}M_{A-n}^F = {}_{\infty}N_A^F(t+n) - {}_{\infty}N_{A-n}^F(t) + {}_{\infty}D_{A-n}^F.$$

Le nombre net d'immigrants parmi les enfants nés entre t et $t+n$ est :

$$\text{Net } M_B^F = {}_nN_0^F(t+n) + D_B^F.$$

Si les effectifs et/ou les coefficients de survie ne sont pas disponibles par groupes d'âge, on évalue comme suit le nombre total net d'immigrants :

$$\text{Net } {}_{\infty}M_0^F = {}_{\infty}N_0^F(t+n) - {}_{\infty}N_0^F(t) + {}_{\infty}D_0^F.$$

B) Mesure de la migration inter-régionale nette à partir de données sur le lieu de naissance

On peut mesurer l'immigration interne nette vers une région déterminée au départ des autres régions du même pays exactement de la même manière que l'immigration internationale, décrite plus haut, en remplaçant la population née à l'étranger par la population née hors de la région considérée.

En outre, en appliquant la même méthode aux données sur l'évolution des nombres de personnes nées dans la région étudiée (plutôt qu'en dehors) et vivant à l'extérieur, on peut évaluer l'émigration interne nette des personnes nées dans la région vers les autres régions du pays. En soustrayant cet effectif de l'immigration interne nette de personnes nées en

dehors de la région étudiée, on obtient une évaluation de l'immigration interne nette totale dans cette région.

Si on a des raisons de penser que les immigrants nés à l'extérieur de la région considérée et les émigrants natifs de cette région ont des schémas de mortalité différents, et si on dispose des coefficients de survie nécessaires, on peut appliquer des coefficients de survie différents à ces deux sous-populations pour évaluer le nombre net de migrants. Cependant, en pratique, il est probable que l'imprécision des données censitaires sur le lieu de naissance lors du recensement précédent l'emportera sur le gain de précision dû à l'utilisation de schémas de mortalité différents.

C) Mesure de la migration interne entre régions d'un même pays à partir des données sur le lieu de résidence lors de l'enquête précédente

On mesure la migration inter-régionale nette directement à partir des nombres de résidents recensés de chaque région qui ont migré depuis le recensement précédent, en fonction de leur lieu (région) de résidence à une date antérieure déterminée (par exemple, lors du précédent recensement). En limitant le calcul aux flux inter-régionaux, la somme des nombres d'immigrants inter-régionaux devrait être égale à la somme des nombres d'émigrants inter-régionaux ; si les immigrants en provenance de l'étranger sont compris dans ces données, on peut élargir l'évaluation de l'immigration interne pour y inclure l'immigration internationale dans chaque région.

Puisque l'un des principaux centres d'intérêt est l'ampleur des flux migratoires inter-régionaux, on accorde autant d'attention aux effectifs totaux de migrants entre régions qu'à la structure par âge des différents flux.

On détermine le nombre des migrants à partir des nombres d'immigrants et d'émigrants internes survivants de la manière suivante :

$${}_5M_x = ({}_5I'_x - {}_5O'_x + ({}_5I'_x - {}_5O'_x)_x / {}_5S_x) / 2,$$

l'apostrophe (') indiquant les nombres de survivants et ${}_5I'_x$ et ${}_5O'_x$ représentant respectivement les nombres d'immigrants internes survivants et d'émigrants internes survivants d'une région donnée à la date du second recensement, qui étaient âgés de x à $x+5$ à cette date.

EXEMPLE

Dans cet exemple, nous exploiterons les effectifs masculins du recensement sud-africain de 2001 et de l'« enquête

tenant lieu de recensement » dite *Community Survey* de 2007. (Bien que l'enquête ait été effectuée environ 5,35 ans après la nuit du recensement de 2001, nous supposons, pour les besoins de notre exposé, qu'elle a eu lieu exactement 5 ans après le recensement.) On trouvera ces exemples dans le fichier Migration_South Africa_males.xlsx (voir le site web).

A) Mesure de l’immigration internationale nette d’étrangers à partir des données sur le lieu de naissance

Étape 1 : Choisir les coefficients de survie

Les coefficients de survie apparaissent dans la cinquième colonne du tableau 35.1. Leurs valeurs ont été calculées à partir (des années vécues dans chaque groupe d’âge) de la table de mortalité alternative du tableur *Models*, au moyen des formules suivantes, respectivement pour les personnes âgées de 20 à 24 ans en années révolues et de 80 ans et plus

au premier recensement, et pour les enfants nés entre les deux recensements :

$${}_5S_{20,5} = \frac{{}_5L_{25}}{{}_5L_{20}} = \frac{4,3382}{4,4975} = 0,96458$$

$${}_{\infty}S_{80,5} = \frac{T_{85}}{T_{80}} = \frac{0,75180}{1,19603} = 0,40912$$

et $S_{B,5} = \frac{{}_5L_0}{{}_5l_0} = \frac{4,707549}{5} = 0,94151$.

Étape 2 : Évaluer le nombre de décès

Comme chaque recensement nous donne les nombres de personnes nées à l’étranger par groupe d’âge, nous pouvons évaluer les nombres de décès de personnes nées à l’étranger survenus entre les deux recensements par groupe d’âge, en utilisant les effectifs d’étrangers de chaque recensement, qui figurent dans les deuxième et troisième colonnes du

TABLEAU 35.1 Estimation du nombre de décès de personnes nées à l’étranger et du nombre net d’immigrants internationaux par groupe d’âge, Afrique du Sud, 2001-2006

| Âge | 2001 | 2006 | x | ${}_5S_x$ | Âge au 2 ^e recensement | D_F | M net |
|-------|---------|---------|-----|-----------|--------------------------------------|--------|---------|
| | | | B | 0,94151 | | | |
| 0-4 | 8 963 | 12 577 | 0 | 0,97896 | 0-4 | 391 | 12 968 |
| 5-9 | 10 390 | 13 724 | 5 | 0,99547 | 5-9 | 242 | 5 003 |
| 10-14 | 13 508 | 13 998 | 10 | 0,99427 | 10-14 | 55 | 3 664 |
| 15-19 | 27 835 | 27 943 | 15 | 0,98602 | 15-19 | 119 | 14 555 |
| 20-24 | 69 787 | 59 493 | 20 | 0,96458 | 20-24 | 616 | 32 275 |
| 25-29 | 87 381 | 95 763 | 25 | 0,93161 | 25-29 | 2 994 | 28 970 |
| 30-34 | 73 338 | 100 450 | 30 | 0,90960 | 30-34 | 6 675 | 19 743 |
| 35-39 | 66 663 | 85 490 | 35 | 0,89780 | 35-39 | 7 563 | 19 715 |
| 40-44 | 59 152 | 75 684 | 40 | 0,89092 | 40-44 | 7 701 | 16 721 |
| 45-49 | 45 184 | 66 113 | 45 | 0,88633 | 45-49 | 7 274 | 14 234 |
| 50-54 | 40 398 | 55 913 | 50 | 0,87224 | 50-54 | 6 154 | 16 883 |
| 55-59 | 30 640 | 42 833 | 55 | 0,84731 | 55-59 | 5 717 | 8 153 |
| 60-64 | 24 376 | 34 433 | 60 | 0,80885 | 60-64 | 5 442 | 9 234 |
| 65-69 | 17 895 | 25 588 | 65 | 0,75468 | 65-69 | 5 353 | 6 564 |
| 70-74 | 13 561 | 18 989 | 70 | 0,66991 | 70-74 | 5 281 | 6 375 |
| 75-79 | 10 238 | 12 850 | 75 | 0,56388 | 75-79 | 5 404 | 4 693 |
| 80-84 | 7 658 | 7 461 | 80+ | 0,40912 | 80-84 | 5 118 | 2 341 |
| 85+ | 4 455 | 5 305 | | | 85+ | 7 410 | 602 |
| TOTAL | 611 423 | 754 608 | | | TOTAL | 79 509 | 222 693 |

tableau 35.1. Pour les personnes de 20-24 ans et de 80 ans et plus à la date du premier recensement, et pour les enfants nés entre les deux recensements, les calculs sont les suivants :

$$\begin{aligned} {}_5D_{20}^F &= \frac{1}{2}({}_5N_{20}^F(2001) \cdot {}_5S_{20,5} + {}_5N_{25}^F(2006))\left(\frac{1}{{}_5S_{20,5}} - 1\right) \\ &= (69\,787 \cdot 0,96458 + 95\,763)\left(\frac{1}{0,96458} - 1\right) = 2\,994 \\ {}_\infty D_{80}^F &= \frac{1}{2}({}_\infty N_{80}^F(2001) \cdot {}_\infty S_{80,5} + {}_\infty N_{85}^F(2006))\left(\frac{1}{{}_\infty S_{80,5}} - 1\right) \\ &= ((7\,658 + 4\,455) \cdot 0,40\,912 + 5\,305)\left(\frac{1}{0,40\,912} - 1\right) = 7\,410 \\ \text{et } D_B^F &= \frac{1}{2}({}_5N_0^F(2006))\left(\frac{1}{S_{B,5}} - 1\right) = 12\,577\left(\frac{1}{0,94151} - 1\right) \\ &= 391. \end{aligned}$$

Si les effectifs et/ou les coefficients de survie n'étaient pas disponibles par groupes d'âge, on pourrait évaluer comme suit le nombre total des décès de personnes nées à l'étranger, sachant que le taux brut de mortalité de la population vaut 14 pour 1 000 :

$$\begin{aligned} {}_\infty D_0^F &= \frac{5}{2}({}_\infty N_0^F(2001) + {}_\infty N_0^F(2006)) {}_\infty m_0 \\ &= \frac{5}{2}(611\,423 + 754\,608) \frac{14}{1000} = 47\,811. \end{aligned}$$

Étape 3 : Évaluer le nombre net d'immigrants étrangers

Puisque nous disposons de données par groupe d'âge pour chaque recensement, l'immigration nette par âge des personnes nées à l'étranger peut être calculée de la manière suivante :

$$\begin{aligned} \text{Net } {}_5M_{20}^F &= {}_5N_{25}^F(2006) - {}_\infty N_{20}^F(2001) + {}_5D_{20}^F \\ &= 95\,763 - 69\,787 + 2\,994 = 28\,970 \\ \text{Net } {}_\infty M_{80}^F &= {}_\infty N_{85}^F(2006) - {}_\infty N_{80}^F(2001) + {}_\infty D_{80}^F \\ &= 5\,305 - (7\,658 + 4\,455) + 7\,410 = 602 \\ \text{Net } M_B^F &= {}_5N_0^F(2006) + D_B^F = 12\,577 + 391 = 12\,968. \end{aligned}$$

Si les effectifs et/ou les coefficients de survie n'étaient pas disponibles par groupe d'âge, on pourrait évaluer comme suit le nombre total net d'immigrants :

$$\begin{aligned} \text{Net } {}_\infty M_0^F &= {}_\infty N_0^F(2006) - {}_\infty N_0^F(2001) + {}_\infty D_0^F \\ &= 754\,608 - 611\,423 + 47\,811 = 190\,996. \end{aligned}$$

B) Mesure de l'immigration inter-régionale nette à partir des données sur le lieu de naissance

Les deuxième et troisième colonnes du tableau 35.2 présentent les nombres de personnes vivant dans la province de Western Cape, en Afrique du Sud, et nées à l'extérieur de la province, dénombrées respectivement par le recensement de 2001 et la *Community Survey* de 2007. Bien que l'on ait utilisé les mêmes coefficients de survie (colonne 5) que dans l'exemple A, il conviendrait de procéder autrement si on devait supposer que les autochtones et les immigrants avaient des schémas de mortalité très différents. La dernière colonne du tableau 35.2 présente, par groupe d'âge, les nombres nets d'immigrants dans la province de Western Cape nés dans une autre province. Au total, 213 911 personnes nées hors de la province y ont immigré (déduction faite de celles d'entre elles qui l'ont quittée).

Les deuxième et troisième colonnes du tableau 35.3 présentent les nombres de personnes nées dans la province de Western Cape et résidant dans une autre province, selon le recensement de 2001 et l'enquête de 2007. Le nombre net d'émigrants internes natifs du Western Cape (c'est-à-dire le nombre de personnes qui sont nées dans cette province et l'ont quittée, moins celles d'entre elles qui y sont revenues) figure dans la huitième colonne. Un nombre négatif signifie que le nombre de natifs du Western Cape qui sont partis vivre dans une autre province pendant la période considérée est inférieur au nombre de natifs de la province qui, vivant ailleurs, y sont revenus. Le total de - 19 017 indique donc que le nombre de personnes qui sont nées au Western Cape, ont vécu ailleurs jusqu'en 2001 et y sont revenues pendant la période étudiée excède de 19 017 celui des personnes qui sont nées au Western Cape et l'ont quitté pour une autre province pendant la période étudiée.

Ces estimations ont été calculées au moyen des mêmes coefficients de survie que ceux qui ont été utilisés pour les personnes qui sont nées hors de la province de Western Cape et y ont immigré, mais si on avait quelque raison de penser que ces deux catégories de migrants avaient des schémas de mortalité différents, il faudrait recourir à deux séries différentes de coefficients de survie pour évaluer la migration totale nette des natifs de la province.

L'immigration interne nette totale de la province figure dans la dernière colonne du tableau 35.3. Donc, globalement, le nombre d'immigrants vers la province de Western Cape excède de 232 928 unités celui des émigrants qui ont quitté cette province pour une autre.

Dans cet exemple, les personnes nées hors de la province comprennent celles qui sont nées à l'étranger, et donc la migration nette inclut les immigrants internationaux qui sont venus s'établir dans la province. En retranchant du tableau 35.2 les personnes nées à l'étranger, on obtiendrait, pour chaque région, le nombre des immigrants internes net des émigrants internes, et la somme de ces effectifs pour toutes les provinces serait nulle.

C) Mesure de la migration interne entre régions d'un même pays à partir des données sur le lieu de résidence lors du recensement précédent

Le tableau 35.4 présente le résultat des réponses à la question, posée aux habitants des provinces lors de la *Community Survey* de 2007, sur leur lieu (ici la province) de résidence à l'époque du recensement de 2001. En réalité, il était demandé à chaque personne si elle habitait au même endroit que lors du dernier recensement, et, si la réponse était négative,

quelle était sa dernière résidence. Cependant, une étude de Dorrington et Moultrie (2009) incite à penser qu'il était assez raisonnable de supposer que les gens n'avaient pas migré plus d'une fois au cours des cinq années suivant le dernier recensement.

Mais le plus grand nombre de migrants est, de loin, celui des personnes qui ont déménagé à l'intérieur d'une même province. Nous les avons exclues du tableau 35.4, car on s'intéresse généralement plus aux migrations interprovinciales qu'aux migrations intra-provinciales.

En plus des effectifs tous âges confondus du tableau 35.4 (en fait, comme c'est souvent le cas, les migrations des enfants nés entre le recensement et l'enquête ont été exclus), on peut également obtenir les nombres d'immigrants et émigrants internes par groupes d'âge, comme on le voit dans le tableau 35.5. Pour l'exhaustivité, ces chiffres incluent des estimations du nombre de migrants qui sont nés depuis le recensement précédent. Cependant, comparés aux chiffres

TABLEAU 35.2 Estimation du nombre net d'immigrants internes parmi les personnes nées à l'extérieur de la province de Western Cape, Afrique du Sud, par groupe d'âge, 2001-2006

| Âge | 2001 | 2006 | x | ${}_5S_x$ | Âge au 2 ^e recensement | D_0 | M net (nés à l'extérieur) |
|-------|---------|---------|-----|-----------|-----------------------------------|--------|-----------------------------|
| | | | B | 0,94151 | | | |
| 0-4 | 16 443 | 19 012 | 0 | 0,97896 | 0-4 | 591 | 19 602 |
| 5-9 | 24 406 | 28 743 | 5 | 0,99547 | 5-9 | 482 | 12 782 |
| 10-14 | 31 134 | 30 792 | 10 | 0,99427 | 10-14 | 125 | 6 511 |
| 15-19 | 44 478 | 53 933 | 15 | 0,98602 | 15-19 | 245 | 23 043 |
| 20-24 | 74 011 | 82 526 | 20 | 0,96458 | 20-24 | 896 | 38 944 |
| 25-29 | 80 187 | 89 522 | 25 | 0,93161 | 25-29 | 2 954 | 18 466 |
| 30-34 | 65 833 | 90 783 | 30 | 0,90960 | 30-34 | 6 074 | 16 670 |
| 35-39 | 56 393 | 76 475 | 35 | 0,89780 | 35-39 | 6 776 | 17 417 |
| 40-44 | 44 420 | 59 692 | 40 | 0,89092 | 40-44 | 6 268 | 9 567 |
| 45-49 | 32 862 | 47 612 | 45 | 0,88633 | 45-49 | 5 338 | 8 529 |
| 50-54 | 28 178 | 37 969 | 50 | 0,87224 | 50-54 | 4 303 | 9 409 |
| 55-59 | 19 983 | 30 205 | 55 | 0,84731 | 55-59 | 4 012 | 6 039 |
| 60-64 | 17 569 | 25 593 | 60 | 0,80885 | 60-64 | 3 832 | 9 442 |
| 65-69 | 11 216 | 20 802 | 65 | 0,75468 | 65-69 | 4 137 | 7 371 |
| 70-74 | 8 365 | 12 612 | 70 | 0,66991 | 70-74 | 3 426 | 4 822 |
| 75-79 | 5 919 | 8 434 | 75 | 0,56388 | 75-79 | 3 458 | 3 528 |
| 80-84 | 4 063 | 5 061 | 80+ | 0,40912 | 80-84 | 3 248 | 2 390 |
| 85+ | 2 152 | 2 183 | | | 85+ | 3 413 | - 620 |
| TOTAL | 567 613 | 721 949 | | | TOTAL | 59 576 | 213 911 |

des autres groupes d'âge, ces nombres paraissent exagérément élevés, nous en examinerons la raison plus loin.

Le nombre net de migrants se calcule, par exemple pour les personnes âgées de 25 à 29 ans au moment de la *Community Survey* (qui avaient donc de 20 à 24 ans lors du recensement de 2001), de la manière suivante :

$${}_5M_x = (20675 - 5649 + (20675 - 5649) / 0,96458 / 2) = 15301.$$

DIAGNOSTICS, ANALYSE ET INTERPRÉTATION

Contrôles et validation

Le contrôle peut-être le plus simple, sur l'allure « acceptable » de la distribution des effectifs (c'est-à-dire leur structure par âge), mais pas sur leur niveau, consiste à regarder si elle s'appa-

rente au profil standard (ou à une de ses variantes). Rogers et Castro (1981a, 1981b) signalent que la distribution des effectifs (ou des taux) d'immigrants et émigrants internes est généralement conforme au standard, avec un pic chez les jeunes adultes (habituellement en lien avec la recherche d'un emploi) et un pic, moins marqué, pour les très jeunes enfants, suivi d'un creux chez les adolescents (son ampleur variant selon que, parmi les adultes jeunes ou d'âge moyen, il s'agit surtout de migrations familiales ou plutôt de migrations individuelles). Il y a aussi parfois une bosse (ou un creux) vers la fin de la vie active, s'il existe un flux important de personnes qui immigrer (ou émigrer) quand elles prennent leur retraite.

Ces schémas (pas nécessairement identiques) s'appliquent séparément aux flux d'immigration et d'émigration internes, mais pas automatiquement à la migration nette (la différence entre les deux flux), à moins que l'un de ces flux soit beaucoup plus important que l'autre.

TABLEAU 35.3 Estimation du nombre net d'émigrants internes parmi les personnes nées dans la province de Western Cape, par groupe d'âge, 2001-2006

| Âge | 2001 | 2006 | x | ${}_5S_x$ | Âge au 2 ^e recensement | D_I | M net (nés dans la province) | M net total |
|-------|---------|---------|-----|-----------|-----------------------------------|--------|--------------------------------|---------------|
| | | | B | 0,94151 | | | | |
| 0-4 | 22 055 | 11 747 | 0 | 0,97896 | 0-4 | 365 | 12 112 | 7 490 |
| 5-9 | 21 895 | 12 509 | 5 | 0,99547 | 5-9 | 367 | - 9 180 | 21 962 |
| 10-14 | 21 382 | 11 593 | 10 | 0,99427 | 10-14 | 76 | - 10 226 | 16 737 |
| 15-19 | 18 265 | 13 455 | 15 | 0,98602 | 15-19 | 100 | - 7 827 | 30 870 |
| 20-24 | 14 645 | 10 477 | 20 | 0,96458 | 20-24 | 202 | - 7 587 | 46 531 |
| 25-29 | 13 501 | 9 534 | 25 | 0,93161 | 25-29 | 434 | - 4 676 | 23 142 |
| 30-34 | 13 118 | 11 047 | 30 | 0,90960 | 30-34 | 867 | - 1 587 | 18 257 |
| 35-39 | 12 121 | 14 614 | 35 | 0,89780 | 35-39 | 1 319 | 2 815 | 14 602 |
| 40-44 | 11 725 | 12 195 | 40 | 0,89092 | 40-44 | 1 311 | 1 384 | 8 183 |
| 45-49 | 10 335 | 10 538 | 45 | 0,88633 | 45-49 | 1 285 | 98 | 8 431 |
| 50-54 | 9 211 | 9 881 | 50 | 0,87224 | 50-54 | 1 221 | 768 | 8 642 |
| 55-59 | 7 264 | 10 568 | 55 | 0,84731 | 55-59 | 1 362 | 2 720 | 3 319 |
| 60-64 | 6 691 | 7 723 | 60 | 0,80885 | 60-64 | 1 250 | 1 710 | 7 732 |
| 65-69 | 4 643 | 5 297 | 65 | 0,75468 | 65-69 | 1 265 | - 128 | 7 499 |
| 70-74 | 3 954 | 3 766 | 70 | 0,66991 | 70-74 | 1 182 | 304 | 4 517 |
| 75-79 | 2 331 | 2 384 | 75 | 0,56388 | 75-79 | 1 240 | - 330 | 3 858 |
| 80-84 | 1 402 | 2 140 | 80+ | 0,40912 | 80-84 | 1 336 | 1 145 | 1 244 |
| 85+ | 707 | 555 | | | 85+ | 1 024 | - 531 | - 89 |
| TOTAL | 195 246 | 160 023 | | | TOTAL | 16 206 | - 19 017 | 232 928 |

TABLEAU 35.4 Migrations interprovinciales, Afrique du Sud, 2001-2006

| Résidence précédente (origine) | Province de résidence (destination) | | | | | | | | | Total |
|--------------------------------------|-------------------------------------|---------------|---------------|---------------|---------------|---------------|----------------|---------------|---------------|----------------|
| | WC | EC | NC | FS | KZ | NW | GT | MP | LM | |
| WC | | 12 173 | 4 060 | 1 745 | 3 221 | 2 113 | 16 400 | 1 405 | 874 | 41 992 |
| EC | 52 239 | | 1 120 | 7 187 | 25 209 | 14 430 | 28 633 | 4 693 | 2 116 | 135 626 |
| NC | 4 813 | 1 942 | | 3 480 | 908 | 3 728 | 4 956 | 1 062 | 357 | 21 246 |
| FS | 2 943 | 3 145 | 2 546 | | 2 352 | 12 733 | 19 920 | 4 293 | 1 963 | 49 896 |
| KZ | 6 762 | 7 015 | 631 | 2 358 | | 3 573 | 50 980 | 8 886 | 1 194 | 81 399 |
| NW | 1 478 | 907 | 9 811 | 5 555 | 2 329 | | 47 633 | 3 090 | 4 337 | 75 140 |
| GT | 24 891 | 12 948 | 3 962 | 11 437 | 18 145 | 32 433 | | 18 598 | 15 133 | 137 547 |
| MP | 2 134 | 1 317 | 280 | 1 724 | 4 546 | 5 767 | 42 941 | | 8 628 | 67 338 |
| LM | 2 754 | 1 583 | 255 | 1 709 | 2 209 | 9 773 | 81 394 | 24 211 | | 123 889 |
| HAS | 21 221 | 5 467 | 1 209 | 9 584 | 10 933 | 11 437 | 51 873 | 8 335 | 9 286 | 129 346 |
| INC | 500 | 3 | 15 | 124 | 132 | 78 | 228 | 89 | 0 | 1 170 |
| NR | 1 058 | 1 029 | 107 | 208 | 875 | 508 | 3 558 | 408 | 633 | 8 384 |
| TOTAL | 120 794 | 47 528 | 23 996 | 45 111 | 70 860 | 96 573 | 348 516 | 75 070 | 44 524 | 872 973 |

WC = Western Cape, EC = Eastern Cape, NC = Northern Cape, FS = Free State, KZ = KwaZulu-Natal,
 NW = North West, GT = Gauteng, MP = Mpumalanga, LM = Limpopo, HAS = Hors d'Afrique du Sud, INC = Inconnu,
 NR = Non-réponse.

