Université catholique de Louvain
Centre de recherche en démographie et sociétés

Bruno Masquelier

Estimation de la mortalité adulte en Afrique subsaharienne à partir de la survie des proches

Apports de la microsimulation

Thèse présentée en vue de l'obtention du grade de Docteur en Sciences Sociales (Démographie)

M^me^	Catherine Gourbin	UCL	(Présidente du jury)
M.	Gilles Pison	INED	(Rapporteur)
M.	Bruno Schoumaker	UCL	(Rapporteur)
M.	Guillaume Wunsch	UCL	(Rapporteur)
M.	Dominique Tabutin	UCL	(Directeur)

Merci à Dominique Tabutin pour sa disponibilité, son encadrement motivant, et ses lectures minutieuses. Merci à Bruno Schoumaker, Guillaume Wunsch et Gilles Pison, pour le temps qu'ils ont consacré à la lecture de cette thèse et pour leurs commentaires encourageants. Merci également à Catherine Gourbin, présidente de ce jury masculin.

Cette thèse doit beaucoup à l'échange d'idées entretenu avec des collègues démographes. Je remercie tout spécialement Patrick Gerland, de la Division de la Population des Nations Unies, qui m'a fourni énormément d'informations. Merci à Carl Mason, de l'Université de Berkeley, qui a mis SOCSIM à ma disposition et modifié le code pour répondre à certaines de mes demandes. Je tiens à remercier également Georges Reniers, Abdramane Soura, Christophe Vandeschrick et François Pelletier.

Les données de recensements utilisées ici l'ont été grâce au projet IPUMS et aux offices statistiques suivants : Afrique du Sud, Kenya, Mali, Ouganda, Rwanda, Tanzanie. Cette recherche n'aurait pas été possible sans les données d'enquêtes EDS, financées notamment par l'USAID, et distribuées gratuitement en ligne. Merci à Bridgette James, de MEASURE DHS, qui m'a fourni les données de l'enquête EDS spécifique menée en Tanzanie en 1995 (Sumve Survey on Adult and Childhood mortality).

Cette recherche s'inscrit dans le cadre d'un mandat d'Aspirant du Fonds National de la Recherche Scientifique (FNRS).

Merci à Sophie et à Gaspard, chez qui j'ai trouvé au quotidien la motivation nécessaire pour mener cette recherche.

Louvain-la-Neuve, le 10 novembre 2010.

Table des matières

Table des matières **iv**

Introduction générale **1**

1 Données disponibles sur la survie des parents et des frères et sœurs **13**
- 1.1 Collecte des données relative à la survie des parents 22
  - 1.1.1 Enquêtes nationales . . . 22
  - 1.1.2 Recensements . . . 26
- 1.2 Collecte des données relative à la survie des frères et sœurs . . . 29
  - 1.2.1 Enquêtes nationales . . . 32
- Conclusion . . . 42

2 La microsimulation : un outil pour exploiter les données imparfaites **43**
- 2.1 Avantages de la microsimulation . . . 46
- 2.2 SOCSIM en bref . . . 50
  - 2.2.1 Le fichier de base . . . 51
  - 2.2.2 Les populations de départ . . . 52
  - 2.2.3 Les fichiers de taux . . . 52
  - 2.2.4 Les populations finales . . . 55
- 2.3 Paramètres démographiques introduits . . . 55
  - 2.3.1 Taille et structures des populations . . . 59
  - 2.3.2 Niveaux et structures de la mortalité "sans sida" . . . 64
  - 2.3.3 Niveaux et structures de la fécondité . . . 69
- 2.4 Intégration du VIH-sida . . . 75
  - 2.4.1 Prévalence du VIH . . . 78
  - 2.4.2 Distributions par sexe et par âge des risques d'infection . 81
  - 2.4.3 Durée de survie des adultes infectés . . . 87
  - 2.4.4 Transmission verticale . . . 93
  - 2.4.5 Réduction de la fécondité chez les femmes séropositives . 94
- Conclusion . . . 96

3 Estimer la mortalité adulte à partir de la survie des parents **99**
- 3.1 Comment formaliser la probabilité d'être orphelin ? . 103
- 3.2 Convertir les proportions de survivants en probabilités de survie . . . 110
  - 3.2.1 Coefficients introduits par Timaeus (1992) . . . 111
  - 3.2.2 Une application pratique . . . 117

3.2.3 Autres modèles et paramètres utilisés 122
3.2.4 Modèles construits pour la mortalité masculine 129
3.2.5 Applications aux microsimulations 133
3.3 Applications aux cas de fécondité et de mortalité non stables 136
3.3.1 Fécondité non stable 136
3.3.2 Mortalité non stable 142
3.3.3 Recours aux cohortes synthétiques 149
3.4 Survie des parents aux âges adultes 154
3.4.1 Survie des parents à 20 ans 156
3.4.2 Survie des parents au mariage 159
3.5 Biais de sélection 163
3.5.1 Surmortalité des orphelins 163
3.5.2 Association entre la survie des mères et des enfants . . . 170
3.5.3 Association entre mortalité des mères et fécondité 174
3.6 Application aux populations simulées pour l'Afrique subsaharienne 176
3.6.1 Comparaison des proportions d'orphelins issues des simulations avec celles issues des enquêtes EDS 176
3.6.2 Application de la méthode des orphelins aux microsimulations 183
Conclusion 194