TABLEAU 35.5 Estimation du nombre net d'immigrants internes par groupes d'âge, Western Cape, Afrique du Sud, 2001-2006

| Âge | Immigrants survivants (I') | Émigrants survivants (O') | x | ${}_5S_x$ | Immigrants internes nets |
|--------------|----------------------------------|---------------------------------|-----|-----------|--------------------------------|
| 0-4 | 20 846 | 11 747 | B | 0,94151 | 9 381 |
| 5-9 | 6 586 | 3 554 | 0 | 0,97896 | 3 065 |
| 10-14 | 6 685 | 2 882 | 5 | 0,99547 | 3 812 |
| 15-19 | 10 402 | 3 967 | 10 | 0,99427 | 6 454 |
| 20-24 | 21 266 | 4 488 | 15 | 0,98602 | 16 897 |
| 25-29 | 20 675 | 5 649 | 20 | 0,96458 | 15 301 |
| 30-34 | 15 584 | 6 008 | 25 | 0,93161 | 9 928 |
| 35-39 | 10 584 | 5 098 | 30 | 0,90960 | 5 758 |
| 40-44 | 7 264 | 3 045 | 35 | 0,89780 | 4 458 |
| 45-49 | 4 648 | 2 714 | 40 | 0,89092 | 2 053 |
| 50-54 | 3 095 | 1 500 | 45 | 0,88633 | 1 698 |
| 55-59 | 3 940 | 935 | 50 | 0,87224 | 3 225 |
| 60-64 | 3 776 | 527 | 55 | 0,84731 | 3 541 |
| 65-69 | 3 127 | 818 | 60 | 0,80885 | 2 582 |
| 70-74 | 1 540 | 437 | 65 | 0,75468 | 1 282 |
| 75-79 | 561 | 206 | 70 | 0,66991 | 442 |
| 80-84 | 797 | 116 | 75 | 0,56388 | 944 |
| 85+ | 264 | 47 | 80+ | 0,40912 | 374 |
| TOTAL | 141 640 | 53 739 | | | 91 194 |

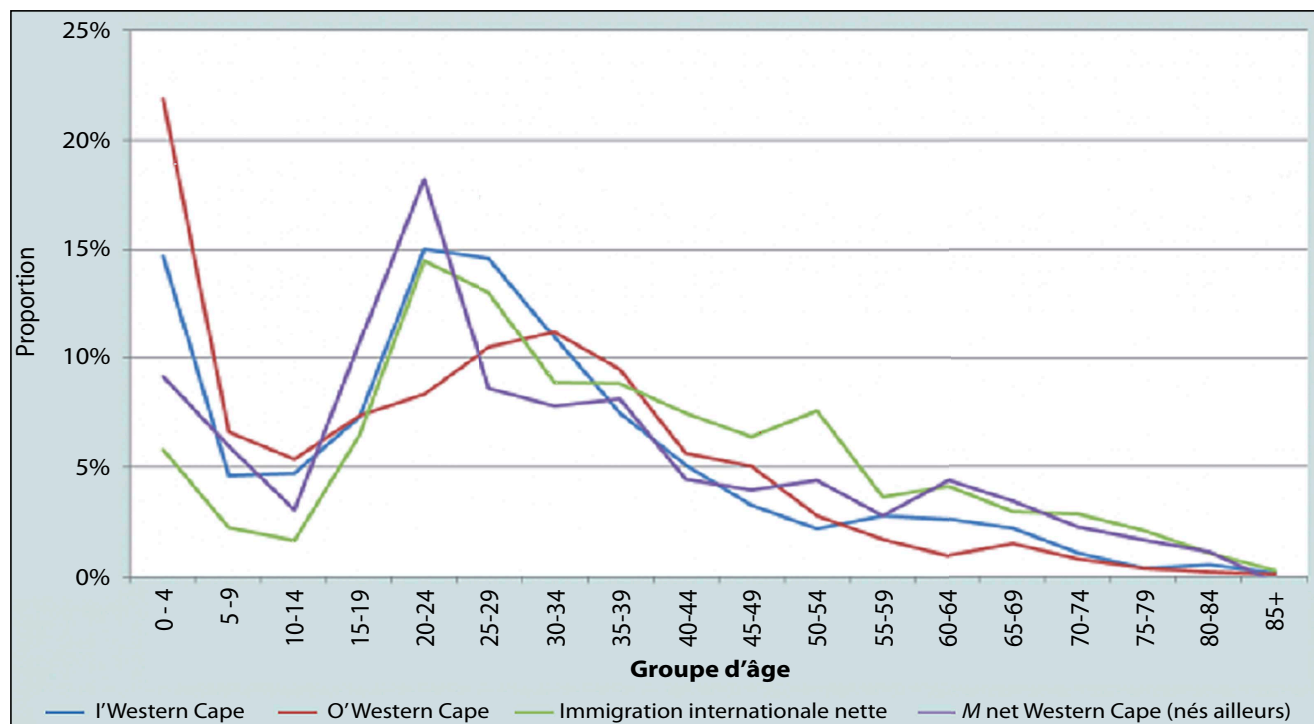
C'est ce qu'illustre la figure 35.1, en utilisant certains des résultats calculés plus haut, sous forme de proportions par rapport au total (pour qu'ils puissent tenir sur une seule figure). On peut y constater que, grosso modo (à l'exception de certains cas où la proportion de migrants aux très jeunes âges semble exagérément élevée), les flux ont le profil attendu.

Les émigrants internes nets nés au Western Cape (non présentés sur la figure pour qu'elle reste lisible) s'écartent d'un schéma standard de migration, ce qui pourrait signifier que ces chiffres ne sont pas très sûrs, mais ils sont peu nombreux en regard des immigrants internes nés hors de la province, et on peut donc tolérer cette distorsion. De plus, deux autres caractéristiques de la figure 35.1 méritent d'être soulignées. Premièrement, l'émigration interne du Western Cape, telle qu'elle a été évaluée à partir de données sur le lieu de résidence au recensement précédent, indique que le pic

des émigrants adultes se situe à un âge un peu plus avancé (et qu'il s'agit peut-être de migrations familiales plutôt qu'individuelles). Deuxièmement, l'immigration internationale nette se conforme au profil standard, ce qui signifie que le flux d'immigration vers le pays est beaucoup plus important que le flux de retour de ces immigrants.

Si, lors du recensement, on a demandé le lieu de naissance et le lieu de résidence au recensement précédent, on peut comparer les deux évaluations de l'immigration interne nette d'une région donnée. Si elles coïncident, cela inspire une certaine confiance dans les résultats. Dans le cas des données sur le lieu de naissance en Afrique du Sud, le nombre net d'immigrants internes au Western Cape est de 232 928 (tableau 35.3), tandis que l'estimation à partir du lieu de résidence au recensement précédent n'est que de 92 194 (tableau 35.4), ce qui incite à penser que l'une au moins de ces deux séries de données est suspecte.

FIGURE 35.1 Structure par âge de quelques flux migratoires, Afrique du Sud, sexe masculin, 2001-2006



Le contrôle le plus élémentaire que l'on puisse faire sur les estimations de la migration consiste à projeter la population (du pays ou de la province) du premier recensement jusqu'à la date du second, en utilisant les estimations de nombres de migrants, et de comparer les résultats obtenus avec les chiffres du second recensement pour voir dans quelle mesure ils concordent, surtout dans les tranches d'âge où se concentre le phénomène migratoire. Dans le cas de l'immigration interne nette au Western Cape, projeter la population de 2001 en utilisant les estimations déduites de la variation des effectifs par lieu de naissance a abouti à une concordance nettement meilleure dans la tranche d'âge de 20 à 29 ans, ce qui veut dire que les données sur le lieu de naissance sont sans doute plus complètes que les données sur le lieu de résidence au recensement précédent. C'est ce que confirme, dans une certaine mesure, la comparaison de la variation intercensitaire du nombre de résidents du pays nés à l'étranger, 222 693 (tableau 35.1), avec le total des résidents qui ont déclaré avoir migré de l'étranger vers l'une des provinces sud-africaines depuis le recensement précédent, 129 346 (tableau 35.4).

Idéalement, si on disposait d'estimations indépendantes des nombres de migrants, on pourrait les comparer avec les estimations obtenues par les méthodes exposées ci-dessus. Malheureusement, les estimations indépendantes fiables sont rares. Bien que la plupart des pays essaient d'enregistrer les émigrants et les immigrants, ces données ne sont généralement pas fiables, surtout dans les pays en développement aux frontières relativement poreuses. À moins que le pays dispose de réglementations extrêmement strictes et tienne un registre exact et complet de sa population, la seule manière alternative de mesurer la migration interne consiste à réaliser des enquêtes spécifiquement consacrées aux migrations, qui s'avèrent beaucoup plus utiles pour analyser les types de migrations (permanentes, temporaires, circulaires, etc.) que pour fournir de bonnes estimations des nombres de migrants, étant donné la situation souvent assez floue dans laquelle vivent les immigrés (surtout les immigrés récents) et leur réticence compréhensible à s'identifier comme migrants.

Interprétation

Si on considère les nombres de migrants évalués à partir des données sur le lieu de résidence au recensement précédent, qui figurent dans le tableau 35.4 (sans oublier qu'ils sous-estiment probablement le volume réel des migrations), de 2 à

4 % de la population ont migré d'une province à l'autre au cours des 5 ans qui séparent le recensement de 2001 de la *Community Survey*. En comptant aussi les personnes qui ont déménagé sans changer de province, ce sont de 7 à 15 % des habitants qui ont changé de résidence dans cette période de 5 ans.

Les principales provinces de destination sont celles de Gauteng (largement en tête) et de Western Cape, qui sont essentiellement urbaines et les plus riches du pays. Les principales provinces d'origine sont celles de Gauteng (l'examen de la répartition par âge des migrants montrerait qu'il s'agit essentiellement de migrations de retour de travailleurs qui prennent leur retraite), d'Eastern Cape et de Limpopo, ces deux dernières étant des provinces pauvres et principalement rurales, que l'on quitte pour aller chercher de l'emploi dans les régions urbaines.

Il semble que les migrations soient majoritairement individuelles (recherche d'emploi) plutôt que familiales.

PROBLÈMES D'INTERPRÉTATION PROPRES À LA MÉTHODE

Erreurs de scanner

À en croire les données basées sur la province de naissance, il y aurait un nombre relativement élevé d'enfants nés après le premier recensement qui auraient migré vers une autre province. Selon toute probabilité, il s'agit d'un effet fortuit du procédé de saisie des données, qui a consisté à scanner les questionnaires remplis manuellement. Dans un petit nombre de cas, le scanner a pu avoir du mal à distinguer le code du Western Cape, « 1 », de celui de Gauteng, « 7 ». Il en est résulté, par exemple, que certains enfants étaient enregistrés comme nés hors de la province où ils étaient recensés, et donc considérés comme migrants, alors qu'ils ne l'étaient probablement pas. Même si cette erreur de scanner ne touche qu'une très petite proportion des naissances, celles-ci peuvent être beaucoup plus nombreuses que les migrants, et cela peut entraîner une quantité non négligeable de données erronées. Comme de plus en plus de pays en développement saisissent les données par scanner, ce genre de problème pourrait être assez répandu.

Quand des erreurs de scanner ou autres empêchent d'obtenir des estimations fiables des nombres de migrants parmi les enfants nés depuis le recensement précédent, on peut recourir aux ratios enfants/femmes (*en anglais* : *CWR*) du second recensement :

$$\text{Net}_5 M_0 = \frac{1}{4} CWR_0 \cdot \text{Net}_{30} M_{15}^f$$

pour les enfants nés au cours des 5 dernières années, et

$$\text{Net}_5 M_5 = \frac{3}{4} CWR_5 \cdot \text{Net}_{30} M_{20}^f$$

pour ceux qui sont nés au cours des 5 années précédant celles-ci si les recensements sont espacés de 10 ans, CWR_x représentant le rapport du nombre d'enfants âgés de x à $x + 5$ au nombre de femmes âgées de $15 + x$ à $45 + x$ (dans la population du pays ou de la province) au moment du second recensement, et ${}_{30}M_x^f$ représentant le nombre de femmes migrantes âgées de x à $x + 30$.

L'application aux données du Western Cape indique que le nombre de migrants nés depuis le premier recensement devrait être inférieur à la moitié de l'estimation obtenue à partir des données sur le lieu de naissance.

DESCRIPTION DÉTAILLÉE DE LA MÉTHODE

Présentation mathématique

L'estimation indirecte de la migration se déduit de l'équation qui relie deux recensements espacés de n années :

$$\begin{aligned} {}_5N_{x+n}(t+n) &= {}_5N_x(t) - {}_5D_x + {}_5I'_x - {}_5O'_x \\ &= {}_5N_x(t) - {}_5D_x + {}_5M'_x \end{aligned}$$

où ${}_5M'_x = {}_5I'_x - {}_5O'_x$ est le nombre net d'immigrants internes (les entrants moins les sortants) âgés de x à $x + 5$ à la date du premier recensement et toujours en vie au second recensement, et ${}_5D_x$, ${}_5I'_x$ et ${}_5O'_x$ sont respectivement les nombres de décédés, d'immigrés survivants et d'émigrés survivants âgés de x à $x + 5$ lors du premier recensement, qui sont morts ou ont migré entre les deux recensements.

Pour les enfants nés après le premier recensement, l'équation devient :

$${}_nN_0(t+n) = B - D_B + M'_B$$

et pour l'intervalle d'âge ouvert :

$${}_\infty N_A(t+n) = {}_\infty N_{A-n}(t) - {}_\infty D_{A-n} + {}_\infty M'_{A-n}$$

où B représente le nombre de naissances enregistrées entre les deux recensements, D_B le nombre de décès survenus parmi ces enfants entre les deux recensements, M'_B le nombre net d'immigrants survivants nés à l'étranger entre les deux

recensements, ${}_\infty D_{A-n}$ le nombre de personnes âgées de $A - n$ et plus à l'époque du premier recensement et qui sont décédées pendant la période intercensitaire, et ${}_\infty M'_{A-n}$ le nombre net d'immigrants qui étaient âgés de $A - n$ et plus lors du premier recensement. Ainsi,

$${}_5M'_x = {}_5N_{x+n}(t+n) - {}_5N_x(t) + {}_5D_x$$

$$M'_B = {}_nN_0(t+n) - B + D_B$$

$${}_\infty M'_{A-n} = {}_\infty N_A(t+n) - {}_\infty N_{A-n}(t) + {}_\infty D_{A-n}$$

ou encore

$${}_5M'_x = {}_5N_{x+n}(t+n) - {}_5N_x(t) {}_5S_x$$

$$M'_B = {}_nN_0(t+n) - B {}_5S_B$$

$${}_\infty M'_{A-n} = {}_\infty N_A(t+n) - {}_\infty N_{A-n}(t) {}_\infty S_{A-n}$$

où ${}_5S_x$, S_B et ${}_\infty S_{A-n}$ représentent respectivement les proportions de survivants au second recensement parmi la population âgée de x à $x + 5$ au premier recensement, parmi les enfants nés pendant la période intercensitaire et parmi les personnes âgées de $A - n$ et plus au moment du premier recensement.

On peut alors évaluer le nombre net d'immigrants à partir du nombre net de survivants au second recensement :

$${}_5M_x = ({}_5M'_x + {}_5M'_x / {}_5S_x) / 2 = {}_5M'_x \frac{({}_5S_x + 1)}{2 {}_5S_x}$$

$$M_B = M'_B \frac{(S_B + 1)}{2 S_B}$$

$${}_\infty M_{A-n} = {}_\infty M'_{A-n} \frac{({}_\infty S_{A-n} + 1)}{2 {}_\infty S_{A-n}}$$

Malheureusement, comme le nombre net d'immigrants est généralement faible par rapport au volume de la population, les erreurs d'âge ou de dénombrement, dans l'un des recensements ou dans les deux, peuvent conduire à des estimations très douteuses. On peut obtenir de meilleures estimations du nombre net d'immigrants dans un pays en limitant l'analyse à la population d'origine étrangère (définie comme l'ensemble des personnes nées à l'étranger) et en posant l'hypothèse que les retours d'autochtones émigrés sont négligeables. On remplace alors les symboles des équations ci-dessus par leurs équivalents en termes de population née à l'étranger. Comme on ne disposera probablement pas d'un compte précis des décès de personnes nées à l'étranger, on devra l'estimer de l'une ou l'autre des manières suivantes :

- *Option 1 (probabilités de survie de la table de mortalité)* : en appliquant les probabilités d'une table-type de mortalité appropriée,

$${}_5S_x = \frac{{}_5L_{x+n}}{{}_5L_x}, S_B = \frac{nL_0}{n \cdot l_0} \text{ et } {}_\infty S_{A-n} = \frac{T_A}{T_{A-n}}.$$

- *Option 2 (probabilités de survie intercensitaires)* : en supposant que l'émigration d'autochtones est négligeable et que les proportions de survivants sont les mêmes chez les étrangers que chez les autochtones

$${}_5S_x = \frac{{}_5N_{x+n}^{nb}(t+n)}{{}_5N_x^{nb}(t)}, S_B = \frac{nN_0^{nb}}{B^{nb}} \text{ et } {}_\infty S_{A-n} = \frac{{}_\infty N_A^{nb}(t+n)}{{}_\infty N_{A-n}^{nb}(t)},$$

l'exposant *nb* identifiant les autochtones.

- *Option 3 (données d'état civil)* : si on dispose des nombres de naissances et de décès provenant d'une autre source, comme l'état civil (sans doute la seule possibilité quand il s'agit de migration interne), on peut travailler directement avec les nombres de naissances et de décès correspondant à la population migrante, plutôt qu'avec des probabilités de survie, pour estimer le nombre net d'immigrés internes survivants. Le nombre net d'immigrants peut également être calculé comme ci-dessus en posant :

$${}_5S_x = 1 - \frac{{}_5D_x}{{}_5N_x(t)}, S_B = \frac{D_B}{B} \text{ et } {}_\infty S_{A-n} = \frac{D_{A-n}}{{}_\infty N_{A-n}(t)}$$

où les nombres de naissances et de décès proviennent de l'état civil.

Mais dans la plupart des pays en développement, et surtout les pays africains, les systèmes d'état civil sont trop incomplets pour être utilisés de cette manière.

La migration interne

Pour ce qui est de la migration interne, on peut évaluer, pour chaque région, l'immigration interne nette de personnes nées hors de la région (c'est-à-dire le nombre de ces personnes qui y ont immigré, moins le nombre de celles d'entre elles qui en sont reparties) en exploitant les données sur le lieu de naissance pour évaluer la variation du nombre de personnes nées ailleurs, de la même façon que ci-dessus. Mais, comme nous connaissons aussi le lieu de résidence des personnes qui sont nées dans la région et ont migré vers une autre région (mais pas vers l'étranger), nous pouvons évaluer l'émigration interne nette des natifs de la région (soit le nombre de natifs de la région qui l'ont quittée pour une autre, moins le nombre

de ceux-ci qui y sont revenus) en appliquant la méthode décrite plus haut à la seule population née dans la région considérée.

Pour l'évaluation de la survie des personnes natives des différentes régions, les probabilités de survie intercensitaires pourraient prendre l'avantage sur celles de la table de mortalité, car toute erreur de dénombrement, en plus ou en moins, de la population des régions peut être compensée par une erreur du même type au niveau de la population du pays et, par suite, les probabilités de survie donneraient une meilleure estimation des nombres de migrants que celles d'une table de mortalité.

En plus de leur lieu de naissance, le recensement peut demander aux personnes qui ont migré depuis le recensement précédent (ou depuis une autre date) où elles résidaient à ce moment-là, ce qui permet de calculer l'émigration interne et aussi l'immigration interne (brute) séparément pour chaque région du pays.

Si le recensement demande l'année de la migration (ou depuis combien de temps la personne réside à l'endroit où elle est recensée), on peut avoir une idée du calendrier de la migration et évaluer des taux annuels de migration. C'est un processus complexe, et nous ne le traiterons pas ici, mais nous renvoyons le lecteur intéressé à l'article de Dorrington et Moultrie (2009).

Travailler avec les seuls effectifs totaux

Si on ne dispose pas des effectifs par âge, ou si la distribution par âge est considérée comme douteuse, on peut encore obtenir des mesures par âge en évaluant le nombre total de migrants comme on l'a décrit plus haut et en répartissant ensuite ce total par groupes d'âge en appliquant soit la structure par âge de la même population à un autre moment (étant donné que la structure par âge des flux migratoires est généralement stable dans le temps) ou (de préférence) une structure-type appropriée, telle que celle présentée par Rogers et Castro (1981a, 1981b). Par conséquent,

$$\text{Net } {}_\infty M_0^F = {}_\infty N_0^F(t+n) - {}_\infty N_0^F(t) + {}_\infty D_0^F$$

où

$${}_\infty D_0^F = \frac{n}{2} ({}_\infty N_0^F(t) + {}_\infty N_0^F(t+n)) {}_\infty m_0$$

et ${}_\infty m_0$ est une estimation du taux brut de mortalité de la population du pays recensé.

Limites

Le principal problème associé à l'utilisation des recensements pour évaluer l'immigration internationale et l'immigration interne nette est la qualité des dénombrements, en particulier l'ampleur des éventuels sous-dénombrements, surtout si elle varie d'un recensement à l'autre. Mais, même si le sous-dénombrement est faible, le recensement peut ne pas saisir tous les migrants. Souvent, les immigrés récents ne sont pas facilement identifiés par un recensement, parce que leur installation n'est pas définitive. De plus, les immigrants peuvent ne pas être enclins à s'identifier comme tels et préférer se soustraire au recensement ou cacher leur origine étrangère.

Ensuite, dans le cas des migrants internes, la déclaration du lieu de naissance et/ou de résidence au recensement précédent pourrait être erronée du fait de modifications de limites administratives ou à cause de l'ignorance (éventuellement le biais) de la personne interrogée.

Le troisième inconvénient du recensement est qu'il ne permet pas de mesurer l'émigration vers l'étranger. Ce phénomène est particulièrement difficile à évaluer dans la plupart des pays, mais une possibilité consiste à appliquer la méthode d'estimation de l'immigration internationale nette d'étrangers, décrite plus haut, aux recensements des principaux pays de destination des émigrants pour évaluer l'évolution des effectifs de migrants vers ces pays. Bien sûr, ce procédé n'est réellement utile que si les recensements de ces divers pays mesurent avec une précision suffisante les nombres de personnes nées à l'étrangers (migrants) par pays de naissance.

En général, les statistiques sur les immigrants, et plus encore celles sur les émigrants, qui sont recueillies aux postes-frontières donnent de mauvaises évaluations des effectifs réels, à moins que les frontières du pays ne soient quasi impénétrables et qu'il n'y ait qu'un petit nombre de points d'entrée parfaitement contrôlés. Même dans ce cas, il peut toujours y avoir un certain nombre de « visiteurs » qui finissent par rester dans le pays.

AUTRES LECTURES ET RÉFÉRENCES

Pour un aperçu général de la problématique des migrations, la définition des termes et des détails sur l'analyse et l'interprétation des données sur les migrations internes, nous renvoyons le lecteur intéressé à l'excellent *Manuel VI* des Nations Unies (UN Population Division 1970).

Une dernière difficulté apparaît quand on travaille sur des données agrégées sans distinction des âges. En général, il faut alors utiliser le taux brut de mortalité de la population du pays recensé pour estimer le nombre de décès de migrants. Cependant, la structure par âge des migrants pouvant être très différente de celle de la population totale, le nombre de décès de migrants calculé peut s'avérer tout à fait faux.

EXTENSIONS DE LA MÉTHODE

Certains recensements posent des questions supplémentaires qui peuvent aider à comprendre la migration, sinon à améliorer l'évaluation de son niveau. La plus fréquente, sans doute, porte sur la date de la migration. Ces données permettent d'évaluer des taux annuels de migration, mais il se peut que les personnes interrogées aient tendance à déclarer des dates de migration plus récentes que ce n'est le cas en réalité (Dorrington et Moultrie 2009).

Quand le recensement demande, comme l'ont fait les derniers recensements sud-africains, aux personnes qui ont immigré depuis le recensement précédent de quand date leur dernière migration et d'où elles venaient, et non où elles résidaient à la date du recensement précédent, on peut rétroprojeter les effectifs de migrants en appliquant les taux annuels de migration entre régions pour en évaluer le nombre par région au moment du recensement précédent (Dorrington et Moultrie 2009). Mais, au moins dans le cas de l'Afrique du Sud, l'hypothèse que la plupart des migrants n'ont fait qu'une migration au cours des cinq dernières années, et donc que le lieu de résidence avant la dernière migration coïncide avec le lieu de résidence au moment du recensement précédent, paraît tout à fait raisonnable (Dorrington et Moultrie 2009).

Si on dispose à la fois de données sur la région de naissance et la région de résidence lors du recensement précédent, on peut croiser ces deux variables et donc classer les migrants récents selon qu'ils sont migrants primaires, migrants secondaires ou migrants de retour.

Le manuel de Shryock et Siegel (1976) ou sa version moderne due à Siegel et Swanson (2004) donnent également une introduction générale à l'étude des migrations et couvrent, en particulier, l'évaluation des migrations internationales.

Aux personnes intéressées par l'évaluation des taux annuels de migration et la rétroprojection des migrants pour en évaluer les effectifs par région de résidence au moment du recensement précédent, à partir des données sur le lieu de résidence avant la dernière migration et la date de celle-ci, nous recommandons la lecture de l'article de Dorrington et Moultrie (2009).

Dorrington RE and TA Moultrie. 2009. "Making use of the consistency of patterns to estimate age-specific rates of interprovincial migration in South Africa," Paper presented at Annual conference of the Population Association of America. Detroit, US, 30 April - 2 May.

Rogers A and LJ Castro. 1981a. "Age patterns of migration: Cause-specific profiles," in Rogers, A (ed). *Advances in Multi-regional Demography (RR-81-006)*. Laxenburg, Austria: International Institute for Applied Systems Analysis, pp. 125-159. <http://webarchive.iiasa.ac.at/Admin/PUB/Documents/RR-81-006.pdf>

Rogers A and LJ Castro. 1981b. *Model Migration Schedules (RR-81-030)*. Laxenburg, Austria: International Institute for Applied Systems Analysis. <http://webarchive.iiasa.ac.at/Admin/PUB/Documents/RR-81-030.pdf>

Shryock HS and JS Siegel. 1976. *The Methods and Materials of Demography (Condensed Edition)*. San Diego: Academic Press.

Siegel JS and D Swanson. 2004. *The Methods and Materials of Demography*. Amsterdam: Elsevier.

Timæus IM. 2004. "Impact of HIV on mortality in Southern Africa: Evidence from demographic surveillance," Paper presented at Seminar of the IUSSP Committee "Emerging Health Threats" HIV, Resurgent Infections and Population Change in Africa. Ougadougou, 12-14 February.

UN Population Division. 1970. *Manual VI: Methods of Measuring Internal Migration*. New York: United Nations, Department of Economic and Social Affairs, ST/SOA/Series A/47. <http://www.un.org/esa/population/techcoop/IntMig/manual6/manual6.html>

Chapitre 36 LE MODÈLE MULTI-EXPONENTIEL DE MIGRATION

Jani Little et Rob Dorrington

DESCRIPTION DE LA MÉTHODE

Ce chapitre explique comment ajuster un modèle multi-exponentiel de migration à des données observées sur les migrations.

Depuis une trentaine d'années, ce modèle, conçu par Rogers et Castro (1981), s'est révélé remarquablement efficace pour la représentation des structures par âge typiques des migrations. En général, on observe les mêmes structures par âge, qu'il s'agisse des migrations internationales et inter-régionales dans leur ensemble ou des migrations relatives à une région particulière. La fonction multi-exponentielle a été élaborée pour exprimer la relation entre la migration et l'âge, et elle le fait au travers d'une série additive de composantes exponentielles à 7, 9, 11 ou 13 paramètres, selon le degré de complexité du schéma de migration et de l'aptitude des données à supporter une paramétrisation croissante.

Appliqué à une série de taux de migration par année d'âge, le modèle de Rogers-Castro fournit une expression optimale et nuancée de la structure par âge de la migration, qui peut servir à lisser une série de taux de migration observés et que l'on peut utiliser directement pour améliorer la compréhension de la dynamique des migrations. Ce résultat peut trouver de nombreuses autres applications, par exemple dans la mise au point de schémas de migration utilisables dans des projections de population multirégionales. Dans l'idéal, on disposera d'indices de migration par année de calendrier et année d'âge, auxquels on ajustera le modèle de Rogers-Castro. Mais, si – comme c'est souvent le cas dans les pays en développement, où la piètre qualité des données de base peut faire obstacle au développement de calculs aussi minutieux – les données ne sont disponibles que par groupes quinquennaux d'âge, on doit estimer par interpolation des taux par année d'âge au moyen de l'une des méthodes décrites ci-après, avant de tenter l'ajustement par un modèle de Rogers-Castro.

DONNÉES REQUISES ET HYPOTHÈSES

Tabulations nécessaires

- Probabilités ou taux de migration par année d'âge (ou, si ce n'est pas possible, par groupes quinquennaux d'âge) de 0 à un âge supérieur à 65 ans.

Idéalement, les données devraient se présenter sous forme de taux par année d'âge. Quand elles sont réparties en groupes quinquennaux d'âge, il faut estimer leurs valeurs par année d'âge au moyen d'une interpolation, avant de tenter l'ajustement par une fonction multi-exponentielle. Le choix de l'âge maximum est quelque peu arbitraire, mais la borne supérieure des données utilisées pour l'ajustement d'un modèle devrait, au moins, être supérieure à l'âge modal de la retraite.

Hypothèses importantes

Le recensement national le plus récent dénombre avec exactitude la population par région de résidence et par lieu de naissance, et identifie les personnes qui ont migré d'une région à l'autre depuis une certaine date antérieure (par exemple le recensement précédent).

TRAVAUX PRÉPARATOIRES ET RECHERCHES PRÉLIMINAIRES

Avant de mettre en œuvre cette méthode, il faut examiner la qualité des données sur au moins deux aspects :

- la structure par âge de la population (éventuellement par région) ; et
- le degré de couverture des recensements (éventuellement par région).

MISE EN GARDE

La méthode ne doit être appliquée aux données de migration nette qu'avec précaution, car la distribution multi-exponen-

tielle des taux par âge modélise les flux migratoires bruts (c'est-à-dire les entrées ou les sorties), mais pas automatiquement les migrations nettes, à moins que l'un des deux flux ne soit nettement plus important que l'autre à tous les âges.

PRÉSENTATION GÉNÉRALE DU MODÈLE MULTI-EXPONENTIEL DE MIGRATION

La fonction multi-exponentielle a été élaborée par Rogers et Castro (1981) pour exprimer le lien entre la migration et l'âge. La migration est généralement forte au cours de la première année de vie ; elle chute jusqu'à un minimum au début de l'adolescence, pour augmenter rapidement jusqu'à son maximum au début de l'âge adulte, après quoi elle diminue, sauf un éventuel sursaut autour de l'âge de la retraite, suivi d'une reprise de la baisse. Dans certains cas, on peut observer une dernière remontée aux âges les plus avancés (Rogers et Castro 1981 ; Rogers et Watkins 1987).

Depuis une trentaine d'années, le système de modélisation de la migration de Rogers-Castro s'est révélé remarquablement efficace pour la représentation des structures par âge des migrations (Little et Rogers 2007 ; Raymer et Rogers 2008 ; Rogers et Castro 1981 ; Rogers et Castro 1986 ; Rogers et Little 1994 ; Rogers, Little et Raymer 2010 ; Rogers et Raymer 1999 ; Rogers et Watkins 1987). Les mêmes structures par âge des migrations ont été observées pour des régions de taille différente et pour des sous-populations distinguées selon le sexe et le groupe ethnique (Rogers et Castro 1981). On les a observées tant dans le cas des migrations interrégionales dans leur ensemble que pour les migrations relatives à une région particulière. Un flux migratoire unidirectionnel (de la région i vers la région j) présente également la même structure. Par exemple, le modèle de Rogers-Castro a été appliqué avec succès aux flux migratoires intercommunaux (entre *local authorities*) en Angleterre (Bates et Bracken 1982 ; 1987), entre zones métropolitaines et non-métropolitaines au Canada (Liaw et Nagnur 1985), entre régions au Japon, en Corée et en Thaïlande (Kawabe 1990), et au niveau national en Afrique du Sud et en Pologne (Hofmeyr 1988 ; Potrykowska 1988).

Appliqué à une série de taux de migration par année d'âge, le modèle de Rogers-Castro fournit une expression optimale et nuancée de la structure par âge de la migration, qui peut être caractérisée par 7, 9, 11 ou 13 paramètres, selon le degré de complexité de cette structure et la robustesse des données. De plus, les fluctuations aléatoires, souvent associées à l'imprécision des taux par âge observés, sont aplanies.

Des modèles de Rogers-Castro ont été utilisés pour réaliser des projections de population au Canada (George 1994), et appliqués à des périodes, des régions ou des sous-populations dont les données migratoires étaient insuffisantes ou inexistantes (Rogers, Little et Raymer 2010).

Dans sa forme la plus complète et la plus complexe, le modèle multi-exponentiel compte 13 paramètres. Si on appelle $M(x)$ le taux de migration à l'âge x , le modèle complet est défini par :

$$\begin{aligned} M(x) = & a_1 \exp(-\alpha_1 x) \\ & + a_2 \exp\{-\alpha_2(x - \mu_2) - \exp[-\lambda_2(x - \mu_2)]\} \\ & + a_3 \exp\{-\alpha_3(x - \mu_3) - \exp[-\lambda_3(x - \mu_3)]\} \\ & + a_4 \exp(\lambda_4 x) \\ & + c. \end{aligned}$$

Il comporte cinq composantes additives. La première, $a_1 \exp(-\alpha_1 x)$, est une fonction exponentielle négative représentant le schéma de migration aux *âges qui précèdent l'entrée sur le marché du travail*. La deuxième, $a_2 \exp\{-\alpha_2(x - \mu_2) - \exp[-\lambda_2(x - \mu_2)]\}$, est une fonction unimodale asymétrique à gauche qui décrit le schéma de migration par âge des personnes d'*âge actif*. La troisième, $a_3 \exp\{-\alpha_3(x - \mu_3) - \exp[-\lambda_3(x - \mu_3)]\}$, est presque une courbe en cloche représentant la fonction de migration par âge *après la retraite*, quand les migrations augmentent brusquement au moment de la retraite avant de décliner à nouveau. Associée à celle-ci, la quatrième composante, $a_4 \exp(\lambda_4 x)$, est une fonction exponentielle positive qui concerne la migration par âge *après la retraite* dans les quelques cas où l'on constate une hausse générale de la migration à cette époque de la vie. On peut observer ce phénomène, par exemple, aux États-Unis, avec les personnes âgées du Nord-Est qui migrent vers les États de la *sunbelt* du Sud-Est et du Sud-Ouest. La dernière composante est un terme constant, c , qui représente les migrations « résiduelles ».

Les recherches effectuées jusqu'à présent ont identifié quatre familles de modèles multi-exponentiels (Rogers, Little et Raymer 2010) ; une seule d'entre elles, qui présente à la fois un pic au moment de la retraite et une augmentation des migrations par la suite, utilise les 13 paramètres et les 5 composantes. Décrite dans des études sur la migration des personnes âgées (Rogers et Watkins 1987), elle est illustrée dans le cadre inférieur droit de la figure 36.1.

Les autres familles de modèles sont des formes restreintes du modèle complet, ce qui signifie qu'au moins une composante en est absente. Par exemple, le schéma le plus courant identifié par Rogers, Little et Raymer (2010) utilise sept paramètres et comporte les deux premières composantes et le terme constant. On l'appelle le schéma *standard* : son profil est présenté dans le cadre supérieur gauche de la figure 36.1.

De nombreux schémas de migration présentent le profil standard plus un pic autour de l'âge de la retraite (Rogers et Castro 1981 ; 1986), ce qui équivaut au modèle à 11 paramètres, qui comprend les composantes 1, 2, 3 et 5 (cadre inférieur gauche de la figure 36.1). Dans les populations qui comptent de nombreux travailleurs migrants, surtout dans les pays en développement, la troisième composante peut

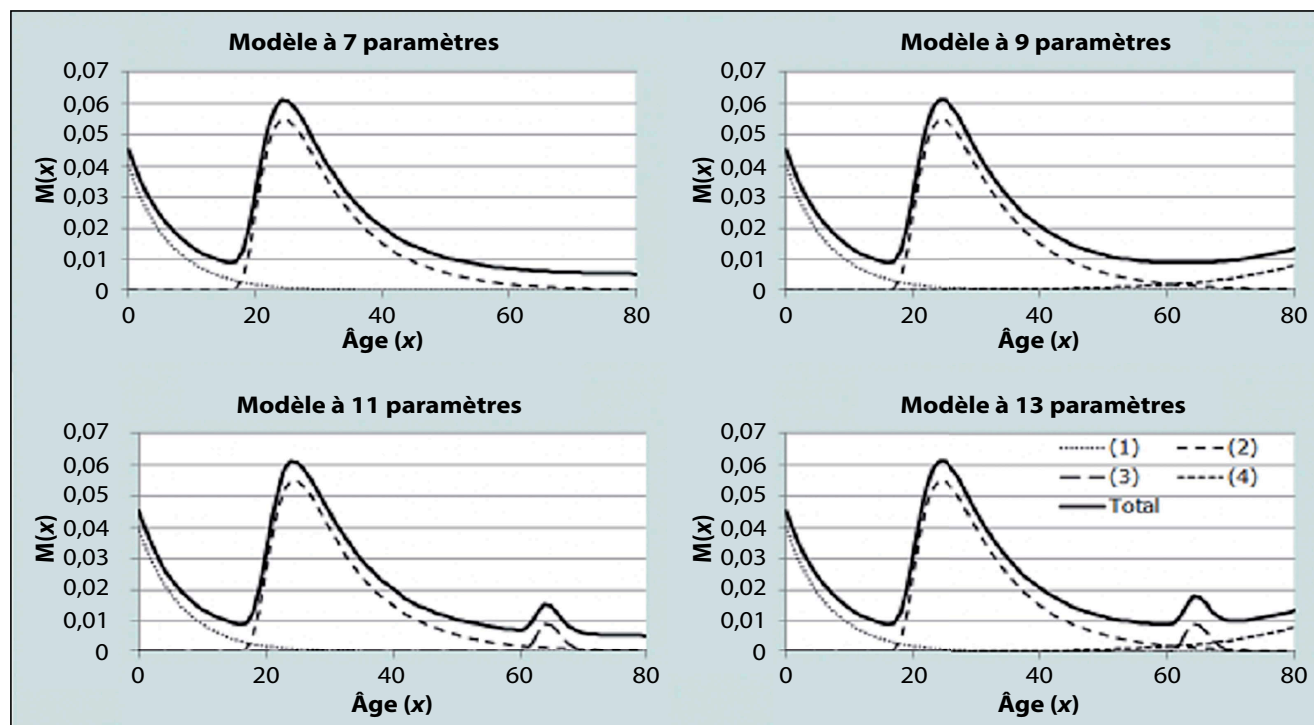
donner lieu à un creux au lieu d'un pic, puisque ces immigrants rentrent chez eux au moment de la retraite.

On emploie le modèle à 9 paramètres quand on observe le schéma standard pour les jeunes et les actifs et un redressement de la courbe au-delà de l'âge de la retraite (cadre supérieur droit de la figure 36.1). Ce cas a été observé dans diverses régions des Pays-Bas en 1974 par Rogers et Castro (1981).

Comme cela devrait être clair d'après ce qui précède, tous les paramètres sont interprétables et peuvent servir à caractériser le modèle auquel ils appartiennent.

Dans la version originale à 11 paramètres du modèle multi-exponentiel de migration, Rogers et Castro (1981) ont illustré leur modèle en l'appliquant aux taux d'émigration

FIGURE 36.1 Les quatre grandes familles de modèles de migration à composantes additives de Rogers-Castro



Source : D'après Raymer et Rogers (2008).

Note. Les courbes présentent, en séquence, (1) le schéma de migration par âge des jeunes qui ne sont pas encore sur le marché du travail ; (2) celui des personnes d'âge actif ; (3) celui des personnes retraitées, avec une hausse suivie d'une diminution ; et (4) celui des personnes retraitées avec hausse généralisée aux âges élevés.

interne masculine de Stockholm de 1974. La figure 36.2 présente les données originales (la courbe irrégulière) et la courbe lisse ajustée à 11 paramètres correspondante.

Cinq des onze paramètres (α_1 , α_2 , α_3 , λ_2 and λ_3) correspondent aux pentes croissantes ou décroissantes des divers segments de la courbe modélisée, tandis que les paramètres de niveau (a_1 , a_2 , a_3 et c) correspondent aux extremums de la courbe : a_1 pour le pic de la première année de vie, a_2 pour celui des migrations de main-d'œuvre, a_3 pour celui des migrations au moment de la retraite, c étant le taux de migration résiduelle. Les paramètres μ_2 et μ_3 indiquent les âges auxquels interviennent respectivement les pics de migration de main-d'œuvre et de migration de retraités.

On peut utiliser divers indices pour décrire soit le schéma observé soit le schéma modélisé. Par exemple, x_l est l'âge auquel la migration est la plus faible avant l'entrée sur le marché du travail, x_b est l'âge auquel elle est la plus forte chez les personnes d'âge actif, et x_r est l'âge du pic de migration au moment de la retraite. On appelle l'écart entre x_l et

x_b « labour force shift », X , et la hausse du taux de migration entre x_l et x_b « jump », B . Le « parental shift », A , décrit l'écart d'âge moyen entre la migration des parents et la migration correspondante des enfants. Le taux brut de « migraproduction » (TBM) est le total des taux à tous les âges (c'est-à-dire la surface sous la courbe), et il mesure le niveau global d'émigration interne d'une région, ou le niveau global d'un flux unidirectionnel (d'une région i vers une région j) (Rogers et Castro 1981).

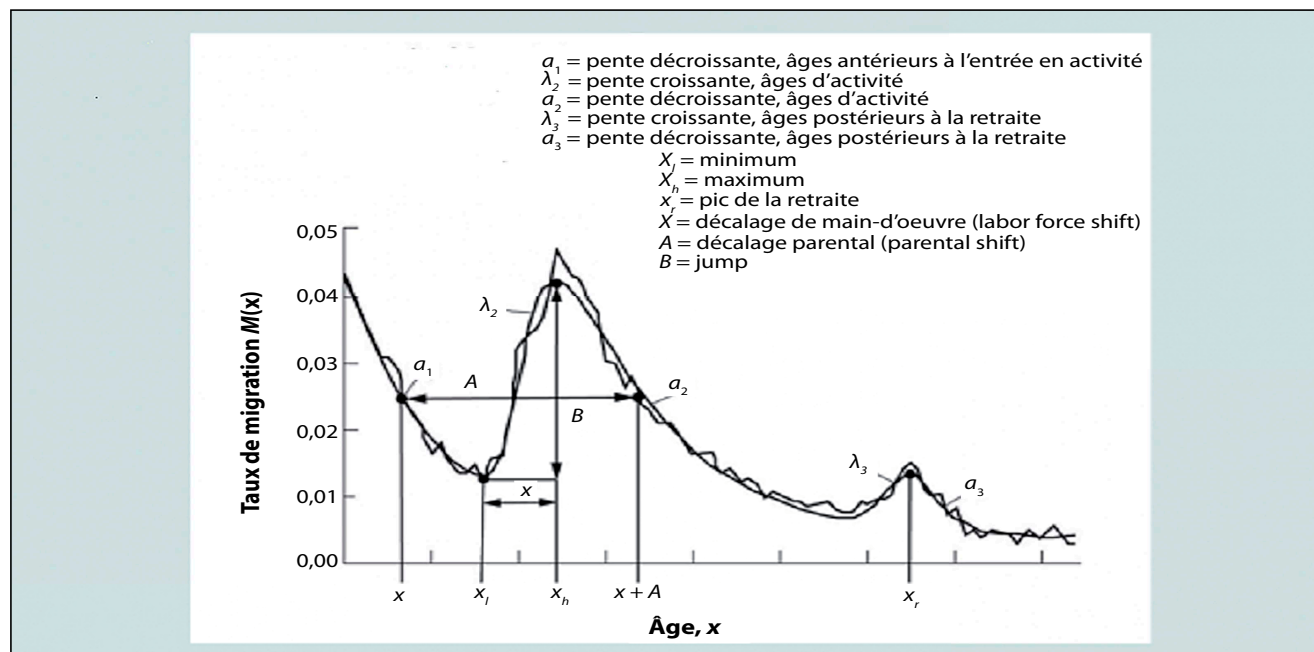
APPLICATION DE LA MÉTHODE

L'application de la méthode comporte les étapes suivantes.

Étape 1 : Préparer une série de taux observés

La première étape du calcul d'un modèle de migration consiste à préparer les données. Le choix des indices de migration à utiliser dépend des sources de données disponibles (registre, recensement ou enquête) et de l'objectif de la recherche. Par exemple, dans une étude comparative de

FIGURE 36.2 Le modèle migratoire à 11 paramètres de Rogers-Castro appliqué aux taux d'émigration interne des hommes vivant dans la région de Stockholm, 1974



Source : Rogers et Castro (1981). Reproduction autorisée par l'International Institute for Applied Systems Analysis (IIASA).

schémas migratoires, n'importe quel indice pourrait convenir pourvu qu'il ait été élaboré de la même manière dans les différents contextes à comparer. Par ailleurs, si les modèles doivent servir au calcul de projections de population par année d'âge, le schéma ajusté devrait représenter des taux de migration par année d'âge et année de calendrier. Mais si on ne dispose pas d'observations par année d'âge et année de calendrier qui évoluent de façon relativement lisse en fonction de l'âge, on doit avant tout convertir les données en indices par année d'âge et année de calendrier. Voici plusieurs situations couramment rencontrées.

A. DONNÉES DE RECENSEMENT, TAUX ANNUELS DE MIGRATION, PÉRIODE QUINQUENNALE D'OBSERVATION

Quand un recensement fournit les nombres de migrants survivants au terme d'une période rétrospective de cinq ans ainsi que l'année de la dernière migration, on peut calculer des taux de migration par année de calendrier et année d'âge au moyen d'une procédure, conceptuellement simple mais algébriquement complexe, de rétroprojection décrite par Dorrington et Moultrie (2009). Leur méthode intègre l'effet de la mortalité en appliquant aux migrants le régime de mortalité de la population générale, et l'effet de la migration interrégionale en appliquant les taux annuels de migration de la dernière année pour évaluer la population des régions un an avant le recensement, puis en utilisant ces estimations pour évaluer les taux de migration deux ans avant le recensement et en appliquant ceux-ci pour évaluer la population des régions deux ans avant le recensement, etc. Cela nécessite des données complémentaires sur la région de naissance des enfants de moins de 5 ans et des estimations annuelles de la population des régions par année d'âge. Les schémas de migration ainsi obtenus peuvent être ajustés et lissés par un modèle de Rogers-Castro et être utilisés dans des projections de population année par année.

B. INTERPOLATION DE PROBABILITÉS ANNUELLES À PARTIR DE PROBABILITÉS QUINQUENNALES

Quelle que soit la période d'observation des migrations, et que l'on utilise des données de recensement ou de registre, les probabilités de migration sont plus fiables quand elles sont évaluées par groupes quinquennaux que par année d'âge (Rogers, Little et Raymer 2010). En outre, les effectifs de migrants par tranches d'âge d'un an ne sont habituellement disponibles que dans les enquêtes par sondage, car les services statistiques nationaux publient généralement les

effectifs de migrants interrégionaux par groupes quinquennaux d'âge.

Pour appliquer le modèle multi-exponentiel quand les indices de migration disponibles se présentent par groupes quinquennaux d'âge, on a besoin d'un procédé de conversion des taux quinquennaux en taux annuels. L'interpolation par spline cubique (McNeil, Trussell et Turner 1977) est l'une de ces techniques, qui fournit une structure par année d'âge lisse. Rogers et Castro (1981) ont utilisé des taux suédois disponibles par année d'âge et par groupes quinquennaux pour tester la précision de cette méthode, et ils ont obtenu des résultats généralement satisfaisants.

Pour obtenir des courbes lisses de migration par année d'âge, les indices de migration observés par groupes quinquennaux d'âge sont attribués à l'âge central de chaque groupe, soit 2 ans, 7 ans, 12 ans, 17 ans, ..., 72 ans, 77 ans, etc. (ou 2,5 ans, 7,5 ans, 12,5 ans, ..., etc., si on travaille avec des taux plutôt qu'avec des probabilités). À partir de cette série de valeurs, on calcule une série continue de probabilités d'émigration internationale par âge au moyen d'une interpolation par spline cubique qui engendre des polynômes du troisième degré passant par la série des points de contrôle prédéfinis (appelés nœuds). On peut aussi utiliser des extensions, payantes ou libres, de Microsoft Excel, comme *XlXtrFun*, pour appliquer l'interpolation par spline cubique.

Une approche alternative consiste en une adaptation de la procédure d'interpolation à 6 paramètres de Beers (1945) pour interpoler les taux, des groupes les plus jeunes aux groupes les plus âgés, y compris l'extrapolation à 0 an et 1 an (ou 0,5 an et 1,5 an). Pour cette extrapolation, on suppose que l'écart entre les indices de migration à 1 et 2 ans est le même qu'entre 2 et 3 ans, et que l'écart entre les indices à 0 et 1 an est le même qu'entre 3 et 4 ans.

Donc, quelle que soit l'approche retenue, on a besoin d'une série de taux de migration par groupes quinquennaux d'âge, de 0-4 ans à 65-69 ans au moins.

Étape 2 : Choisir le type de modèle multi-exponentiel

Une fois que la série d'indices observés est prête, on doit choisir le type de modèle multi-exponentiel à adopter. Le système de modèles multi-exponentiels de migration présenté plus haut décrit les caractéristiques des modèles à 7, 9, 11 et 13 paramètres. Ce choix doit se fonder sur un examen visuel de la courbe, en gardant à l'esprit que le modèle est supposé représenter le profil réel de la courbe

de migration. Parfois, une fois tracée la représentation graphique, on ne voit pas clairement comment modéliser au mieux la partie correspondant à l'époque de la retraite et aux grands âges. Par exemple, il se peut qu'un modèle classique à 7 ou un modèle à 9 paramètres (migration en hausse aux âges élevés) convienne. Dans une telle situation, le choix en faveur du modèle à 9 paramètres pourrait se fonder sur la perspective théorique d'une hausse de la migration dans les dernières années de la vie. D'un autre côté, le modèle à 9 paramètres pourrait être écarté sur la base des indices de qualité d'ajustement, comme étant insuffisamment économe s'il ne donne pas un meilleur ajustement que le modèle à 7 paramètres. Pour bien choisir le type de modèle, il est recommandé de comparer la qualité d'ajustement d'un modèle simple à celle d'un modèle plus complexe (par exemple, comparer l'ajustement d'un modèle à 7 paramètres à celui d'un modèle à 11 paramètres). En règle générale, et sans perdre de vue la robustesse présumée des données de base, il faut une amélioration sensible de l'ajustement pour justifier le choix d'un modèle plus complexe.

Dans la plupart des pays en développement, en particulier quand l'âge de la « retraite » n'est pas concentré entre 60 et 65 ans et qu'il existe une tendance à exagérer les âges élevés, les données ne sont sans doute pas assez sûres pour s'adapter à un modèle plus complexe que la version à 7 paramètres.

Étape 3 : Ajustement du modèle au moyen du Solveur

Étant donné le nombre de paramètres (entre 7 et 13) des modèles multi-exponentiels de migration, il n'est pas conseillé de déterminer l'ajustement optimal *ab initio*, par essais et erreurs. Il vaut mieux utiliser des algorithmes analytiques. La procédure que nous présentons ci-dessous utilise un algorithme, *Solver*, présent dans le logiciel Microsoft Excel. *Solver* n'est pas toujours fourni dans les configurations habituelles de Microsoft Excel. Pour l'activer, il faut cliquer sur « Fichier → Options → Compléments → Gérer Compléments Excel → Atteindre ... » et s'assurer que la case « Complément solveur » est cochée.

Les caractéristiques de la fonction *Solver* et les conditions et contraintes qu'il faut accepter pour l'employer ont été sélectionnées dans le fichier Excel disponible en ligne. Pour appliquer le programme à une feuille de calcul, cliquer sur « Données → Solveur → Résoudre ».

Le modèle ajusté dans le fichier Excel disponible en ligne est organisé de manière à permettre à l'utilisateur de décider

si la « cible » à minimiser sera la somme des carrés des écarts entre taux observés et taux ajustés ou le chi-carré.

Par défaut, *Solver* est prévu pour utiliser tous les paramètres. Si on souhaite ajuster une courbe avec seulement quelques-uns des paramètres, on doit le spécifier dans la fenêtre « Cellules variables » et attribuer aux autres paramètres des valeurs fixes pertinentes (qui peuvent être ou non égales à zéro, selon les exigences de la procédure d'ajustement). Nous présentons ci-dessous un exemple dans lequel une telle optimisation sous contrainte peut être nécessaire.

La somme des carrés des écarts se calcule ainsi :

$$\frac{1}{n} \sum_{i=1}^n (O_i - F_i)^2$$

où O_i représente le taux observé à l'âge i , F_i la valeur ajustée à l'âge i et n le nombre de groupes d'âge.

Le chi-carré se calcule ainsi :

$$\frac{1}{n} \sum_{i=1}^n \frac{(O_i - F_i)^2}{F_i}.$$

Le chi-carré est plus sensible aux mauvais ajustements aux âges où les taux sont faibles (ce qui engendre une erreur proportionnellement plus importante), et il constitue donc un meilleur indicateur de la qualité d'ajustement quand on tente d'ajuster la « bosse de la retraite » (la troisième composante).

LE CHOIX DES VALEURS INITIALES

DE LA PROCÉDURE D'AJUSTEMENT

Le choix des valeurs de départ est la principale difficulté dans l'estimation de paramètres non-linéaires. Dans l'idéal, à partir d'un ensemble de valeurs initiales, l'algorithme procède par itérations et débouche sur une série corrigée de valeurs « optimales ». Mais cet optimum peut n'être qu'un optimum local et non général. Un meilleur choix des valeurs des paramètres de départ peut entraîner une qualité d'ajustement améliorée et aboutir à une série différente de valeurs finales. Un choix moins pertinent de valeurs initiales peut empêcher toute convergence, même vers un optimum local.

Gardant ceci à l'esprit, la méthode la plus efficace pour s'assurer que les résultats d'une procédure d'ajustement seront réellement « optimaux » est de choisir des valeurs des paramètres utilisées antérieurement pour une courbe « similaire ». À cet effet, on peut commencer avec les valeurs déjà intro-

duites dans le fichier Excel disponible en ligne, qui ont servi à ajuster les courbes dans les exemples développés ci-dessous.

La convergence peut être plus difficile à atteindre avec les modèles à 11 et 13 paramètres. Quand le recours à ces modèles plus lourds est justifié, on peut essayer d'ajuster d'abord aux données un modèle classique à 7 paramètres (ce qui assure déjà l'ajustement pour le pic de la courbe et pour les âges de la première moitié de la vie). Puis on peut figer ces 7 paramètres à la valeur obtenue lors de cette première étape (et donc les traiter à partir de là comme des constantes) et ensuite évaluer les paramètres restants. Une autre procédure efficace consiste à appliquer d'abord une méthode d'estimation linéaire, qui ne comporte pas d'algorithme itératif. Cette méthode a été décrite pour la première fois par Rogers et Castro (1981) et incluse ensuite dans la série de techniques alternatives présentée par Rogers, Castro et Lea (2005).

La recherche d'une solution optimale comporte un autre défi : le choix d'un critère d'arrêt de l'algorithme itératif. Quand la procédure itérative converge vers une solution, le chi-carré, qui mesure les écarts entre valeurs observées et valeurs estimées, diminue. On a atteint une solution acceptable quand le chi-carré ne diminue plus que d'une quantité négligeable entre une itération et la suivante. La valeur de cette faible variation, appelée « seuil de tolérance », est définie par l'utilisateur. On peut être tenté de la fixer à un niveau très bas, donc très proche de zéro, pour aller jusqu'à une valeur véritablement minimale du chi-carré. Mais le risque de cette attitude est qu'un seuil de tolérance aussi bas ne soit jamais atteint, même quand une solution a été obtenue. Press, Flannery, Teukolsky *et al.* (1986) proposent 0,001 comme valeur raisonnable du seuil de tolérance. Si le programme ne converge pas, on peut alléger les critères de convergence, en relevant le seuil de tolérance ou en essayant une nouvelle série de valeurs de départ.

Pour le choix des valeurs de départ, une autre méthode, procédant par essais et erreurs, s'appuie sur les graphiques du fichier Excel associé à ce chapitre (en ligne). En introduisant votre propre série de données observées dans l'une des feuilles de calcul, vous pouvez formuler des « hypothèses initiales » sur les valeurs des paramètres et les introduire dans les cellules destinées à leurs valeurs finales. Ensuite, après examen visuel de l'ajustement obtenu et identification des paramètres les moins satisfaisants, relancez la procédure avec de nouvelles valeurs de ceux-ci. Continuez ainsi jusqu'à ce que la courbe ajustée soit raisonnablement proche de la

courbe observée. Vous aurez alors en main des valeurs initiales acceptables à introduire dans la procédure d'estimation non-linéaire par les moindres carrés.

Étape 4 : Évaluer la qualité d'ajustement du modèle

On évalue la qualité d'ajustement du modèle en calculant l'écart absolu moyen en pourcentage (*EAMP*; en anglais : *mean absolute percent error*) :

$$EAMP = 100 \frac{1}{n} \sum_{i=1}^n \left[\frac{|F_i - O_i|}{O_i} \right].$$

L'*EAMP* a tendance à exagérer les erreurs, surtout quand les données observées comportent de nombreuses valeurs proches de zéro (Morrison, Bryan et Swanson 2004).

En plus de l'*EAMP*, on calcule également le R^2 , le carré du coefficient de corrélation entre les valeurs des O_i et celles des F_i . À titre heuristique, on considère couramment qu'un ajustement raisonnable est atteint avec un *EAMP* inférieur ou égal à 15 % et un R^2 sensiblement supérieur à 90 %.

Par ailleurs, puisque la méthode suppose que le modèle de Rogers-Castro calculé représente l'allure réelle de la courbe de migration, il devrait représenter la structure sous-jacente des données observées.

Étape 5 : Interprétation des résultats de l'ajustement

Si l'objectif est de décrire le schéma de migration et si on a pu ajuster un modèle multi-exponentiel aux données, on peut utiliser n'importe quel indice résumé (par exemple le taux brut de migration, X , B ou A), ainsi que les estimations des paramètres, pour caractériser la série. Les indices résumés et les interprétations des paramètres sont donnés dans la présentation générale, plus haut dans ce chapitre.

EXEMPLES

Dans les exemples développés ci-dessous, on applique des modèles multi-exponentiels de migration à diverses séries de données, de qualité et de complexité variable, provenant de différentes sources. Tous ces exemples sont fournis dans le fichier excel associé à ce chapitre et disponible en ligne.

C'est parce qu'on doit recourir à des méthodes itératives pour ajuster une table-type de mortalité aux probabilités conditionnelles de survie aux âges adultes que les exemples ne sont pas développés en détail dans le texte. Nous renvoyons le lecteur à l'explication, fournie dans la section précédente, de

l'utilisation de *Solver*, dans Microsoft Excel, pour déterminer les ajustements optimaux. Le fichier Excel est prêt à employer *Solver* pour le calcul des résultats présentés.

Données de recensement, migrations sur une période d'un an

La figure 36.3 présente un exemple basé sur des probabilités de migration par année d'âge sur une période d'un an, à partir des données d'un recensement. Les données proviennent de l'*American Community Survey* (ACS) de 2005, une enquête nationale réalisée chaque année par l'US Census Bureau. Même dans le cas de la Californie, un État fortement peuplé, les probabilités de migration par année d'âge sur une période d'un an sont très instables. L'*EA*MP est égal à 17 % et le R^2 à 0,92.

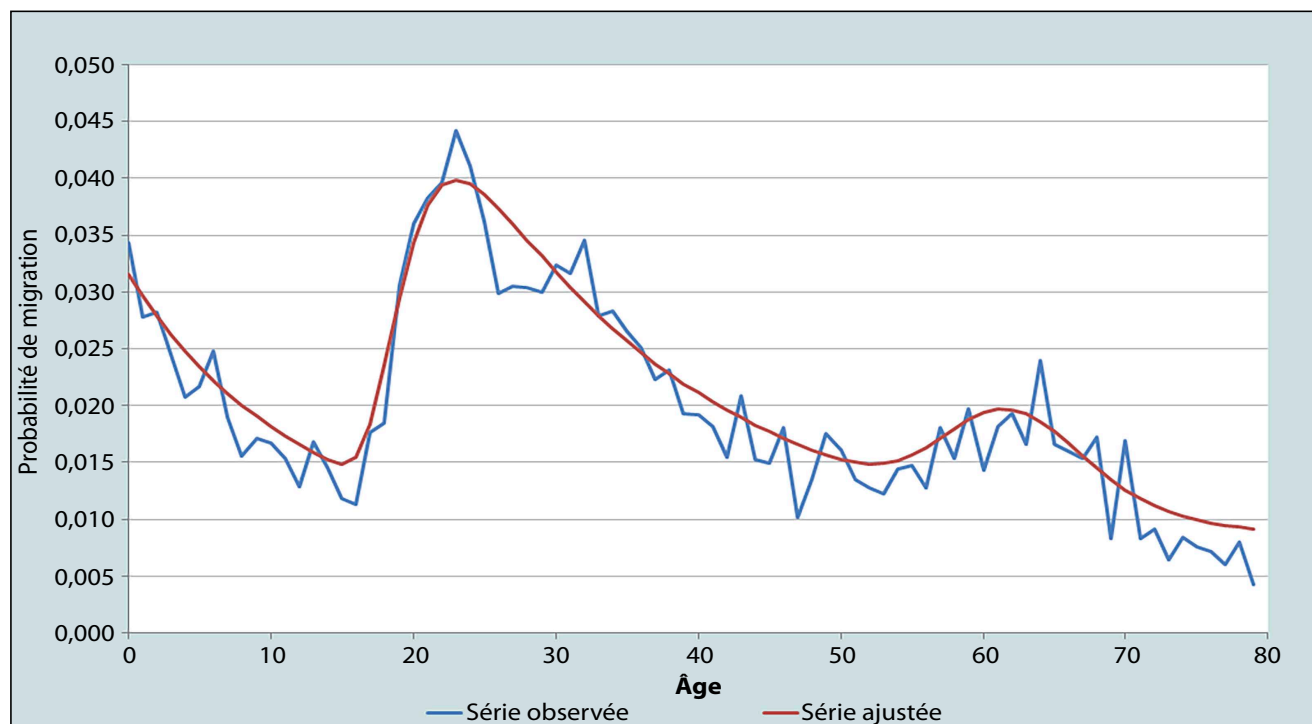
Il faut être prudent quand on utilise des probabilités de migration par année d'âge sur des périodes d'un an. Pour chaque année d'âge, les effectifs soumis au risque de migration et les effectifs de migrants peuvent être faibles, et cela engendre des probabilités irrégulières et instables. Il peut

être préférable de calculer des probabilités par groupes quinquennaux d'âge, qui s'avèrent plus fiables que les probabilités par année d'âge (Rogers, Little et Raymer 2010), et calculer ensuite par interpolation les probabilités par année d'âge, soit par spline cubique, soit par la formule de Beer, comme on l'a vu dans la section consacrée à l'application de la méthode.

Données de recensement, migrations sur une période de cinq ans

L'exemple de la figure 36.4 utilise des données de recensement relatives à l'État du New Hampshire. L'échantillon public de données individuelles (PUMS) à 1 % du US Bureau of Census est un échantillon relativement petit extrait du recensement, et le New Hampshire est l'un des États les moins peuplés. Les probabilités de migration par année d'âge sont très instables, leurs fluctuations sont énormes, mais le modèle donne une estimation lissée du véritable profil de la courbe. L'*EA*MP est égal à 52 % et R^2 à 0,68.

FIGURE 36.3 Émigration interne sur une période d'un an, de 2004 à 2005, et ajustement du modèle à 11 paramètres, Californie, ACS 2005



La figure 36.5 présente l'application de la méthode d'interpolation des splines cubiques aux probabilités de migration par groupes quinquennaux d'âge du New Hampshire, calculées à partir des données de l'échantillon public à 1 % du recensement de 2000. La courbe des valeurs interpolées à partir des taux quinquennaux est nettement plus lisse et donne des estimations plus fiables que la courbe des valeurs observées par année d'âge de la figure 36.4, elle constitue donc une meilleure série d'estimations à comparer avec la courbe ajustée multi-exponentielle. L'*EAMP*, qui valait 52 % avec les probabilités par année d'âge, tombe à 15 % avec les valeurs interpolées à partir des probabilités par groupes quinquennaux d'âge, et R^2 passe de 0,68 à 0,94.

Plusieurs facteurs expliquent pourquoi les niveaux des probabilités de migration sont nettement plus élevés au New Hampshire (figure 36.5) qu'en Californie (figure 36.3). L'exemple californien présente les migrations sur une période d'un an et l'exemple du New Hampshire porte sur une période de cinq ans. De plus, le territoire du New Hampshire est beaucoup plus petit que celui de la Cali-

fornie, et on s'attend généralement à ce que la migration soit d'autant plus intense que la zone géographique est plus petite.

DIAGNOSTICS, ANALYSE ET INTERPRÉTATION

Contrôles et validation

Il est important de vérifier par examen visuel si les taux de migration par âge ont un profil compatible avec les modèles de Rogers-Castro. Si ce n'est pas le cas, il est probable que ces modèles ne donneront pas un ajustement satisfaisant. De même, il vaut la peine de contrôler l'existence de valeurs extrêmes, en particulier pour les âges élevés, ce qui pourrait fausser le choix des paramètres ou même le choix du nombre de paramètres à prendre en compte pour l'ajustement. Si les données observées sont particulièrement erratiques, il est bon de les regrouper par classes quinquennales d'âge et de calculer ensuite une distribution lissée en utilisant soit l'interpolation à 6 paramètres de Beer soit une courbe spline.

Figure 36.4 Émigration interne par âge sur une période de 5 ans, de 1995 à 2000, et ajustement d'un modèle à 7 paramètres, New Hampshire, échantillon PUMS à 1 % du recensement de 2000

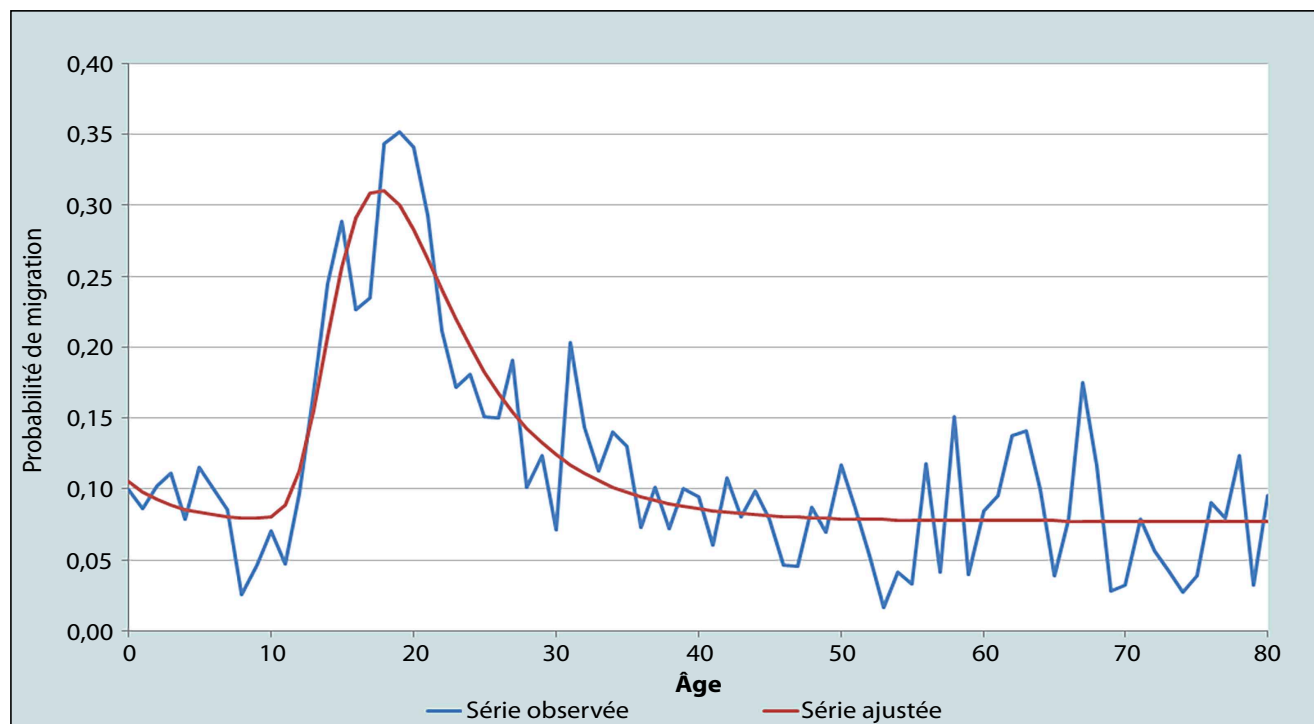
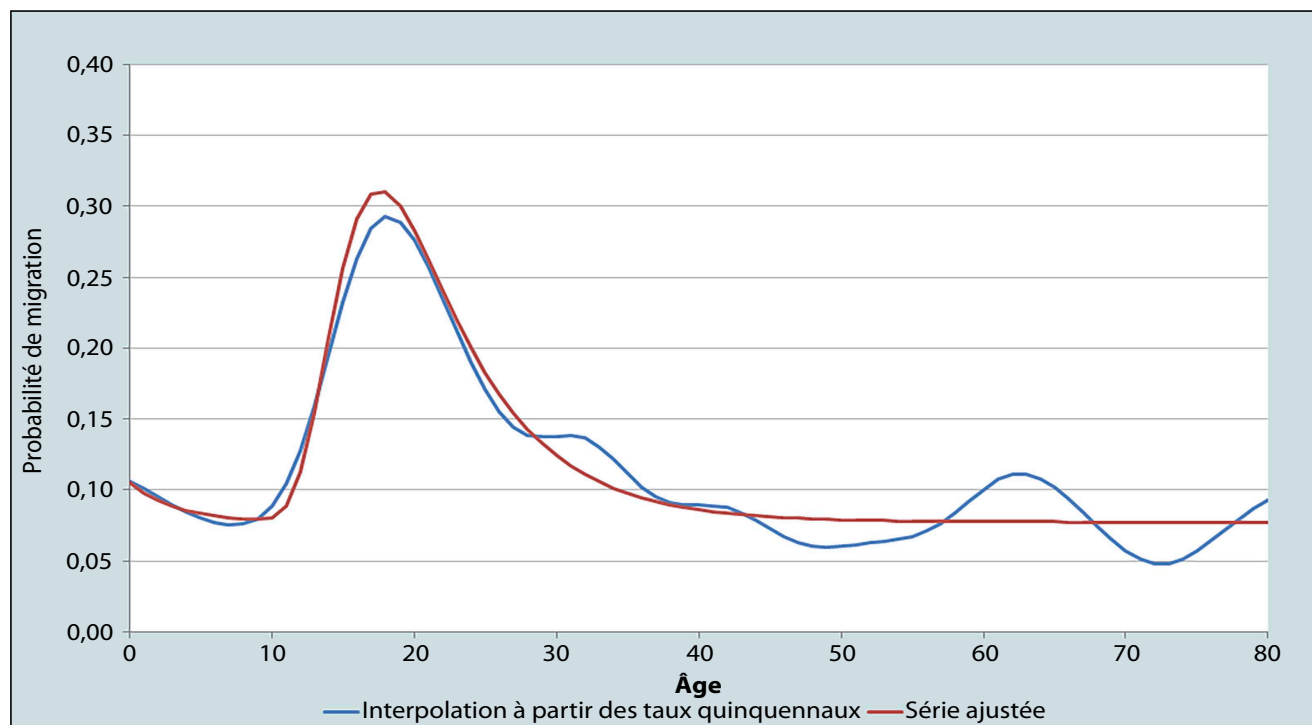


FIGURE 36.5 Émigration interne par âge sur une période de 5 ans, de 1995 à 2000, et ajustement d'un modèle à 7 paramètres, New Hampshire, échantillon PUMS à 1 % du recensement de 2000 - interpolations annuelles*



* Interpolations annuelles à partir des taux de migration par groupes quinquennaux d'âge.

REMARQUES SUR L'APPLICATION DE LA MÉTHODE

La théorie du modèle multi-exponentiel a été présentée plus haut dans ce chapitre, nous ne la répéterons pas ici. Dans cette section, nous développerons en détail certains éléments à examiner attentivement avant toute mise en application de la méthode.

Préparation des données

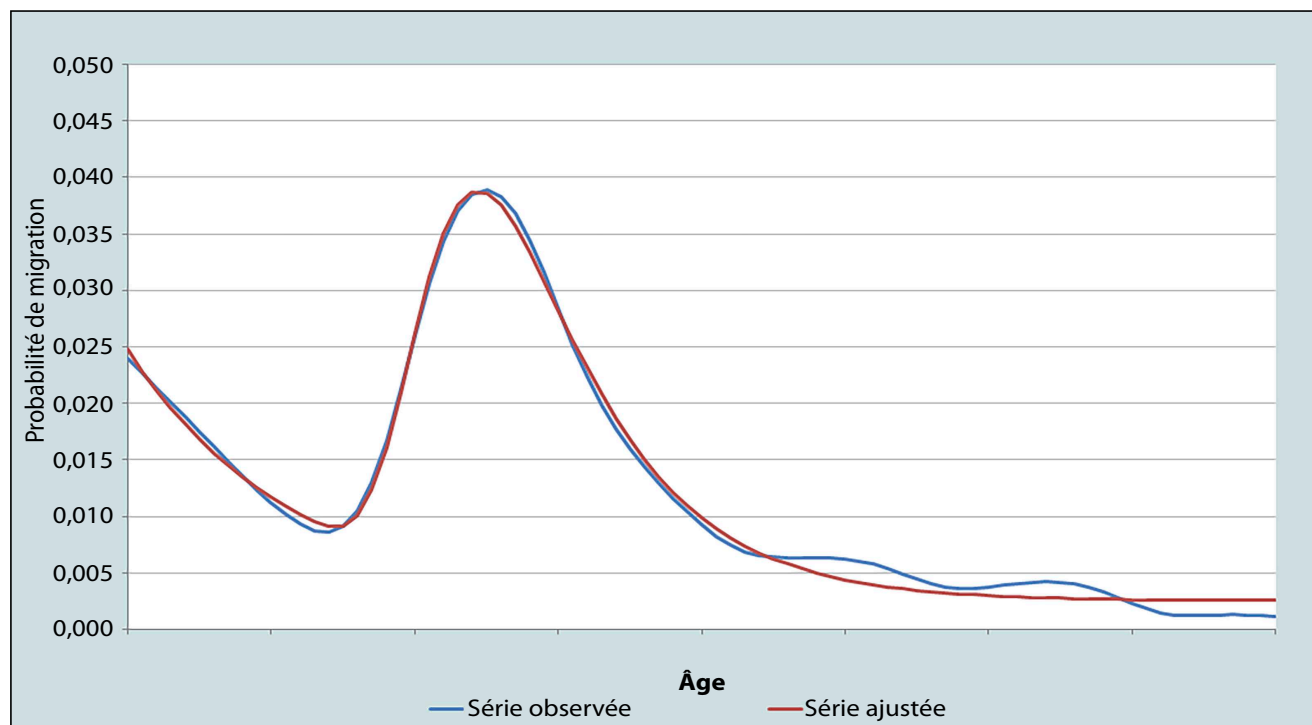
Le modèle multi-exponentiel s'applique à des séries de taux de migration par année d'âge, généralement de 0 an jusqu'à 65 ans, voire au-delà de 65 ans pour couvrir pleinement le schéma de migration des personnes âgées. Les données de migration par âge peuvent concerner un flux unidirectionnel (c'est-à-dire d'une région i vers une région j) ou l'émigration interne totale (c'est-à-dire d'une région i vers toutes les autres régions), ou l'ensemble des migrations interrégionales sans distinction d'origine ou de destination. Habi-

tuellement, les données sur les migrations proviennent des recensements nationaux (ou, dans les pays développés, des registres de population). Le modèle multi-exponentiel peut s'appliquer à divers indices de migration par année d'âge calculés à partir de l'une quelconque de ces sources.

Quand il est calculé à partir des données d'un registre national, le taux de migration des personnes qui avaient l'âge x au début d'une période déterminée est le rapport du nombre de migrations observées au cours de cette période au nombre moyen de personnes-années exposées au risque de migration. Certaines personnes peuvent migrer plus d'une fois en cours de période. Ce sont des taux du type événements/exposition au risque, bien que les migrations de personnes qui n'ont pas survécu jusqu'à la fin de la période puissent ne pas figurer au numérateur (Rogers et Castro 1981).

Les données de la courbe de migration de la figure 36.2, sur l'émigration masculine de Stockholm sur une période

FIGURE 36.6 Migrations intercommunales masculines sur la période de 5 ans de 1968 à 1973 en Suède, et ajustement d'un modèle à 11 paramètres



d'un an, provenaient du registre national suédois. Par contre, la figure 36.6 présente les séries observée et estimée par le modèle de l'ensemble des migrations masculines intercommunales suédoises sur une période de cinq ans. Comme il fallait s'y attendre, les niveaux sont beaucoup plus élevés sur la figure 36.6, du fait que les migrations sont plus intenses quand toutes les régions sont envisagées que quand il s'agit seulement de la région de Stockholm. On s'attend également à observer davantage de migrations sur une période de cinq ans que sur une seule année. Rees (1977) a constaté que les probabilités de migration sur cinq ans sont généralement moins que cinq fois (entre trois et cinq fois) supérieures à la probabilité sur un an. La courbe des taux observés est aussi plus lisse et plus proche de celle du modèle sur la figure 36.6, ce qui signifie que les taux par année d'âge sont plus fiables quand ils sont mesurés sur une longue période.

D'un autre côté, les recensements dénombrent les migrants survivants (et non les migrations). Les migrants sont les personnes qui déclarent résider dans une région au début de la période rétrospective considérée et dans une autre région au moment du recensement. Une personne qui a fait plusieurs migrations, consignées dans le registre national, peut être considérée comme non-migrante lors du recensement si elle est revenue à son point de départ. En général, les effectifs de migrants dénombrés par un recensement sont inférieurs aux nombres de migrations, surtout quand la période d'observation est longue et permet donc qu'il y ait de nombreux retours et de nombreux migrants décédés. C'est pourquoi une série de mesures de la migration fondée sur le registre de population n'est pas directement comparable à une série basée sur un recensement (Rogers et Castro 1986).

Habituellement, les recensements enregistrent la résidence actuelle de chaque individu et lui demandent où il

résidait un an ou cinq ans auparavant. Avec ces données et l'âge des individus au moment du recensement, on établit les nombres de migrants survivants et de survivants qui étaient exposés au risque de migration. Le rapport du nombre de migrants survivants au nombre de survivants exposés au risque de migrer est parfois qualifié de « conditionnel », parce que tant les migrants que les personnes exposées au risque de migration doivent avoir survécu jusqu'à la fin de la période rétrospective considérée pour être comptabilisés par le recensement (Rogers, Little et Raymer 2010). Comme ce n'est pas un taux du type événements/exposition au risque, nous parlerons plutôt ici de probabilité de migrer.

Données de recensement, migrations sur une période d'un an

Pour calculer des probabilités de migration par année d'âge quand le recensement demande à chaque personne où elle vivait un an auparavant, toutes les personnes sont « renvoyées » à la région où elles habitaient un an plus tôt quand elles avaient un an de moins, ce qui nous donne le nombre de personnes soumises au risque de migrer au départ de cette région. Par exemple, un enfant d'1 an au recensement de 2010 avait 0 an en 2009. Si les âges en 2010 vont de 1 à 85 ans, ceux de 2009 vont de 0 à 84 ans. (Note : seules les personnes âgées d'1 an et plus auront déclaré un lieu de résidence un an auparavant.) Revenir un an en arrière donne le nombre de personnes qui, ayant survécu jusqu'au recensement de 2010, étaient exposées au risque d'émigrer de la région *i* en 2009. Le nombre d'émigrants est celui des personnes qui ont déclaré habiter la région *i* en 2009 et ont été recensées dans une autre région en 2010. Pour chaque année d'âge, le rapport du nombre de migrants au nombre de personnes exposées au risque de migrer donne la probabilité d'émigration interne par âge sur l'année précédant le recensement. Quand le numérateur est un nombre de migrants unidirectionnels, c'est-à-dire de personnes qui ont migré de la région *i* vers la région *j*, ce rapport constitue la probabilité de migrer de la région *i* vers la région *j*, par âge.

Il faut être prudent quand on utilise des probabilités de migration par année d'âge sur une période d'un an. Pour chaque année d'âge, les effectifs soumis au risque de migration et les effectifs de migrants peuvent être faibles, et cela engendre des probabilités irrégulières et instables. Il peut être préférable de calculer des probabilités par groupes quinquennaux d'âge, qui s'avèrent plus fiables que les probabilités par année d'âge (Rogers, Little et Raymer 2010), et calculer ensuite par interpolation les probabilités par année d'âge.

Données de recensement, migrations sur une période de cinq ans

Si le recensement demande à chaque individu où il habitait cinq ans auparavant, on peut calculer des probabilités de migrer par année d'âge sur la période de cinq ans tant que l'âge de chacun est connu. Pour ce faire, on « renvoie » tous les individus à la région où ils vivaient cinq ans avant, quand ils avaient cinq ans de moins. Les enfants âgés de 5 ans lors d'un recensement effectué en 2000, par exemple, avaient 0 an en 1995. Si les âges en 2000 vont de 5 à 85 ans, ceux de 1995 vont de 0 à 80 ans. Le nombre d'émigrants est celui des personnes qui ont déclaré habiter la région *i* en 1995 et ont été recensées dans une autre région en 2000. Pour chaque année d'âge, le rapport du nombre de migrants au nombre de personnes exposées au risque de migrer donne la probabilité d'émigration interne par âge sur les cinq années précédant le recensement.

Données de recensement, taux annuels de migration sur une période de cinq ans

Quand le recensement donne les nombres de migrants qui ont survécu jusqu'au terme d'une période de cinq ans, on peut calculer des taux de migration par année d'âge et année de calendrier au moyen d'une procédure de rétroprojection décrite par Dorrington et Moultrie (2009). Cette méthode intègre l'effet de la mortalité en appliquant aux migrants le régime de mortalité de la population générale, et l'effet des migrations répétées en appliquant les taux annuels de migration de la dernière année pour évaluer la population des régions un an avant le recensement, puis en utilisant ces estimations pour évaluer les taux de migration deux ans avant le recensement et en appliquant ceux-ci pour évaluer la population des régions deux ans avant le recensement, etc. Cela nécessite des données complémentaires sur la région de naissance des enfants de moins de 5 ans et des estimations annuelles de la population des régions par année d'âge. Les schémas de migration ainsi obtenus peuvent être ajustés et lissés par un modèle de Rogers-Castro et être utilisés dans des projections de population année par année.

Limites

À moins que l'on ait des données exactes ne présentant aucune irrégularité, le modèle multi-exponentiel ne fournit pas d'ajustement très précis et peut donc être surparamétré, en ce sens que plusieurs séries différentes de paramètres peuvent s'ajuster aussi bien les unes que les autres aux don-

nées observées. En pareil cas, il pourrait être utile de donner une valeur constante à un ou deux paramètres et de procéder à l'ajustement avec les autres ; nous recommandons donc une certaine modération quant au nombre de paramètres.

Extensions de la méthode

L'application du modèle multi-exponentiel ne se limite pas aux séries de taux ou probabilités de migration. Plusieurs études ont montré que les structures par âge de migrants (et de migrations si on dispose de données de registre) ont souvent une forme multi-exponentielle et peuvent être fidèlement représentées par un modèle de Rogers-Castro (Little et Rogers 2007 ; Rogers, Little et Raymer 2010).

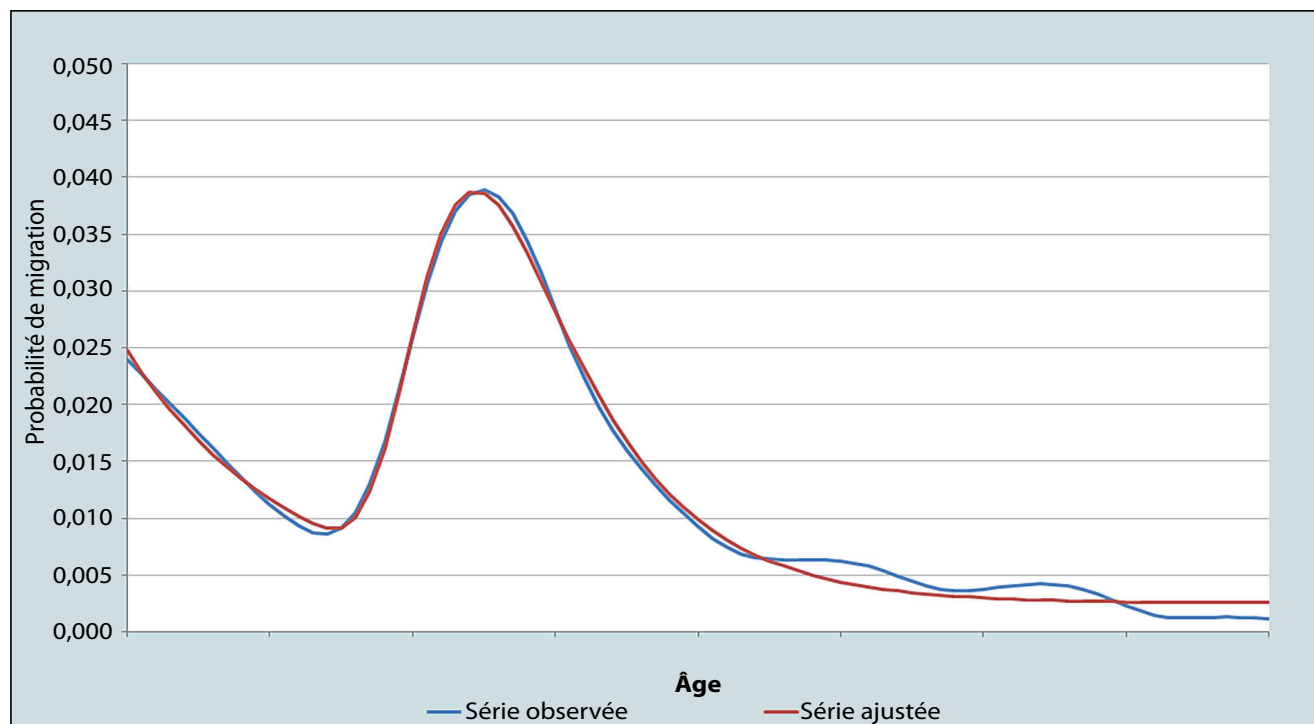
On peut déterminer les nombres de migrants ou de migrations par année d'âge en utilisant l'une ou l'autre des sources de données et des méthodes présentées plus haut, car ce sont simplement les numérateurs des taux ou probabilités de migration. Les données observées ajustées par les modèles sont les rapports des nombres de migrants ou de migrations par année d'âge à leurs totaux respectifs. (Note : si les effectifs de migrants sont distribués par classes quinquennales

d'âge, l'un ou l'autre type d'interpolation sera nécessaire ; si on utilise l'interpolation par spline cubique, l'effectif associé à chaque centre de classe doit être l'effectif de la classe quinquennale divisé par cinq.)

Par exemple, la figure 36.7 présente la distribution de fréquences des migrations suédoises par âge. Elle se révèle très lisse et fiable, à l'exception des âges les plus élevés. Un modèle à 7 paramètres donne un très bon ajustement, avec un R^2 de 0,99 et un *EAMP* de 29 %. Mais cet exemple montre comment l'*EAMP* peut exagérer l'imperfection du modèle, car il « gonfle » excessivement en présence d'une série de petites variations dans les données observées.

Il existe deux autres logiciels qui permettent d'ajuster la courbe multi-exponentielle et offrent une alternative au fichier Excel disponible en ligne sur le site associé à cet ouvrage. Ce sont : (1) *Data Master 2003*, un programme gratuit d'ajustement de courbes qui applique l'algorithme de Levenberg-Marquardt ; et (2) *R* (R Development Core Team 2012), un environnement logiciel, gratuit également, destiné aux calculs statistiques et représentations graphiques en tous genres, et qui demande donc un temps d'apprentis-

FIGURE 36.7 Structure par âge des migrations intercommunales en Suède, de 1968 à 1973, et ajustement d'un modèle à 7 paramètres



sage assez long avant de pouvoir être utilisé en toute sécurité. L'Appendice à ce chapitre, disponible en ligne sur le site web associé à cet ouvrage fournit des instructions élémentaires

pour la définition des fonctions de R nécessaires au calcul des modèles à 7 et 11 paramètres en utilisant l'algorithme de Gauss-Newton.

RÉFÉRENCES

- Bates J and I Bracken. 1982. "Estimation of migration profiles in England and Wales", *Environment and Planning A* **14**(7):889-900. doi: <http://dx.doi.org/10.1068/a140889>
- Bates J and I Bracken. 1987. "Migration age profiles for local-authority areas in England, 1971-1981", *Environment and Planning A* **19**(4):521-535. doi: <http://dx.doi.org/10.1068/a190521>
- Beers H. 1945. "Six-term formulas for routine actuarial interpolation", *The Record of the American Institute of Actuaries* **33**(2):245-260.
- Dorrington R and TA Moultrie. 2009. "Making use of the consistency of patterns to estimate age-specific rates of interprovincial migration in South Africa," Paper presented at Annual Meeting of the Population Association of America. Detroit, Michigan, 29 April - 2 May 2009.
- George MV. 1994. *Population projections for Canada, provinces and territories, 1993-2016*. Ottawa: Statistics Canada, Demography Division, Population Projections Section.
- Hofmeyr BE. 1988. "Application of a mathematical model to South African migration data, 1975-1980", *Southern African Journal of Demography* **2**(1):24-28.
- Kawabe H. 1990. *Migration rates by age group and migration patterns: Application of Rogers' migration schedule model to Japan, The Republic of Korea, and Thailand*. Tokyo: Institute of Developing Economies.
- Liaw K-L and DN Nagnur. 1985. "Characterization of metropolitan and nonmetropolitan outmigration schedules of the Canadian population system, 1971-1976", *Canadian Studies in Population* **12**(1):81-102.
- Little JS and A Rogers. 2007. "What can the age composition of a population tell us about the age composition of its out-migrants?", *Population, Space and Place* **13**(1):23-19. doi: <http://dx.doi.org/10.1002/psp.440>
- McNeil DR, TJ Trussell and JC Turner. 1977. "Spline interpolation of demographic data", *Demography* **14**(2):245-252. doi: <http://dx.doi.org/10.2307/2060581>
- Morrison PA, TM Bryan and DA Swanson. 2004. "Internal migration and short-distance mobility," in Siegel, JS and DA Swanson (eds). *The Methods and Materials of Demography*. San Diego: Elsevier pp. 493-521.
- Potrykowska A. 1988. "Age patterns and model migration schedules in Poland", *Geographia Polonica* **54**:63-80.
- Press WH, BP Flannery, SA Teukolsky and WT Vetterling. 1986. *Numerical Recipes: The Art of Scientific Computing*. Cambridge: Cambridge University Press.
- R Development Core Team. 2012. *R: A language and environment for statistical computing: Reference Index*. Vienna, Austria: R Foundation for Statistical Computing. <http://www.mendeley.com/research/r-language-environment-statistical-computing-13/>
- Raymer J and A Rogers. 2008. "Applying model migration schedules to represent age-specific migration flows," in Raymer, J and F Willekens (eds). *International Migration in Europe: Data, Models and Estimates*. Chichester: Wiley, pp. 175-192.
- Rees PH. 1977. "The measurement of migration, from census data and other sources", *Environment and Planning A* **9**(3):247-272. doi: <http://dx.doi.org/10.1068/a090247>
- Rogers A and LJ Castro. 1981. *Model Migration Schedules*. Laxenburg, Austria: International Institute for Applied Systems Analysis. <http://webarchive.iiasa.ac.at/Admin/PUB/Documents/RR-81-030.pdf>
- Rogers A and LJ Castro. 1986. "Migration," in Rogers, A and F Willekens (eds). *Migration and Settlement: A Multiregional Comparative Study*. Dordrecht: D. Reidel, pp. 157-208.
- Rogers A, LJ Castro and M Lea. 2005. "Model migration schedules: Three alternative linear parameter estimation methods", *Mathematical Population Studies* **12**(1):17-38. doi: <http://dx.doi.org/10.1080/08898480590902145>
- Rogers A and JS Little. 1994. "Parameterizing age patterns of demographic rates with the multiexponential model schedule", *Mathematical Population Studies* **4**(3):175-195. doi: <http://dx.doi.org/10.1080/08898489409525372>
- Rogers A, JS Little and J Raymer. 2010. *The Indirect Estimation of Migration: Methods for Dealing with Irregular, Inadequate, and Missing Data*. Dordrecht: Springer.
- Rogers A and J Raymer. 1999. "Estimating the regional migration patterns of the foreign-born population in the United States: 1950-1990", *Mathematical Population Studies* **7**(3):181-216. doi: <http://dx.doi.org/10.1080/08898489909525457>
- Rogers A and J Watkins. 1987. "General versus elderly interstate migration and population redistribution in the United States", *Research on Aging* **9**(4):483-529. doi: <http://dx.doi.org/10.1177/0164027587094002>

Chapitre 37 LES MODÈLES LOG-LINÉAIRES DE FLUX MIGRATOIRES

Jani Little et James Raymer

INTRODUCTION AUX APPLICATIONS DES MODÈLES

Le système de modèles log-linéaires offre plusieurs techniques intéressantes pour l'étude et l'estimation des flux migratoires dans le cadre d'un ensemble de régions. Jusqu'à présent, ces méthodes ont été le plus souvent appliquées à l'étude des migrations internes dans des groupes de régions définies comme unités administratives infranationales. Mais elles ne se limitent pas aux migrations internes, et on peut les appliquer également aux migrations internationales (Raymer 2007).

On définit un flux migratoire comme le nombre de migrations d'une région vers une autre observées sur une période déterminée. Il existe diverses manières de comptabiliser les migrations, et elles peuvent donner des résultats différents. Par exemple, Rees et Willekens (1986) font une distinction entre les registres de population, qui dénombrent les changements de résidence interrégionaux sur une période donnée, et les recensements, qui comptent les individus qui, à la date du recensement, vivent en un lieu différent de l'endroit où ils résidaient au début de la période d'observation.

Quelle que soit la méthode de mesure des flux migratoires, il est de tradition de les présenter sous forme de tableaux de contingence. Ce sont des tableaux carrés où sont consignés les nombres de déplacements entre régions d'origine et régions de destination. Les flux d'une telle matrice de migrations peuvent être parfaitement reproduits par le modèle à composantes multiplicatives, qui est un modèle log-linéaire saturé (c'est-à-dire où il y a autant de paramètres estimés que de points de données). Il a été utilisé par Willekens (1983), Rogers, Willekens, Little *et al.* (2002) et Rogers, Little et Raymer (2010) pour représenter la matrice des flux entre régions, et par Raymer et Rogers (2007), Raymer, Bonaguidi et Valentini (2006) et Rogers, Little et Raymer (2010) pour modéliser la structure des flux interrégionaux par

groupes d'âge. Les composantes multiplicatives sont susceptibles d'interprétation et facilement utilisables pour décrire le régime des migrations entre les régions étudiées (Rogers, Willekens, Little *et al.* 2002). Si on les a calculées pour plusieurs séries de flux interrégionaux, relatives, par exemple, à des périodes différentes ou à des sous-populations différentes définies par l'âge, le sexe ou la race, les composantes multiplicatives peuvent servir à comparer les régimes de migration de ces diverses populations.

On peut utiliser les méthodes log-linéaires pour justifier des représentations simplifiées des régimes de migration, qui sont moins lourdes que le modèle saturé. La pertinence d'un modèle restreint est déterminée par l'adéquation entre les flux calculés et les flux observés et par le recours aux méthodes statistiques d'évaluation de la qualité d'ajustement. Si le modèle restreint convient, c'est-à-dire s'il s'ajuste bien aux données, on peut l'utiliser pour une estimation indirecte des flux. Le modèle d'indépendance, par exemple, repose sur l'hypothèse que les flux interrégionaux sont distribués conformément à la répartition qui aurait pu être prédite sur la base des distributions marginales des flux par régions d'origine et de destination. Si le modèle d'indépendance est validé, on peut prédire les flux interrégionaux et les estimer indirectement, mais avec précision, si les flux totaux d'entrée et de sortie de chacune des régions sont connus.

Parfois, on fait l'hypothèse que la structure des migrations est invariante par rapport aux facteurs tels que le temps, l'âge, le sexe et la race. Ces hypothèses peuvent être traduites et testées par des modèles log-linéaires. Portant sur des niveaux de migration différents, plusieurs recherches ont montré la remarquable stabilité des régimes de migration, en particulier les taux par âge, en fonction du temps (Mueser 1989 ; Nair 1985 ; Snickars et Weibull 1977). D'autres travaux ont montré la cohérence des structures par âge des migrations interrégionales au cours du temps (Raymer et Rogers 2007).

En outre, la courbe de migration par âge des groupes les plus jeunes, que l'on peut élaborer à partir des effectifs par lieu de naissance, s'est révélée, dans certains cas, être un « proxy » du niveau de migration et a permis d'évaluer la migration des groupes les plus âgés (Raymer et Rogers 2007 ; Rogers, Little et Raymer 2010).

Tous ces travaux ont préparé le terrain pour l'élaboration de la méthode des *offsets* en tant qu'excellent outil d'estimation indirecte des flux migratoires. C'est une application particulière de la modélisation log-linéaire, qui plaque une structure de migration connue sur une série de flux migratoires interrégionaux qui peut comporter des lacunes ou des valeurs douteuses. Avec cette méthode, le schéma de migration connu d'une certaine période peut être adapté de celui d'une autre période. De plus, quand les flux sont répartis par âge, on peut appliquer la structure des flux interrégionaux par âge d'une période à une autre période. Enfin, Raymer et Rogers (2007) ont montré que le niveau de la migration des petits enfants peut être utilisé, avec la méthode des *offsets*, pour évaluer indirectement les flux migratoires des personnes âgées.

Diverses applications des modèles log-linéaires, avec les hypothèses correspondantes, sont détaillées dans les sections qui suivent, en commençant par le cas à deux variables, c'est-à-dire l'analyse des flux par origine et destination. Dans cette section, ayant défini le modèle log-linéaire dans le cadre de matrices de flux à deux dimensions, nous calculerons et nous interpréterons les formes multiplicative et additive du modèle saturé. Le modèle log-linéaire d'indépendance et le modèle de quasi-indépendance « limité aux migrants » seront exposés, avec quelques illustrations et une brève description des méthodes d'évaluation de la qualité d'ajustement.

Cette section s'achèvera sur une illustration de l'emploi de la méthode des *offsets* pour estimer indirectement les flux interrégionaux d'une période à partir de la structure de ceux d'une autre période. Si on dispose des flux migratoires de deux périodes différentes, on peut tester l'hypothèse d'invariance par rapport au temps avec un modèle log-linéaire et la méthode des *offsets*. Nous examinerons les modèles qui répartissent les flux d'origine et de destination par groupes d'âge. Nous terminerons par une illustration de la manière dont le modèle multiplicatif prenant en compte l'âge peut, en association avec la méthode des *offsets*, servir à l'estimation indirecte des flux interrégionaux par âge d'une autre période.

APPLICATIONS DU MODÈLE

À DEUX VARIABLES

Pour illustrer le modèle log-linéaire à deux variables, nous considérons les migrations, observées en 1973 et en 1976, entre communes néerlandaises groupées en six catégories selon leur degré d'urbanisation. Les données, publiées par Willekens (1983), sont présentées dans le tableau 37.1. Dans cet exemple, il y a deux variables, la région d'origine (O) et la région de destination (D). Aucune d'elles n'est désignée comme la variable dépendante. Le résultat peut prendre la forme des flux migratoires interrégionaux, notés n_{ij} , dans la variante multiplicative du modèle, ou celle des logarithmes naturels des flux, notés $\ln(n_{ij})$, dans la variante additive.

Les décompositions du modèle saturé, reproduisant chacune parfaitement les données observées, sont décrites dans les sous-sections présentant le modèle à composantes multiplicatives et le modèle additif linéaire, et trois techniques d'estimation indirecte sont illustrées dans les trois sous-sections suivantes, consacrées au modèle d'indépendance, au modèle de quasi-indépendance et à la méthode des *offsets*.

Application 1 : Le modèle à composantes multiplicatives

La forme multiplicative du modèle log-linéaire saturé, dite modèle à composantes multiplicatives, reproduit les éléments de la matrice des flux de la manière suivante :

$$n_{ij} = (T)(O_i)(D_j)(OD_{ij}). \quad (1)$$

Comme tous les modèles saturés, ce n'est pas, à proprement parler, un modèle, mais une manière de représenter les données. n_{ij} est le flux observé de la région i vers la région j , et les paramètres mesurant l'effet des facteurs intervenants sont T , O_i , D_j et OD_{ij} . Ainsi, chaque flux de i vers j , occupant une cellule dans les matrices 6×6 du tableau 37.1 (hors marges), peut être représenté par une équation de la même forme que l'équation 1, avec la série de paramètres correspondante. T représente l'effet global de l'ensemble des facteurs intervenants, O_i l'effet de la région d'origine, D_j celui de la région de destination, et OD_{ij} l'effet de l'interaction entre O_i et D_j . Pris tous ensemble, les paramètres du modèle saturé représentent la structure spatiale de la migration (Rogers, Willekens, Little *et al.* 2002).

Deux séries différentes de paramètres satisfaisant le modèle à composantes multiplicatives ont été utilisées dans les recherches sur la migration. Nous les présentons toutes les deux ici. Chacune offre une manière différente de représenter et d'interpréter le régime de migration. On appelle

TABLEAU 37.1 Migrations entre communes selon le degré d'urbanisation*, Pays-Bas, 1973 et 1976

| A. Migrations 1973 | | | | | | | |
|--|---------|---------|---------|--|---------|---------|---------|
| Destination | | | | | | | |
| Origine | 1 | 2 | 3 | 4 | 5 | 6 | Total |
| 1 | 50 498 | 23 829 | 8 566 | 21 846 | 16 264 | 18 856 | 139 859 |
| 2 | 25 005 | 27 536 | 6 953 | 14 326 | 16 212 | 18 282 | 108 314 |
| 3 | 15 675 | 10 710 | 13 874 | 6 266 | 9 819 | 19 701 | 76 045 |
| 4 | 23 457 | 14 169 | 4 431 | 10 209 | 9 386 | 10 973 | 72 625 |
| 5 | 29 548 | 25 267 | 11 802 | 13 160 | 15 979 | 20 406 | 116 162 |
| 6 | 46 815 | 39 123 | 42 399 | 25 012 | 26 830 | 23 304 | 203 483 |
| TOTAL | 190 998 | 140 634 | 88 025 | 90 819 | 94 490 | 111 522 | 716 488 |
| B. Migrations 1976 | | | | | | | |
| Destination | | | | | | | |
| Origine | 1 | 2 | 3 | 4 | 5 | 6 | Total |
| 1 | 14 473 | 14 327 | 6 077 | 11 689 | 10 618 | 9 897 | 67 081 |
| 2 | 14 833 | 36 258 | 13 289 | 17 391 | 20 899 | 21 869 | 124 539 |
| 3 | 8 330 | 17 764 | 25 113 | 10 489 | 18 171 | 29 220 | 109 087 |
| 4 | 11 315 | 16 498 | 8 935 | 10 537 | 10 762 | 12 519 | 70 566 |
| 5 | 11 875 | 24 370 | 19 151 | 12 312 | 16 724 | 22 591 | 107 023 |
| 6 | 16 582 | 32 336 | 52 415 | 22 264 | 28 182 | 27 810 | 179 589 |
| TOTAL | 77 408 | 141 553 | 124 980 | 84 682 | 105 356 | 123 906 | 657 885 |
| *1 : communes purement rurales | | | | 4 : villes à caractère rural et petites villes | | | |
| 2 : communes rurales industrielles | | | | 5 : villes moyennes | | | |
| 3 : communes de résidence des navetteurs | | | | 6 : grandes villes (plus de 100 000 habitants) | | | |

Source : Bureau central des statistiques, La Haye.

la première « calcul des effets des facteurs sur la base de la moyenne géométrique » (*en anglais : geometric mean effect coding*) (Knoke et Burke 1980 ; Willekens 1983), et la seconde « calcul des effets des facteurs sur la base du total général » (*en anglais : total sum reference coding*) (Raymer et Rogers 2007 ; Rogers, Little et Raymer 2010). Un troisième modèle à composantes multiplicatives est développé dans la sous-section consacrée au modèle log-linéaire additif.

Application 2 : Calcul des effets des facteurs sur la base de la moyenne géométrique

Le calcul des effets des facteurs sur la base de la moyenne géométrique a été la première décomposition de l'équa-

tion 1 utilisée dans l'analyse des migrations. Proposé par Birch (1963), il est, par sa forme, équivalent au modèle de gravitation (Willekens 1983). Le tableau 37.2 présente les composantes multiplicatives résultant de l'application du calcul des effets des facteurs sur la base de la moyenne géométrique aux données néerlandaises du tableau 37.1. Notons que la composante globale (T) se trouve dans les totaux généraux du tableau, les composantes de la région d'origine (O_i) dans les totaux de lignes, les composantes de la région de destination (D_j) dans les totaux de colonnes, et les composantes de l'interaction origine-destination (OD_{ij}) dans les cellules intérieures des deux matrices (hors marges).

TABLEAU 37.2 Composantes multiplicatives obtenues par application du calcul des effets des facteurs sur la base de la moyenne géométrique

| A. Migrations 1973 | | | | | | | |
|--------------------|-------|-------------|-------|-------|-------|-------|------------|
| | | Destination | | | | | |
| Origine | 1 | 2 | 3 | 4 | 5 | 6 | Total |
| 1 | 1,457 | 0,940 | 0,656 | 1,352 | 0,933 | 0,882 | 1,180 |
| 2 | 0,885 | 1,332 | 0,653 | 1,087 | 1,140 | 1,048 | 0,962 |
| 3 | 0,771 | 0,720 | 1,811 | 0,661 | 0,959 | 1,570 | 0,692 |
| 4 | 1,275 | 1,052 | 0,639 | 1,190 | 1,014 | 0,966 | 0,627 |
| 5 | 0,943 | 1,102 | 1,000 | 0,901 | 1,013 | 1,055 | 1,067 |
| 6 | 0,838 | 0,957 | 2,015 | 0,960 | 0,954 | 0,676 | 1,903 |
| TOTAL | 1,711 | 1,252 | 0,644 | 0,798 | 0,861 | 1,056 | 17 168,003 |
| B. Migrations 1976 | | | | | | | |
| | | Destination | | | | | |
| Origine | 1 | 2 | 3 | 4 | 5 | 6 | Total |
| 1 | 1,753 | 0,984 | 0,571 | 1,317 | 0,979 | 0,787 | 0,656 |
| 2 | 0,986 | 1,366 | 0,686 | 1,075 | 1,057 | 0,954 | 1,195 |
| 3 | 0,655 | 0,792 | 1,533 | 0,767 | 1,088 | 1,508 | 1,010 |
| 4 | 1,277 | 1,055 | 0,783 | 1,106 | 0,925 | 0,927 | 0,704 |
| 5 | 0,900 | 1,047 | 1,127 | 0,868 | 0,965 | 1,124 | 1,048 |
| 6 | 0,769 | 0,850 | 1,888 | 0,960 | 0,995 | 0,847 | 1,712 |
| TOTAL | 0,768 | 1,354 | 0,989 | 0,825 | 1,008 | 1,169 | 16 401,919 |

L'effet global, T , est décrit comme la constante de proportionnalité ou l'effet principal de la *taille* (Willekens 1983). C'est la moyenne géométrique de tous les flux interrégionaux :

$$T = \left[\prod_{ij} n_{ij} \right]^{\frac{1}{m \times m}},$$

où m est le nombre de régions d'origine (lignes), égal au nombre de régions de destination (colonnes). T est égal à 17 168,003 en 1973 et à 16 401,919 en 1976.

Pour une région déterminée i , l'effet principal de cette région d'origine est le rapport de la moyenne géométrique des flux partant de i à la moyenne géométrique générale.

$$O_i = \frac{1}{T} \left[\prod_j n_{ij} \right]^{\frac{1}{m}}.$$

L'effet principal O_i mesure l'importance relative de la région i comme source de migrations (Alonso 1986). Par exemple,

selon les données de 1973, l'effet de région d'origine des communes de la 4^e catégorie est égal à :

$$O_4 = \frac{1}{17168,003} [23457 \times 14169 \times 4431 \times 10209 \times 9386 \times 10973]^{\frac{1}{6}} = 0,627.$$

C'est le plus faible des effets de région d'origine (lignes), ce qui signifie que la 4^e catégorie était la moins importante des sources de migrations en 1973.

De même, l'effet principal de la région de destination, D_j , mesure l'importance relative de la région j comme destination des migrants. C'est le rapport de la moyenne géométrique de la colonne j à la moyenne géométrique générale, et sa formule est :

$$D_j = \frac{1}{T} \left[\prod_i n_{ij} \right]^{\frac{1}{m}}.$$

Par exemple, pour les communes de la 4^e catégorie, l'effet de région de destination en 1973 est égal à :

$$D_4 = \frac{1}{17168,003} [21846 \times 14326 \times 6266 \times 10209 \times 13160 \times 25012]^{\frac{1}{6}} = 0,798.$$

Tous les autres effets, en ligne comme en colonne, se calculent de la même manière. Chaque effet est égal au rapport de la moyenne géométrique des éléments d'une ligne (ou d'une colonne) à la moyenne géométrique générale. Ils sont l'équivalent des facteurs d'équilibre du modèle de gravitation (Willekens 1983).

On peut comparer ces effets entre régions et entre périodes. Par exemple, la 6^e catégorie était la plus importante source de migrations en 1973 (1,903 est supérieur à tous les autres effets de région d'origine) et en 1976 (1,712 est supérieur à tous les autres effets de région d'origine). La 1^{re} catégorie était moins importante comme destination des migrations en 1976 qu'en 1973 (0,768 est inférieur à 1,711), et, en 1973, elle était moins importante comme origine que comme destination (1,180 est inférieur à 1,711).

Les cadres A et B du tableau 37.2 sont parfois appelés matrices d'interaction spatiale. Ils contiennent les effets d'interaction OD_{ij} de l'équation 1, et chaque effet est égal au rapport du flux observé entre i et j au flux calculé, qui est le produit des trois autres paramètres. La formule est :

$$OD_{ij} = \frac{n_{ij}}{(T)(O_i)(D_j)}.$$

Chaque OD_{ij} exprime l'écart entre le flux observé, n_{ij} , et le flux calculé sous l'hypothèse d'indépendance entre région de destination j et région d'origine i , c'est-à-dire $(T)(O_i)(D_j)$. On interprète ces coefficients comme des indicateurs d'accessibilité, ou de facilité d'interaction ou d'attractivité entre deux régions (Rogers, Willekens, Little *et al.* 2002).

Des valeurs égales à 1,0 signifient l'indépendance, c'est-à-dire l'absence d'association entre origine et destination. Selon l'équation 1, si un coefficient OD_{ij} vaut 1,0, n_{ij} est déterminé uniquement par les valeurs de T , O_i et D_j . Un écart par rapport 1,0 dans n'importe quel sens est le signe d'une association entre origine et destination. Des valeurs supérieures à 1,0 représentent des niveaux d'accessibilité/attractivité supérieurs aux attentes, et des valeurs inférieures

à 1,0 représentent des niveaux d'accessibilité/attractivité inférieurs aux attentes.

Comme les effets de la diagonale de 1973 sont généralement supérieurs à 1,0, il semble que, de manière imprévue, les migrants aient été attirés par des destinations appartenant à la même catégorie que leurs communes d'origine. La 6^e catégorie fait exception : les migrants originaires de grandes villes de plus de 100 000 habitants étaient plus attirés par des communes de navetteurs (3^e catégorie) que par d'autres grandes villes (2,015 est supérieur à 0,676).

Le tableau 37.2 présente tous les paramètres dont on a besoin pour reproduire les flux de 1973 et de 1976. Pour vérifier que n'importe quel flux du tableau 37.1 peut être reproduit par les composantes multiplicatives, prenons, par exemple, le flux de la 2^e catégorie vers la 3^e en 1973 :

$$n_{2,3} = 6953 = 17168,003 \times 0,962 \times 0,644 \times 0,653.$$

Les valeurs des paramètres ne sont pas toutes indépendantes les unes des autres. En d'autres termes, les valeurs de certains paramètres peuvent être déduites de celles des autres. Pour une année d'observation et pour toutes les combinaisons de i et de j , il y a 36 effets d'interaction, 6 effets principaux de la région d'origine et 6 de la région de destination, ainsi qu'un effet général, comme on le voit dans le tableau 37.2. Mais ces 49 paramètres, pour chaque année, ont été établis à partir de 36 flux observés seulement, ce qui veut dire que 13 paramètres sont redondants. Autrement dit, 13 des 49 paramètres peuvent être déduits des 36 autres, et la relation entre les paramètres est déterminée par les contraintes suivantes associées au calcul des effets des facteurs sur la base de la moyenne géométrique. La première série de contraintes oblige les produits des effets principaux de la région d'origine (et ceux de la région de destination) à être égaux à 1. Ce qui s'écrit :

$$\prod_i O_i = 1 \text{ et } \prod_j D_j = 1.$$

La deuxième série de contraintes concerne les éléments d'interaction de chaque ligne et de chaque colonne : les produits des éléments intérieurs de chaque ligne (et de chaque colonne) doivent être égaux à 1. En d'autres termes, si cinq des effets d'interaction associés à une région d'origine (ou de destination) sont connus, le sixième s'en déduit automatiquement. Ceci s'écrit :

$$\prod_i O_{ij} = 1 \text{ et } \prod_j D_{ij} = 1.$$

En général, s'il y a m régions, il y a m^2 paramètres linéairement indépendants et $1 + m + m + (m \times m)$ composantes multiplicatives. Pour l'ensemble des calculs des effets des facteurs sur la base de la moyenne géométrique, voir le tableau 37.2 de la feuille de calcul *Multiplicative Components* du fichier Excel associé à ce chapitre et disponible en ligne.

Application 3 : Calcul des effets des facteurs sur la base du total général

Le calcul des effets des facteurs sur la base de la moyenne géométrique, qui utilise la moyenne géométrique comme valeur de référence, est le plus ancien mode de décomposition employé pour décrire la migration (Rogers, Willekens, Little *et al.* 2002 ; Willekens 1983). Mais récemment, le calcul des effets des facteurs sur la base du total général est devenu plus populaire (Raymer et Rogers 2007 ; Rogers, Little et Raymer 2010). Les deux types de décomposition satisfont l'équation 1, mais le calcul des effets des facteurs sur la base du total général est plus transparent. Par exemple, l'effet global T est maintenant le nombre total de migrants, noté n_{++} , O_i est la proportion de l'ensemble des migrants qui sortent de la région i (soit n_{i+}/n_{++}), et D_j est la proportion de tous les migrants qui entrent dans la région j (soit n_{+j}/n_{++}). La composante d'interaction, OD_{ij} , est maintenant définie comme $n_{ij}/[(T)(O_i)(D_j)]$, soit le rapport du nombre observé de migrants, n_{ij} , au nombre calculé, $(T)(O_i)(D_j)$. Tous ces effets pris ensemble constituent une autre manière de représenter la structure spatiale des migrations.

Les composantes multiplicatives obtenues par le calcul des effets des facteurs sur la base du total général sont présentées dans le tableau 37.3. Considérons, par exemple, la décomposition des 8 566 migrations des communes de la 1^{re} catégorie vers celles de la 3^e en 1973 selon les quatre composantes multiplicatives :

$$\begin{aligned} n_{13} &= (T)(O_1)(D_3)(OD_{13}) \\ &= n_{++} \left(\frac{n_{1+}}{n_{++}} \right) \left(\frac{n_{+3}}{n_{++}} \right) \left[\frac{n_{13}}{(n_{++}) \left(\frac{n_{1+}}{n_{++}} \right) \left(\frac{n_{+3}}{n_{++}} \right)} \right] \\ &= (716\,488) \left(\frac{139\,859}{716\,488} \right) \left(\frac{88\,025}{716\,488} \right) \left(\frac{8\,566}{17\,183} \right) \\ &= 716\,488(0,102)(0,190)(0,477) \\ &= 8\,566. \end{aligned}$$

Les interprétations de ces composantes sont relativement évidentes. La composante globale est le nombre total de migrations constatées en 1973, soit 716 488. La composante de la région d'origine représente la proportion du total des migrants sortant de chaque région (par exemple, 10 % de toutes les migrations sont des sorties de communes de la 1^{re} catégorie). La composante de la région de destination représente la proportion du total des migrants entrant dans chaque région (par exemple, 19 % de toutes les migrations sont des entrées dans des communes de la 3^e catégorie). Finalement, la composante d'interaction représente le rapport du nombre de migrants observés au nombre de migrants attendus ; il y a eu ainsi, en gros, 48 migrations observées entre les régions 1 et 3 pour 100 attendues. Le flux attendu est basé sur le total marginal, $(T)(O_1)(D_3)$.

Comme dans le cas du calcul des effets des facteurs sur la base de la moyenne géométrique, la décomposition obtenue par le calcul des effets des facteurs sur la base du total général donne plus de paramètres que de points de données originaux. Les contraintes qui définissent les relations entre les paramètres, et ainsi permettent le calcul des paramètres redondants, sont les suivantes :

$$\begin{aligned} \sum_i O_i &= 1; \sum_j D_j = 1; \\ \frac{\sum_i O_i \sum_j OD_{ij}}{m} &= 1, \text{ et} \\ \frac{\sum_j D_j \sum_i OD_{ij}}{m} &= 1. \end{aligned}$$

où m est le nombre de régions (Raymer, Bonaguidi et Valentin 2006).

Pour l'ensemble des calculs des effets des facteurs sur la base du total général, voir le tableau 37.3 de la feuille de calcul *Multiplicative Components* du fichier Excel associé à ce chapitre et disponible en ligne.

Comparaison des deux modèles à composantes multiplicatives

Si on applique le même mode de décomposition à deux séries de flux migratoires dans un ensemble de régions, tous les paramètres sauf T sont des nombres sans dimension. Cela signifie que mesurer le rapport entre deux séries de composantes fournit une méthode simple d'évaluation de la stabilité du régime de migration indépendamment de l'élévation ou de la diminution du niveau général de la migration.

TABLEAU 37.3 Composantes multiplicatives obtenues par application du calcul des effets des facteurs sur la base du total général

| A. Migrations 1973 | | | | | | | |
|--------------------|-------|-------|-------|-------|-------|-------|---------|
| Destination | | | | | | | |
| Origine | 1 | 2 | 3 | 4 | 5 | 6 | Total |
| 1 | 1,354 | 0,868 | 0,499 | 1,232 | 0,882 | 0,866 | 0,195 |
| 2 | 0,866 | 1,295 | 0,523 | 1,043 | 1,135 | 1,084 | 0,151 |
| 3 | 0,773 | 0,718 | 1,485 | 0,650 | 0,979 | 1,664 | 0,106 |
| 4 | 1,212 | 0,994 | 0,497 | 1,109 | 0,980 | 0,971 | 0,101 |
| 5 | 0,954 | 1,108 | 0,827 | 0,894 | 1,043 | 1,129 | 0,162 |
| 6 | 0,863 | 0,980 | 1,696 | 0,970 | 1,000 | 0,736 | 0,284 |
| TOTAL | 0,267 | 0,196 | 0,123 | 0,127 | 0,132 | 0,156 | 716 488 |
| B. Migrations 1976 | | | | | | | |
| Destination | | | | | | | |
| Origine | 1 | 2 | 3 | 4 | 5 | 6 | Total |
| 1 | 1,834 | 0,993 | 0,477 | 1,354 | 0,988 | 0,783 | 0,102 |
| 2 | 1,012 | 1,353 | 0,562 | 1,085 | 1,048 | 0,932 | 0,189 |
| 3 | 0,649 | 0,757 | 1,212 | 0,747 | 1,040 | 1,422 | 0,166 |
| 4 | 1,363 | 1,087 | 0,667 | 1,160 | 0,952 | 0,942 | 0,107 |
| 5 | 0,943 | 1,058 | 0,942 | 0,894 | 0,976 | 1,121 | 0,163 |
| 6 | 0,785 | 0,837 | 1,536 | 0,963 | 0,980 | 0,822 | 0,273 |
| TOTAL | 0,118 | 0,215 | 0,190 | 0,129 | 0,160 | 0,188 | 657 885 |

TABLEAU 37.4 Rapports entre les composantes multiplicatives de 1976 et celles de 1973

| Destination | | | | | | | |
|-------------|-------|-------|-------|-------|-------|-------|-------|
| Origine | 1 | 2 | 3 | 4 | 5 | 6 | Total |
| 1 | 1,354 | 1,144 | 0,957 | 1,099 | 1,121 | 0,904 | 0,522 |
| 2 | 1,169 | 1,045 | 1,075 | 1,040 | 0,923 | 0,860 | 1,252 |
| 3 | 0,839 | 1,055 | 0,816 | 1,149 | 1,062 | 0,854 | 1,562 |
| 4 | 1,125 | 1,093 | 1,342 | 1,046 | 0,972 | 0,970 | 1,058 |
| 5 | 0,988 | 0,955 | 1,139 | 1,000 | 0,936 | 0,993 | 1,003 |
| 6 | 0,909 | 0,854 | 0,906 | 0,993 | 0,980 | 1,117 | 0,961 |
| TOTAL | 0,441 | 1,096 | 1,546 | 1,015 | 1,214 | 1,210 | 0,918 |

tion (Rogers, Willekens, Little *et al.* 2002). Le tableau 37.4 présente les rapports des composantes de 1976 à celles de 1973. Plusieurs s'écartent sensiblement de 1, ce qui indique que le régime de migration a changé en trois ans, de 1973 à 1976. Par exemple, le rapport des composantes pour OD_{11}

est égal à 1,354, ce qui signifie que la migration entre communes de la 1^{re} catégorie était plus attractive en 1976 qu'en 1973. Par contre, le rapport des composantes pour OD_{33} est égal à 0,816, c'est-à-dire que la migration entre communes de la 3^e catégorie était moins attractive en 1976 qu'en 1973.

Application 4 : Le modèle log-linéaire additif

À côté du modèle à composantes multiplicatives, une forme alternative du modèle log-linéaire saturé est le modèle linéaire additif. Que l'on utilise la forme additive ou la forme multiplicative, les paramètres représentent la structure spatiale de la migration (Rogers, Willekens, Little *et al.* 2002), et la valeur de chaque flux peut être parfaitement reproduite par ces paramètres.

Étant donné que le modèle multiplicatif est, dans la forme, équivalent au modèle de gravitation (Willekens 1983), on le considère comme plus adéquat que le modèle linéaire additif pour la représentation des structures spatiales de la migration. D'un autre côté, la forme linéaire additive est beaucoup utilisée en statistique, et quand on emploie un logiciel statistique classique (par exemple *SPSS*, *Stata*, *R*) pour calculer un modèle log-linéaire, on obtient toujours les paramètres de la forme linéaire additive. C'est pourquoi nous présentons dans cette sous-section les calculs et interprétations habituels des paramètres du modèle linéaire additif.

Le modèle additif est une combinaison linéaire de logarithmes, ce qui justifie son appellation de modèle log-linéaire (Knoke et Burke 1980). Mathématiquement équivalent au modèle à composantes multiplicatives, il consiste à prendre les logarithmes des deux membres de l'équation 1 :

$$\ln(n_{ij}) = \ln(T) + \ln(O_i) + \ln(D_j) + \ln(OD_{ij})$$

ou, sous une forme plus condensée :

$$\ln(n_{ij}) = \lambda + \lambda_i^O + \lambda_j^D + \lambda_{ij}^{OD}. \quad (2)$$

Les λ sont simplement les logarithmes naturels des paramètres de l'équation 1. Les lettres *O*, *D* et *OD* servent à distinguer les paramètres (ce ne sont pas des exposants), et les indices *i* et *j* renvoient respectivement aux catégories des régions d'origine et de destination.

En prenant les logarithmes naturels des paramètres des tableaux 2 et 3, on obtiendrait les séries correspondantes de paramètres du modèle linéaire additif. Mais, tout comme il y a au moins deux décompositions du modèle à composantes multiplicatives, le calcul des effets des facteurs sur la base de la moyenne géométrique et le calcul des effets des facteurs sur la base du total général, il y a plusieurs stratégies pour parvenir à des séries de paramètres qui satisfont le modèle linéaire additif (Powers et Xie 2008), et les approches retenues par les logiciels statistiques classiques ne sont pas simplement les transformations logarithmiques des composantes multiplicatives calculées précédemment.

Rappelons-nous qu'un système de migration à *m* régions a $m \times m$ paramètres linéairement indépendants. Les modèles à composantes multiplicatives décrits plus haut donnent des valeurs interprétables de $1 + m + m + (m \times m)$ paramètres, mais qui ne sont pas tous linéairement indépendants les uns des autres. Par ailleurs, les procédures statistiques de *SPSS*, *Stata* et *R* calculent uniquement des paramètres linéairement indépendants, soit 1 valeur pour λ^T , $m - 1$ valeurs pour λ_i^O , $m - 1$ valeurs pour λ_j^D , et $(m - 1) \times (m - 1)$ valeurs pour λ_{ij}^{OD} .

La série de valeurs calculées des paramètres dépend du mode de codage de contraste (*en anglais : contrast coding*) utilisé par le logiciel. Le codage de contraste fige une région en donnant à tous ses paramètres linéaires additifs la valeur 0. *SPSS*, par exemple, annule les paramètres de la dernière région, numérotée *m*, ici :

$$\lambda_m^O = \lambda_m^D = \lambda_{mj}^{OD} = \lambda_{im}^{OD} = 0.$$

Les paramètres obtenus par *SPSS* à partir des données néerlandaises sont détaillés dans le tableau 37.5. Les instructions de *SPSS* qui conduisent à ces résultats pour la matrice de migrations de 1973, ainsi que la sortie *SPSS*, sont présentées dans l'Appendice 1 (en ligne sur le site web *Tools for Demographic Estimation*). Le tableau 37.5 et les formules Excel de calcul des paramètres sont disponibles sur la feuille de calcul *Contrast coding* du fichier Excel associé à ce chapitre et disponible en ligne.

Les paramètres de la dernière région sont tous égaux à 0, et donc n'apportent rien à l'équation 2. L'interprétation des paramètres du tableau 37.5 est un peu compliquée, car ils sont exprimés en logarithmes. Les retransformer en composantes multiplicatives par exponentiation donne une nouvelle série de composantes multiplicatives qui satisfont l'équation 1. Ces composantes multiplicatives associées au codage de contraste de la « dernière région » figurent dans le tableau 37.6. En général, elles ne servent pas à décrire la structure spatiale des migrations, mais elles ont leur utilité pour la description des systèmes de migration, car les paramètres d'interaction, OD_{ij} , sont assimilables à des *odds ratios*.

Par exemple, le paramètre global λ^T pour les migrations de 1973 du tableau 37.5 donne le logarithme naturel des migrations observées de la région de référence :

- $\ln(n_{66}) = 10,056$, et à partir du tableau 37.6, le paramètre correspondant *T* donne le flux migratoire n_{66} :
 - $n_{66} = \exp(10,056) = 23\,304$.

TABLEAU 37.5 Paramètres linéaires additifs obtenus par codage de contraste de la « dernière région »

| A. Migrations 1973 | | | | | | | |
|--------------------|-------------|---------|---------|---------|---------|-------|---------|
| Origine | Destination | | | | | | Total |
| | 1 | 2 | 3 | 4 | 5 | 6 | |
| 1 | 0,288 | − 0,284 | − 1,388 | 0,076 | − 0,289 | 0,000 | − 0,212 |
| 2 | − 0,384 | − 0,109 | − 1,565 | − 0,315 | − 0,261 | 0,000 | − 0,243 |
| 3 | − 0,926 | − 1,128 | − 0,949 | − 1,216 | − 0,837 | 0,000 | − 0,168 |
| 4 | 0,062 | − 0,262 | − 1,505 | − 0,143 | − 0,297 | 0,000 | − 0,753 |
| 5 | − 0,327 | − 0,304 | − 1,146 | − 0,509 | − 0,385 | 0,000 | − 0,133 |
| 6 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 |
| TOTAL | 0,698 | 0,518 | 0,598 | 0,071 | 0,141 | 0,000 | 10,056 |
| B. Migrations 1976 | | | | | | | |
| Origine | Destination | | | | | | Total |
| | 1 | 2 | 3 | 4 | 5 | 6 | |
| 1 | 0,897 | 0,219 | − 1,122 | 0,389 | 0,057 | 0,000 | − 1,033 |
| 2 | 0,129 | 0,355 | − 1,132 | − 0,007 | − 0,059 | 0,000 | − 0,240 |
| 3 | − 0,738 | − 0,648 | − 0,785 | − 0,802 | − 0,488 | 0,000 | 0,049 |
| 4 | 0,416 | 0,125 | − 0,971 | 0,050 | − 0,165 | 0,000 | − 0,798 |
| 5 | − 0,126 | − 0,075 | − 0,799 | − 0,385 | − 0,314 | 0,000 | − 0,208 |
| 6 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 | 0,000 |
| TOTAL | − 0,517 | 0,151 | 0,634 | − 0,222 | 0,013 | 0,000 | 10,233 |

Un autre exemple concernant les migrations de 1973 du tableau 37.5 montre comment les effets principaux des régions d'origine, λ_i^O , s'ajoutent au paramètre global pour reproduire les migrations des communes de 1^{re} catégorie vers celles de la catégorie de référence, la 6^e, du tableau 37.1. Par exemple :

- $\ln(n_{16}) = 10,056 - 0,212 = 9,845$, et la composante multiplicative correspondante, O_1 multipliée par le T du tableau 37.6 donne :
- $n_{16} = 27\,810 \times 0,356 = 18\,856$.

Avec la même approche, on peut reproduire les logarithmes de tous les flux migratoires en appliquant l'équation 1, avec les paramètres appropriés du tableau 37.6, ou reproduire les flux observés en appliquant l'équation 2 avec les paramètres du tableau 37.5.

Les paramètres d'interaction du modèle linéaire, λ_{ij}^{OD} , sont des logarithmes d'*odds ratios* (LOR) : un LOR est le logarithme du rapport de deux cotes (*odds*) : (1) la cote de la migration vers la région j par rapport à la migration vers la région de référence, sous la condition que la région d'origine

soit la région i , et (2) la cote de la migration vers la région j par rapport à la migration vers la région de référence, sous la condition que la région d'origine soit la région de référence. Par exemple, à partir de la matrice de 1973 du tableau 37.5, $\lambda_{23}^{OD} = -1,565$, calculé ainsi :

$$\lambda_{23}^{OD} = \ln \left[\frac{\frac{n_{23}}{n_{26}}}{\frac{n_{63}}{n_{66}}} \right] = \ln \left[\frac{\frac{6953}{18282}}{\frac{42399}{23304}} \right] = -1,565.$$

Pour le dire avec des mots, ce paramètre est le logarithme du rapport entre la cote de la migration vers une commune de la 3^e catégorie plutôt que de la 6^e, pour un migrant sortant d'une commune de 2^e catégorie et cette même cote pour un migrant sortant d'une commune de la 6^e catégorie.

Les *odds ratios* mesurent la probabilité d'un événement comparativement à celle d'un autre événement. Comme ils sont plus classiques que les LOR, il peut être plus facile de passer à l'exponentielle des LOR et d'interpréter les para-

TABLEAU 37.6 Composantes multiplicatives obtenues par codage de contraste de la « dernière région »

| A. Migrations 1973 | | | | | | | |
|--------------------|-------------|-------|-------|-------|-------|-------|--------|
| Origine | Destination | | | | | | Total |
| | 1 | 2 | 3 | 4 | 5 | 6 | |
| 1 | 1,333 | 0,753 | 0,250 | 1,079 | 0,749 | 1,000 | 0,809 |
| 2 | 0,681 | 0,897 | 0,209 | 0,730 | 0,770 | 1,000 | 0,785 |
| 3 | 0,396 | 0,324 | 0,387 | 0,296 | 0,433 | 1,000 | 0,845 |
| 4 | 1,064 | 0,769 | 0,222 | 0,867 | 0,743 | 1,000 | 0,471 |
| 5 | 0,721 | 0,738 | 0,318 | 0,601 | 0,680 | 1,000 | 0,876 |
| 6 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 |
| TOTAL | 2,009 | 1,678 | 1,819 | 1,073 | 1,151 | 1,000 | 23 304 |
| B. Migrations 1976 | | | | | | | |
| Origine | Destination | | | | | | Total |
| | 1 | 2 | 3 | 4 | 5 | 6 | |
| 1 | 2,453 | 1,245 | 0,326 | 1,475 | 1,059 | 1,000 | 0,356 |
| 2 | 1,138 | 1,426 | 0,322 | 0,993 | 0,943 | 1,000 | 0,786 |
| 3 | 0,478 | 0,523 | 0,456 | 0,448 | 0,614 | 1,000 | 1,051 |
| 4 | 1,516 | 1,133 | 0,379 | 1,051 | 0,848 | 1,000 | 0,450 |
| 5 | 0,882 | 0,928 | 0,450 | 0,681 | 0,731 | 1,000 | 0,812 |
| 6 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 | 1,000 |
| TOTAL | 0,596 | 1,163 | 1,885 | 0,801 | 1,013 | 1,000 | 27 810 |

mètres d'interaction du tableau 37.6 comme des *odds ratios*. Par exemple, le paramètre OD_{23} , dans le cas des données de 1973, se calcule ainsi :

$$OD_{23} = \exp(-1,565) = \left[\frac{n_{23}}{n_{26}} \frac{n_{63}}{n_{66}} \right] = 0,209.$$

En une phrase, les chances qu'un migrant sortant d'une commune de 2^e catégorie entre dans une commune de 3^e catégorie plutôt que dans une commune de 6^e catégorie valent approximativement 1/5 des chances qu'un migrant sortant d'une commune de 6^e catégorie opte pour une commune de 3^e catégorie plutôt que pour une commune de 6^e catégorie. Les *odds ratios* sont toujours positifs et dépendent toujours du choix de la catégorie de référence. Un *odds ratio* égal à 1 signifie l'absence de toute relation, c'est-à-dire l'indépendance statistique. Une valeur supérieure à 1 indique une

association positive, et une valeur inférieure à 1 indique une association négative.

Stata et *R* utilisent le codage de contraste de la « première région », alors que *SPSS* emploie le codage de contraste de la « dernière région ». Dans ces deux logiciels, on donne la valeur 0 aux paramètres de la première région, celle à qui est affectée la première modalité, soit

$$\lambda_1^O = \lambda_1^D = \lambda_{1j}^{OD} = \lambda_{i1}^{OD} = 0.$$

Les instructions de *Stata* et de *R* pour le calcul des paramètres linéaires additifs, ainsi que les résultats correspondants, sur l'exemple des flux migratoires de 1973, peuvent être téléchargés à partir de l'Appendice 1 associé à ce chapitre (en ligne sur le site web *Tools for Demographic Estimation*).

Toutes les variantes du modèle saturé et toutes les méthodes statistiques d'estimation des paramètres d'interaction sont concordantes, et leurs résultats sont très largement similaires. Les formules de calcul des paramètres sont disponibles dans la feuille de calcul *Linear Additive Parameters* du fichier Excel

associé à ce chapitre et disponible en ligne. De plus, *SPSS* et *Stata* vérifient automatiquement que chaque paramètre d'interaction linéaire additif est égal à 0. On trouvera ces résultats dans l'Appendice 1 associé à ce chapitre (en ligne sur le site web *Tools for Demographic Estimation*), et ils montrent que chaque paramètre d'interaction non redondant est statistiquement significatif. Pour les variances des estimations, voir Agresti et Finlay (2009) et Powers et Xie (2008).

Application 5 : Le modèle d'indépendance

Tous les modèles vus jusqu'à présent sont des modèles saturés, et par conséquent, ils représentent parfaitement les flux observés. En général, les paramètres les plus intéressants sont les paramètres d'interaction, parce qu'ils indiquent les associations entre paires de régions. Le modèle d'indépendance, quant à lui, repose sur l'hypothèse que les paramètres d'interaction sont sans intérêt et superflus, parce que tous les paramètres d'interaction multiplicatifs, OD_{ij} , sont égaux à 1, ou, ce qui revient au même, tous les paramètres d'interaction additifs, λ_{ij}^{OD} , sont égaux à 0. Le modèle d'indépendance implique que les termes d'interaction doivent sortir du modèle, ce qui réduit celui-ci à la forme plus modeste d'un modèle à deux variables, soit $n_{ij} = (T)(O_i)(D_j)$ ou $\ln(n_{ij}) = \lambda + \lambda_i^O + \lambda_j^D$.

L'examen visuel des paramètres d'interaction du modèle log-linéaire saturé est un moyen de tester l'hypothèse d'indépendance. Une autre méthode consiste à calculer les distributions conditionnelles de ligne ou de colonne. Si les distributions conditionnelles des lignes (origine) sont identiques, il y a indépendance entre régions d'origine et de destination. En outre, l'indépendance étant une propriété symétrique, si les distributions conditionnelles des lignes (origines) sont identiques, celles des colonnes (destinations) le sont également (Agresti et Finlay 2009 ; Powers et Xie 2008). Sur la feuille de calcul *Independence* du fichier Excel disponible en ligne, on a calculé la répartition en pourcentage des migrations néerlandaises dans chaque colonne (destinations). Ces pourcentages de colonne sont très variés, et cela signifie, comme l'indiquent aussi les paramètres d'interaction, que l'indépendance statistique n'est pas vérifiée dans ce cas.

L'hypothèse d'indépendance implique que chaque flux interrégional puisse être déterminé à partir du volume des flux marginaux. Soit N_{ij} le flux attendu entre les régions i et j si l'hypothèse d'indépendance est vérifiée. N_{ij} est alors égal au total des flux du système de migration, n_{++} , multiplié par la proportion de l'ensemble des migrants qui

sortent de la région i , n_{i+}/n_{++} , et par la proportion de l'ensemble des migrants qui entrent dans la région j , n_{+j}/n_{++} , soit $N_{ij} = n_{++}(n_{i+}/n_{++})(n_{+j}/n_{++})$. Si on peut soutenir l'hypothèse d'indépendance, N_{ij} est une bonne estimation d'un flux interrégional, et le problème de l'estimation des flux migratoires interrégionaux s'en trouve réellement simplifié.

Les différences entre flux observés, n_{ij} , et flux attendus, N_{ij} , sont à la base de l'évaluation de la qualité d'ajustement et du chi-carré de Pearson, χ^2 , largement utilisé pour résumer ces différences. Il se calcule par la formule :

$$\chi^2 = \sum \frac{(n_{ij} - N_{ij})^2}{N_{ij}},$$

où la sommation porte sur toutes les cellules intérieures de la matrice de migration. Quand les flux observés et les flux attendus concordent parfaitement, dans toutes les cellules du tableau, le χ^2 est égal à 0, ce qui signifie que le modèle d'indépendance s'ajuste parfaitement aux données observées. Plus l'écart entre n_{ij} et N_{ij} est important, plus la valeur du χ^2 est élevée, et plus il est clair que le modèle d'indépendance ne convient pas. De manière générale, les faibles valeurs indiquent un bon ajustement et les valeurs élevées un ajustement médiocre.

Si l'hypothèse d'indépendance est vérifiée, le χ^2 est déterminé par la distribution de probabilités du χ^2 à $(m-1) \times (m-1)$ degrés de liberté. C'est sur la base de cette distribution que l'on teste la significativité du χ^2 (Agresti 2007 ; Agresti et Finlay 2009). Si le χ^2 se situe vers l'extrémité droite de sa distribution, cela signifie qu'il y a peu de chances (par exemple $p < 0,05$) que l'hypothèse d'indépendance soit vérifiée, et le modèle est alors rejeté. Les valeurs du χ^2 associées au modèle d'indépendance appliqué aux données néerlandaises du tableau 37.1 ont été calculées dans la feuille de calcul *Independence* du fichier Excel disponible en ligne. Pour les instructions *SPSS*, *Stata* et *R* permettant de tester le modèle d'indépendance avec les données de 1973, voir l'Appendice 2 (disponible sur le site web *Tools for Demographic Estimation*).

La valeur du χ^2 associée aux données de 1973 de notre exemple est 47 623, avec 25 degrés de liberté (df). La valeur de p correspondante est inférieure à 0,000, et l'hypothèse d'indépendance est rejetée. (Cependant, on trouvera plus loin des commentaires sur les limites de ce test quand on a affaire à un grand échantillon.) Cela ne doit pas surprendre, vu les trois décompositions multiplicatives des données néer-

landaises présentées dans les tableaux 2, 3 et 6. Les données montrent de façon constante de fortes associations entre régions, et beaucoup de paramètres d'interaction multiplicatifs sont très différents de 1. De plus, les variances fournies par *SPSS* et *Stata*, présentées dans l'Appendice 1 (disponible sur le site web *Tools for Demographic Estimation*), indiquent que les paramètres d'interaction linéaires additifs sont nettement différents de 0.

Une alternative au χ^2 est le rapport de vraisemblance, ou déviation, ou coefficient G^2 . Ce sont des appellations différentes du même test, et le choix de l'une ou l'autre dépend des préférences des auteurs de manuels et des concepteurs de logiciels. Par simplicité, nous emploierons ici G^2 . Le G^2 est similaire au χ^2 en ce que des valeurs proches de zéro indiquent un bon ajustement du modèle, et des valeurs élevées un ajustement médiocre. Si l'hypothèse du modèle d'indépendance est valable, le G^2 a une distribution du type de celle du χ^2 .

L'utilité du G^2 dépasse largement le test du modèle d'indépendance dans une analyse log-linéaire. Il sert souvent à comparer un modèle simple à un modèle plus complexe. Il se calcule à partir du rapport entre deux probabilités : (1) la probabilité que le modèle contraint (ici le modèle d'indépendance) s'ajuste bien aux données, et (2) la probabilité que le modèle non contraint (ici le modèle saturé) s'ajuste bien aux données. Si ce rapport est proche de 1, on préférera le modèle contraint, plus simple et plus parcimonieux, car il représente les données aussi bien que le modèle plus complexe.

Le rapport de ces deux probabilités n'a pas de distribution de type χ^2 . Mais, si le rapport est transformé en logarithme naturel et multiplié par -2 , il devient G^2 , variable qui a une distribution de type χ^2 avec $(m-1) \times (m-1)$ degrés de liberté. Si on appelle L_c la probabilité associée au modèle contraint (le modèle d'indépendance) et L_u la probabilité associée au modèle non contraint (le modèle saturé), G^2 se calcule comme suit :

$$G^2 = -2 \ln \left(\frac{L_c}{L_u} \right) = -2 \ln L_c + 2 \ln L_u.$$

Puisque le modèle saturé s'ajuste parfaitement aux données (c'est-à-dire que $L_u = 1$), $G^2 = -2 \ln L_c$. Ses valeurs, selon l'exemple traité et le logiciel choisi, se trouvent dans l'Appendice 2 (voir le site web). Pour *SPSS* et *Stata*, il s'appelle « *Deviance* » et vaut 46 477,63 ; pour *R*, il s'appelle « *Residual Deviance* » et est arrondi à 46 480. Avec 25 degrés de liberté,

la probabilité que le modèle d'indépendance convienne est bien égale à 0.

Les paramètres χ^2 et G^2 sont asymptotiquement équivalents (Powers et Xie 2008) et constituent, respectivement, les bases du test du chi-carré de Pearson et du test du rapport de vraisemblance. Comme pour tous les tests inférentiels, l'utilisateur doit être attentif aux hypothèses sous-jacentes et aux limites. Ces deux tests reposent sur l'hypothèse que chaque flux interrégional de la matrice de migrations suit une distribution de Poisson propre (Powers et Xie 2008), et ils ont tous les deux des limites importantes liées à la taille de l'échantillon. Le χ^2 est artificiellement « gonflé » quand l'échantillon est grand. Le chi-carré de Pearson n'est donc pas indiqué en présence de gros échantillons. Le G^2 et le test du rapport de vraisemblance sont préférables en pareil cas (Powers et Xie 2008). On opte plutôt pour le test du chi-carré de Pearson quand les effectifs attendus sont, en moyenne, compris entre 1 et 10, mais aucun test n'est vraiment satisfaisant quand la plupart des effectifs attendus sont inférieurs à 5 (Agresti et Finlay 2009 ; Powers et Xie 2008).

On a également mis en doute l'efficacité du G^2 dans le cas de grands échantillons (Raftery 1986, 1995), et on s'accorde de plus en plus à estimer que des mesures d'information doivent être envisagées parallèlement aux tests de significativité traditionnels pour évaluer la qualité d'ajustement d'un modèle. Le critère d'information bayésien (*en anglais : Bayesian Information Criterion, BIC*) est étroitement relié à G^2 , et il se calcule en *Stata* par la formule :

$$BIC = G^2 - df \ln(m \times m),$$

et en *SPSS* par la formule :

$$BIC = -2 \ln L_c + p \ln(m \times m),$$

où p est le nombre de paramètres estimés par le modèle d'indépendance, soit $2m - 1$. Une valeur faible incite à préférer le modèle d'indépendance au modèle saturé (Powers et Xie 2008).

Le critère d'information d'Akaike (*en anglais : Akaike's Information Criterion, AIC*) est un outil alternatif qui accepte les valeurs faibles afin d'apprécier l'ajustement des meilleurs modèles, car il évalue la proximité entre valeurs ajustées et valeurs attendues (Agresti 2007). En *SPSS* et en *R*, il se calcule ainsi :

$$AIC = -2(\ln L_c - p),$$

où p est le nombre de paramètres estimés dans le modèle d'indépendance, soit $2m - 1$. En *Stata*, la formule est :

$$AIC = \frac{-2(\ln L_c - p)}{m \times m}.$$

On peut constater dans l'Appendice 2 (disponible sur le site web *Tools for Demographic Estimation*) que *SPSS* et *Stata* calculent le *BIC* et l'*AIC*, tandis que *R* ne donne qu'une valeur arrondie de l'*AIC*. Comme on l'a déjà signalé, les formules utilisées présentent des différences. La valeur du *BIC* est 46 934,237 selon *SPSS* et 46 388,04 selon *Stata*. L'*AIC*, seul fourni par *R*, vaut 46 920, sa valeur arrondie étant 46 916,818 selon *SPSS*. *Stata* en donne une estimation beaucoup plus faible : 1 303,245. Toutes les valeurs du *BIC* et de l'*AIC* sont grandes, et renforcent l'idée que le modèle d'indépendance ne convient pas à cet exemple.

Le modèle de quasi-indépendance

Le modèle d'indépendance s'ajuste rarement bien à des données migratoires. Cela est dû en partie à la très forte tendance des gens à rester dans la région où ils vivent. Le modèle de quasi-indépendance permet d'éliminer ces « effets d'immobilité » (Powers et Xie 2008), et cela a souvent comme résultat d'améliorer l'estimation des flux interrégionaux attendus. Ce modèle a été appliqué avec succès à des données migratoires provenant de recensements nationaux (Agresti 1990 ;

Rogers, Little et Raymer 2010 ; Rogers, Willekens, Little et al. 2002), les personnes qui ont déclaré habiter la même région qu'au début de la période de référence figurant dans la diagonale de la matrice de migrations.

À titre d'illustration, le cadre A du tableau 37.7 présente les migrations effectuées aux États-Unis entre 1985 et 1990 par des Américains de naissance. Manifestement, les effectifs des quatre cases de la diagonale de la matrice (hors marges) sont nettement supérieurs à ceux des autres cases, ce qui signifie que la propension à rester dans la région où l'on vit est beaucoup plus répandue que la migration d'une région à l'autre.

Cette concentration des effectifs le long de la diagonale est largement responsable du mauvais ajustement du modèle d'indépendance, et le poids écrasant des personnes qui ne quittent pas leur région d'origine a incité les chercheurs à les retirer du modèle. Les migrants étant définis comme les personnes qui changent de région de résidence, ce type de matrice des flux est parfois appelé « matrice limitée aux migrants ». Elle est particulièrement utile pour l'étude de la structure de la migration, puisqu'elle écarte les personnes qui n'ont pas migré ou qui ont déménagé sans changer de région. Le tableau 37.7 présente, dans le cadre B, la matrice des flux où les éléments de la diagonale ont été annulés et les totaux marginaux adaptés en conséquence.

| A. Matrice des migrations complète | | | | | |
|------------------------------------|-------------|------------|------------|------------|-------------|
| | Destination | | | | |
| Origine | Nord-Est | Midwest | Sud | Ouest | Total |
| Nord-Est | 40 262 319 | 336 091 | 1 645 843 | 479 819 | 42 724 072 |
| Midwest | 351 029 | 50 677 007 | 1 692 687 | 958 696 | 53 679 419 |
| Sud | 778 868 | 1 197 134 | 69 563 871 | 1 150 649 | 72 690 522 |
| Ouest | 348 892 | 668 979 | 1 082 104 | 37 872 893 | 39 972 868 |
| TOTAL | 41 741 108 | 52 879 211 | 73 984 505 | 40 462 057 | 209 066 881 |
| B. Matrice limitée aux migrants | | | | | |
| | Destination | | | | |
| Origine | Nord-Est | Midwest | Sud | Ouest | Total |
| Nord-Est | 0 | 336 091 | 1 645 843 | 479 819 | 2 461 753 |
| Midwest | 351 029 | 0 | 1 692 687 | 958 696 | 3 002 412 |
| Sud | 778 868 | 1 197 134 | 0 | 1 150 649 | 3 126 651 |
| Ouest | 348 892 | 668 979 | 1 082 104 | 0 | 2 099 975 |
| TOTAL | 1 478 789 | 2 202 204 | 4 420 634 | 2 589 164 | 10 690 791 |

TABLEAU 37.7 Flux migratoires des personnes nées aux États-Unis, 1985-1990

Les composantes multiplicatives, avec calcul des effets des facteurs sur la base du total général, fournies par ces deux types de matrices sont présentées dans le tableau 37.8. Les valeurs des paramètres du modèle à composantes multiplicatives basées sur la matrice complète s'éloignent manifestement de celles que l'on attendait sous l'hypothèse d'indépendance. Elles sont largement supérieures à 1,0 dans la diagonale, et très nettement inférieures à 1,0 en dehors. Par comparaison, quand on donne aux composantes multiplicatives basées sur la matrice limitée aux migrants la valeur 0 pour reproduire les 0 structurels de la diagonale, les composantes hors de la diagonale sont plus proches de 1,0.

Le modèle de quasi-indépendance exige que seules les migrations d'une région à l'autre satisfassent l'hypothèse d'indépendance. On s'en assure par deux voies différentes mais équivalentes. Dans la première méthode, on considère la matrice complète (cadre A du tableau 37.7) et on donne la valeur 0 aux pondérations des effets d'interaction, OD_{ij} , quand la région de destination est identique à la région d'origine : $n_{ij} = 0$ si $i = j$. On parle alors de zéros structurels. Quand les régions d'origine et de destination sont différentes ($i \neq j$), les effets d'interaction reçoivent la valeur 1,0, ce qui

rejoint le modèle d'indépendance que nous connaissons et donne les flux hors diagonale attendus sous l'hypothèse de quasi-indépendance. L'Appendice 3 (disponible sur le site web *Tools for Demographic Estimation*) présente l'application de cette méthode en *SPSS*, *Stata* et *R*.

La deuxième méthode s'appuie sur la matrice limitée aux migrants (cadre B du tableau 37.7). Elle se présente de préférence sous la forme additive : $\ln(n_{ij}) = \lambda + \lambda_i^O + \lambda_j^D + \delta_i I$, où I est une variable dichotomique qui prend la valeur 1 pour les flux de la diagonale, donc quand $i = j$, et la valeur 0 pour les flux hors diagonale, donc quand $i \neq j$ (Agresti 2002). Un paramètre supplémentaire, δ_i , est donc nécessaire pour estimer chaque flux de la diagonale ; pour les autres flux interrégionaux, le terme $\delta_i I$ disparaît, et le modèle de quasi-indépendance se ramène alors au modèle d'indépendance. Par conséquent, exactement comme dans le modèle d'indépendance, les termes d'interaction hors diagonale sont nécessairement égaux à 0 dans la forme additive du modèle (et à 1 dans la forme multiplicative). L'Appendice 3 présente l'application de cette méthode en *Stata* (sur le site web *Tools for Demographic Estimation*).

TABLEAU 37.8 Composantes multiplicatives des flux migratoires des personnes nées aux États-Unis, 1985-1990

| A. Matrice des migrations complète | | | | | |
|--|-------------|---------|-------|-------|-------------|
| Origine | Destination | | | | Total |
| | Nord-Est | Midwest | Sud | Ouest | |
| Nord-Est | 4,720 | 0,031 | 0,109 | 0,058 | 0,204 |
| Midwest | 0,033 | 3,733 | 0,089 | 0,092 | 0,257 |
| Sud | 0,054 | 0,065 | 2,704 | 0,082 | 0,348 |
| Ouest | 0,044 | 0,066 | 0,076 | 4,896 | 0,191 |
| TOTAL | 0,200 | 0,253 | 0,354 | 0,194 | 209 066 881 |
| B. Matrice des migrations limitée aux migrants | | | | | |
| Origine | Destination | | | | Total |
| | Nord-Est | Midwest | Sud | Ouest | |
| Nord-Est | 0,000 | 0,663 | 1,617 | 0,805 | 0,230 |
| Midwest | 0,845 | 0,000 | 1,363 | 1,318 | 0,281 |
| Sud | 1,801 | 1,859 | 0,000 | 1,520 | 0,292 |
| Ouest | 1,201 | 1,547 | 1,246 | 0,000 | 0,196 |
| TOTAL | 0,138 | 0,206 | 0,413 | 0,242 | 10 690 791 |

Dans la première méthode, le modèle de quasi-indépendance assigne la valeur 0 aux m paramètres OD_{ii} pour $i = 1$ à m . Dans la seconde méthode, m paramètres supplémentaires, les δ_i , sont estimés et leurs exponentielles sont très proches de 0. Quelle que soit la méthode utilisée, le modèle de quasi-indépendance a m paramètres de plus que le modèle d'indépendance, et m degrés de liberté de moins.

L'Appendice 3 montre comment calculer le modèle de quasi-indépendance à l'aide des logiciels statistiques *SPSS*, *Stata* et *R*, en utilisant les données de migration aux États-Unis entre 1985 et 1990 des Américains de naissance. Quand le modèle d'indépendance exploite la totalité des données, tous les indices de qualité de l'ajustement sont, comme prévu, extrêmement élevés : $\chi^2 = 544\,479\,395$ ($df=9$) ; $G^2 = 461\,411\,576$ ($df=9$) ; dans *Stata*, les valeurs du *BIC* et de l'*AIC* sont respectivement 461 000 000 et 28 800 000. Quand on calcule le modèle de quasi-indépendance, toutes ces valeurs sont considérablement réduites : $\chi^2 = 327\,233$ ($df=5$) ; $G^2 = 330\,220$ ($df=5$) ; dans *Stata*, les valeurs du *BIC* et de l'*AIC* sont respectivement 330 207 et 27 535.

Les tests inférentiels restent significatifs, et le modèle de quasi-indépendance ne peut pas être accepté comme véritable modèle de migration. Il ne faut pas comparer inférieurement les modèles d'indépendance et de quasi-indépen-

dance avec le test du rapport de vraisemblance, car ce ne sont pas des modèles emboîtés. Cependant, les mesures d'information sont directement comparables. Tant le *BIC* que l'*AIC* sont nettement réduits, ce qui avantage le modèle de quasi-indépendance par rapport au modèle d'indépendance.

Par ailleurs, le tableau 37.9 met en parallèle les flux attendus du modèle d'indépendance et ceux du modèle de quasi-indépendance. La comparaison visuelle des flux attendus du tableau 37.9 avec les données observées du tableau 37.7 montre combien le modèle de quasi-indépendance est plus proche des données. Deux indices résumés supplémentaires sont présentés : R^2 et l'*EAMP* (écart absolu moyen en pourcentage). La comparaison des R^2 montre que le modèle d'indépendance explique 10 % de la variation des données observées, tandis que le modèle de quasi-indépendance en explique 95 %. De plus, l'*EAMP* du modèle de quasi-indépendance, égal à 28, est considérablement plus faible que celui du modèle d'indépendance, qui vaut 2 492.

Comme l'ajustement du modèle de quasi-indépendance aux données observées n'est pas suffisamment précis, on ne peut pas considérer celui-ci comme le « vrai » modèle. Mais, en l'absence de données migratoires observées, le modèle de quasi-indépendance peut toujours constituer une solution de secours raisonnable pour estimer les flux interrégionaux.

| A. Indépendance | | | | |
|-----------------------|------------|------------|---------------------|------------|
| Destination | | | | |
| Origine | 1 | 2 | 3 | 4 |
| 1 | 8 530 046 | 10 806 184 | 15 119 178 | 8 268 664 |
| 2 | 10 717 328 | 13 577 116 | 18 996 052 | 10 388 923 |
| 3 | 14 512 977 | 18 385 588 | 25 723 693 | 14 068 264 |
| 4 | 7 980 756 | 10 110 323 | 14 145 583 | 7 736 206 |
| $R^2 = 0,104$ | | | $EAMP = 2\,492,322$ | |
| B. Quasi-indépendance | | | | |
| Destination | | | | |
| Origine | 1 | 2 | 3 | 4 |
| 1 | 0 | 535 839 | 1 349 561 | 576 353 |
| 2 | 442 768 | 0 | 1 793 640 | 766 005 |
| 3 | 720 681 | 1 159 163 | 0 | 1 246 806 |
| 4 | 315 340 | 507 201 | 1 277 434 | 0 |
| $R^2 = 0,945$ | | | $EAMP = 27,575$ | |

TABLEAU 37.9 Flux migratoires attendus des Américains de naissance, sous les hypothèses d'indépendance et de quasi-indépendance, 1985-1990

Application 6 : La méthode des offsets

On peut évaluer la validité des modèles d'indépendance et de quasi-indépendance à l'aide des tests inférentiels associés aux résultats du modèle log-linéaire, et, même quand les modèles ne sont pas étayés par des tests de significativité, on peut les utiliser dans certains contextes pour obtenir des estimations valables des flux migratoires. La méthode des offsets suppose que les données auxiliaires ont une structure implicite des relations interrégionales similaire à la structure des migrations que l'on cherche à déterminer. Cette méthode emprunte la structure des données auxiliaires pour en déduire des estimations des flux migratoires inconnus.

Dans les recherches réalisées jusqu'à présent, c'est généralement une matrice de flux migratoires d'une autre période qui sert d'information auxiliaire (Rogers, Little et Raymer 2010 ; Rogers, Willekens, Little *et al.* 2002 ; Rogers, Willekens et Raymer 2003 ; Willekens 1983), mais elle pourrait concerner une autre catégorie d'âge (Raymer et Rogers 2007), de sexe ou de race. Il pourrait aussi bien s'agir de données provenant d'une autre source, comme les déclarations d'impôt ou les registres d'immatriculation des véhicules à moteur.

Les flux auxiliaires étant notés n_{ij}^* , le modèle log-linéaire avec offset prend la forme : $\ln(\tilde{n}_{ij}) = \lambda + \lambda_i^O + \lambda_j^D + \ln(n_{ij}^*)$. Il permet d'estimer des flux \hat{n}_{ij} dont la structure est aussi proche que possible de celle des flux auxiliaires, et, en même temps, ces flux estimés sont ajustés de manière à ce que leur somme corresponde aux totaux marginaux préalablement fixés par le chercheur. De cette manière, la méthode des offsets est similaire aux modèles d'indépendance et de quasi-indépendance en ce sens qu'elle fournit une distribution attendue des flux telle que les totaux marginaux de ligne et de colonne soient égaux aux estimations *a priori*.

Pour illustrer le fonctionnement de la méthode des offsets, prenons la matrice des flux migratoires aux Pays-Bas en 1976 (tableau 37.1). Supposons que nous voulons conserver les totaux marginaux, mais en même temps remplacer les effets d'interaction observés pendant cette année par ceux de 1973, en utilisant la méthode des offsets. Quelle serait la série correspondante de paramètres log-linéaires ? Le cadre A du tableau 37.10 présente la matrice des flux attendus obtenue par la méthode des offsets, et le cadre B, les composantes multiplicatives correspondantes avec calcul des effets des facteurs sur la base du total général. Notons que les valeurs de T , O_i et D_j de la matrice des flux attendus, c'est-à-dire le cadre B du tableau 37.10, sont identiques à celles de la

matrice des flux observés de 1976 (cadre B du tableau 37.3). Mais les autres termes (les effets d'interaction OD_{ij}) reflètent l'influence de la structure des données observées de 1973 (cadre A du tableau 37.3), ainsi que les totaux marginaux des données de 1976. La méthode des substitutions applique la structure des données auxiliaires, celles de 1973 dans ce cas-ci, aux flux des cellules intérieures de la matrice tout en conservant le volume total des flux observés en 1976.

Les valeurs attendues qui figurent au cadre A du tableau 37.10 ont été empruntées aux résultats des programmes *SPSS*, *Stata* et *R* de mise en œuvre de la méthode des offsets (Appendice 4, sur le site web *Tools for Demographic Estimation*). D'autres exemples de calculs sont disponibles sur la feuille Excel *Method of offsets* du fichier Excel disponible en ligne.

Étant donné que les flux ont été observés directement en 1976, nous avons plusieurs moyens d'évaluer la capacité de la méthode des offsets à reproduire les données. Une méthode simple consiste à examiner visuellement les rapports des composantes multiplicatives d'interaction, comme le montre le tableau 37.4. Une autre méthode s'appuie sur les tests inférentiels et les mesures d'information des procédures log-linéaires. Il s'agit de tester l'hypothèse que la structure des flux migratoires, c'est-à-dire les paramètres d'interaction, n'a pas changé entre 1973 et 1976. Dans l'exemple du tableau 37.10, le G^2 correspondant est égal à 5 914 ($df = 25$), et l'hypothèse que les données auxiliaires représentent le même régime de migration que celui des données observées doit donc être rejetée. Une dernière méthode proposée ici utilise R^2 et l'*EAMP* pour évaluer l'adéquation entre flux attendus et flux observés. Ces deux paramètres, présentés au cadre A du tableau 37.10, valent respectivement 0,97 et 8,36. Avec les rapports du tableau 37.4, ces valeurs indiquent que cette application de la méthode des offsets donne une série d'estimations des flux migratoires de 1976 qui peut être tout à fait satisfaisante.

L'importance accordée aux indices de qualité de l'ajustement dépend de la qualité des flux observés introduits dans l'application de la méthode des offsets. Si la méthode doit être utile dans des cas pratiques, elle doit être applicable quand les flux interrégionaux n'ont pas été observés directement. En l'absence de données sur les flux, la méthode nécessite toujours des estimations préalables des totaux marginaux. D'ailleurs, si on applique la méthode comme le montre l'Appendice 4 (disponible sur le site web *Tools for Demographic Estimation*), on doit introduire des estimations

TABLEAU 37.10 Flux migratoires interrégionaux aux Pays-Bas (1976), calculés par la méthode des *offsets* à partir des totaux marginaux de 1976 et des flux de 1973

| Cadre A. Calcul par la méthode des substitutions | | | | | | | |
|---|--------|---------|---------|--------|----------------|---------|---------|
| Destination | | | | | | | |
| Origine | 1 | 2 | 3 | 4 | 5 | 6 | Total |
| 1 | 12 344 | 13 769 | 6 890 | 12 199 | 10 361 | 11 518 | 67 081 |
| 2 | 13 329 | 34 695 | 12 195 | 17 445 | 22 522 | 24 353 | 124 539 |
| 3 | 9 728 | 15 711 | 28 330 | 8 883 | 15 881 | 30 553 | 109 087 |
| 4 | 11 281 | 16 107 | 7 011 | 11 216 | 11 764 | 13 187 | 70 566 |
| 5 | 12 609 | 25 486 | 16 570 | 12 828 | 17 770 | 21 760 | 107 023 |
| 6 | 18 116 | 35 786 | 53 984 | 22 110 | 27 058 | 22 535 | 179 589 |
| TOTAL | 77 408 | 141 553 | 124 980 | 84 682 | 105 356 | 123 906 | 657 885 |
| $R^2 = 0,966$ | | | | | $MAPE = 8,364$ | | |
| Cadre B. Composantes multiplicatives avec calcul des effets des facteurs sur la base du total général | | | | | | | |
| Destination | | | | | | | |
| Origine | 1 | 2 | 3 | 4 | 5 | 6 | Total |
| 1 | 1,564 | 0,954 | 0,541 | 1,413 | 0,964 | 0,912 | 0,102 |
| 2 | 0,910 | 1,295 | 0,515 | 1,088 | 1,129 | 1,038 | 0,189 |
| 3 | 0,758 | 0,669 | 1,367 | 0,633 | 0,909 | 1,487 | 0,166 |
| 4 | 1,359 | 1,061 | 0,523 | 1,235 | 1,041 | 0,992 | 0,107 |
| 5 | 1,001 | 1,107 | 0,815 | 0,931 | 1,037 | 1,080 | 0,163 |
| 6 | 0,857 | 0,926 | 1,582 | 0,956 | 0,941 | 0,666 | 0,273 |
| TOTAL | 0,118 | 0,215 | 0,190 | 0,129 | 0,160 | 0,188 | 657 885 |

initiales des flux interrégionaux. Les estimations provisoires des totaux marginaux doivent donc être distribuées dans les cellules intérieures des lignes et colonnes correspondantes de la matrice des flux. Le cadre A du tableau 37.11 présente un scénario classique, quoiqu'il utilise toujours les totaux marginaux des données néerlandaises de 1976, qui sont des données observées. Une solution simple consiste à distribuer les flux conformément au modèle d'indépendance, c'est-à-dire $\hat{n}_{ij} = (T)(O_i)(D_j)$, ce qui donne les estimations initiales des flux qui figurent au cadre B du tableau 37.11.

Tant que les flux interrégionaux initiaux ont pour sommes les totaux marginaux, les flux attendus ne sont pas affectés par la méthode de répartition des flux dans les cellules intérieures de la matrice, car, en fin de compte, les flux attendus auront été calculés à partir des données auxiliaires par la méthode des *offsets*, en utilisant l'algorithme itératif

d'ajustement proportionnel (Agresti 1990 ; Deming et Stephan 1940). En d'autres termes, les estimations initiales des flux néerlandais de 1976, introduites dans le modèle log-linéaire avec *offsets*, pourraient être les valeurs des cellules internes du cadre B du tableau 37.1, ou celles du cadre B du tableau 37.11. L'une et l'autre séries d'estimations initiales donnera les flux attendus qui figurent au cadre A du tableau 37.10.

D'autre part, il est important de noter que les tests inférentiels et les mesures d'information qui accompagnent la méthode des *offsets* doivent être interprétés en référence aux estimations initiales des flux. Par exemple, si les flux initiaux sont ceux du cadre B du tableau 37.11, les paramètres χ^2 et G^2 correspondants testent l'hypothèse que les valeurs attendues sont distribuées d'une manière compatible avec le modèle d'indépendance.

TABLEAU 37.11 Inputs de la méthode des offsets en l'absence de flux observés

| Cadre A. Totaux marginaux pré-estimés, Pays-Bas, 1976 | | | | | | | |
|---|--------|---------|---------|--------|---------|---------|---------|
| Destination | | | | | | | |
| Origine | 1 | 2 | 3 | 4 | 5 | 6 | Total |
| 1 | | | | | | | 67 081 |
| 2 | | | | | | | 124 539 |
| 3 | | | | | | | 109 087 |
| 4 | | | | | | | 70 566 |
| 5 | | | | | | | 107 023 |
| 6 | | | | | | | 179 589 |
| TOTAL | 77 408 | 141 553 | 124 980 | 84 682 | 105 356 | 123 906 | 657 885 |
| Cadre B. Distribution des estimations initiales des flux selon le modèle d'indépendance | | | | | | | |
| Destination | | | | | | | |
| Origine | 1 | 2 | 3 | 4 | 5 | 6 | Total |
| 1 | 7 893 | 14 433 | 12 744 | 8 635 | 10 743 | 12 634 | 67 081 |
| 2 | 14 654 | 26 796 | 23 659 | 16 030 | 19 944 | 23 456 | 124 539 |
| 3 | 12 835 | 23 472 | 20 724 | 14 042 | 17 470 | 20 545 | 109 087 |
| 4 | 8 303 | 15 183 | 13 406 | 9 083 | 11 301 | 13 290 | 70 566 |
| 5 | 12 593 | 23 027 | 20 331 | 13 776 | 17 139 | 20 157 | 107 023 |
| 6 | 21 131 | 38 641 | 34 117 | 23 116 | 28 760 | 33 824 | 179 589 |
| TOTAL | 77 408 | 141 553 | 124 980 | 84 682 | 105 356 | 123 906 | 657 885 |

Il n'est pas difficile d'aménager la méthode des substitutions pour l'appliquer au calcul des valeurs attendues d'une matrice limitée aux migrants. Les programmes *SPSS*, *Stata* et *R* ne demandent que de légères modifications, détaillées dans les commentaires de l'Appendice 4 (en ligne sur le site web *Tools for Demographic Estimation*). Un exemple est

traité dans la feuille de calcul *Method of offsets, migrants only* du fichier Excel disponible en ligne (voir le site web). Il se base sur les flux observés aux États-Unis entre 1985 et 1990, pour estimer rétrospectivement les flux de migrants de la période 1975-1980 publiés par Rogers, Willekens, Little *et al.* (2002).

RÉFÉRENCES

Agresti A. 1990. *Categorical Data Analysis*. New York: Wiley.

Agresti A. 2002. *Categorical Data Analysis*. New York: Wiley-Interscience.

Agresti A. 2007. *An Introduction to Categorical Data Analysis*. Hoboken, NJ: Wiley-Interscience.

Agresti A and B Finlay. 2009. *Statistical Methods for the Social Sciences*. Upper Saddle River, NJ: Pearson Prentice Hall.

Alonso W. 1986. *Systemic and log-linear models: From here to there, then to now, and this to that*. Discussion paper 86-10. Cambridge, MA: Harvard University, Center for Population Studies.

Birch MW. 1963. "Maximum likelihood in three-way contingency tables", *Journal of the Royal Statistical Society Series B-Statistical Methodology* **25**(1):220-233.

Deming WE and FF Stephan. 1940. "On a least squares adjustment of a sampled frequency table when the expected marginal totals are known", *Annals of Mathematical Statistics* **11**(4):427-444. doi: <http://dx.doi.org/10.1214/aoms/1177731829>

Knoke D and PJ Burke. 1980. *Log-linear Models*. Beverly Hills, CA: Sage Publications.

Mueser P. 1989. "The spatial structure of migration: An analysis of flows between states in the USA over three decades", *Regional Studies* **23**(3):185-200. doi: <http://dx.doi.org/10.1080/00343408912331345412>

Nair PS. 1985. "Estimation of period-specific gross migration flows from limited data: Bi-proportional adjustment approach", *Demography* **22**(1):133-142. doi: <http://dx.doi.org/10.2307/2060992>

Powers DA and Y Xie. 2008. *Statistical Methods for Categorical Data Analysis*. Bingley, UK: Emerald.

- Raftery AE. 1986. "Choosing models for cross-classifications", *American Sociological Review* **51**(1):145-146. doi: <http://dx.doi.org/10.2307/2095483>
- Raftery AE. 1995. "Bayesian model selection in social research", *Sociological Methodology* **25**(1):111-163. doi: <http://dx.doi.org/10.2307/271063>
- Raymer J. 2007. "The estimation of international migration flows: A general technique focused on the origin-destination association structure", *Environment and Planning A* **39**(4):985-995. doi: <http://dx.doi.org/10.1068/a38264>
- Raymer J, A Bonaguidi and A Valentini. 2006. "Describing and projecting the age and spatial structures of interregional migration in Italy", *Population, Space and Place* **12**(5):371-388. doi: <http://dx.doi.org/10.1002/psp.414>
- Raymer J and A Rogers. 2007. "Using age and spacial flow structures in the indirect estimation of migration streams", *Demography* **44**(2):199-223. doi: <http://dx.doi.org/10.1353/dem.2007.0016>
- Rees P and FJ Willekens. 1986. "Data and accounts," in Rogers, A and FJ Willekens (eds). *Migration and Settlement: A Multiregional Comparative Study*. Dordrecht: D. Reidel, pp. 19-58.
- Rogers A, JS Little and J Raymer. 2010. *The Indirect Estimation of Migration: Methods for Dealing with Irregular, Inadequate, and Missing Data*. Dordrecht: Springer.
- Rogers A, F Willekens, JS Little and J Raymer. 2002. "Describing migration spatial stucture", *Papers in Regional Science* **81**(1):29-48.
- Rogers A, FJ Willekens and J Raymer. 2003. "Imposing age and spatial structures on inadequate migration-flow datasets", *The Professional Geographer* **55**(1):56-69.
- Snickars F and JW Weibull. 1977. "A minimum information principle: Theory and practice", *Regional Science and Urban Economics* **7**(1-2):137-168. doi: [http://dx.doi.org/10.1016/0166-0462\(77\)90021-7](http://dx.doi.org/10.1016/0166-0462(77)90021-7)
- Willekens F. 1983. "Log-linear modeling of spatial interaction", *Papers of the Regional Science Association* **52**:187-205. doi: <http://dx.doi.org/10.1007/BF01944102>