4 Orphelins et crise du VIH-sida **195**
4.1 Perturbations induites par le VIH-sida 198
4.1.1 Biais de sélection associés au VIH-sida 198
4.1.2 Schémas de mortalité atypiques 203
4.2 Calcul du nombre d'orphelins de mère et de père au niveau macro 207
4.3 Retour sur les probabilités de décès 217
4.3.1 Calcul de nouveaux coefficients 218
4.3.2 Application aux microsimulations 219
4.3.3 Application aux enquêtes EDS 221
Conclusion 222

5 Estimer la mortalité adulte à partir de la survie des frères et sœurs **223**
5.1 Estimation des taux de mortalité en biographique (approche directe) 225
5.1.1 Régression de Poisson 226
5.1.2 Modélisation des tendances 229
5.2 Biais de sélection 234
5.2.1 Démonstration mathématique de Trussell et Rodriguez (1990) 234
5.2.2 Applications aux micro-simulations 238
5.2.3 Association entre mortalité et taille des fratries 243
5.2.4 Correction de Gakidou et King (2006) 245
5.3 Exploitation indirecte des informations portant sur la survie des frères et sœurs 267
5.3.1 Formalisation de la proportion de frères et sœurs survivants 268

5.3.2 Conversion des proportions de survivants en probabilités de survie . . . 269
5.3.3 Application aux données d'enquêtes . . . 274

6 Qualité des données et principaux biais 277
6.1 Données portant sur les parents . . . 278
6.1.1 Biais d'adoption . . . 278
6.1.2 Données manquantes sur la survie des parents . . . 282
6.1.3 Erreurs sur les âges . . . 288
6.2 Données portant sur la fratrie . . . 290
6.2.1 Données manquantes . . . 291
6.2.2 Sous-déclaration différentielle par sexe . . . 295
6.2.3 Sous-déclaration des frères et sœurs nés avant les enquêtés 297
6.2.4 Taille des fratries et descendance finale . . . 298
6.2.5 Erreurs faites sur les âges des frères et sœurs . . . 303
6.2.6 Sous-déclaration des décès dans le passé lointain . . . 304
Conclusion . . . 308

7 Estimations nationales basées sur la survie des proches 311
7.1 Afrique de l'Ouest . . . 320
7.1.1 Mali . . . 323
7.1.2 Niger . . . 324
7.1.3 Sénégal . . . 325
7.2 Afrique Centrale . . . 326
7.2.1 Cameroun . . . 327
7.2.2 Tchad . . . 328
7.3 Afrique de l'Est . . . 329
7.3.1 Madagascar . . . 331
7.3.2 Ouganda . . . 332
7.3.3 Rwanda . . . 333
7.4 Afrique australe . . . 334
7.4.1 Afrique du Sud . . . 336
7.4.2 Malawi . . . 337
7.4.3 Zambie . . . 338
7.4.4 Zimbabwe . . . 339
Conclusion . . . 340

Conclusion générale 341

Bibliographie 356

A Annexes 378
A.1 Annexes au chapitre 1 . . . 380
A.1.1 Classification UNECA des pays africains . . . 380
A.1.2 Enquêtes nationales disposant d'informations sur la survie des parents . . . 381
A.1.3 Enquêtes nationales disposant d'informations sur la survie des frères et sœurs . . . 385
A.1.4 Notes sur l'appariement des frères et sœurs adultes dans les fichiers EDS . . . 388
A.2 Annexes au chapitre 2 . . . 394

A.2.1 Comparaison des différentes options disponibles pour modéliser la fécondité en micro-simulation 394
A.2.2 Notes sur le modèle de Lee et Carter (1992) 406
A.2.3 Comparaison des taux de croissance équivalents aux populations stables (paramètres de 1950-1955) avec les taux de croissance estimés en 1950-1955 par les Nations Unies 407
A.2.4 Notes sur les fonctions splines 408
A.2.5 Distribution des décès infantiles utilisée dans les microsimulations 409
A.2.6 Notes sur le calcul des taux de fécondité par rang à partir des rapports de taux 410
A.2.7 Notes sur l'intégration de la nuptialité dans les microsimulations africaines 412
A.2.8 Notes sur les distributions "phase-type" 424
A.3 Annexes au chapitre 3 426
A.3.1 Calcul des probabilités ${}_np_{25}$ à partir de la survie des mères : 6 recensements est-africains 426
A.3.2 Coefficients utilisés pour les méthodes indirectes 428
A.3.3 Modèle polynomial de Brass 431
A.3.4 Notes sur l'estimation indirecte de l'âge moyen à la naissance des enfants des pères 432
A.4 Annexes au chapitre 4 435
A.4.1 Formules de Timaeus et Nunn (1997) 435
A.5 Annexes au chapitre 5 438
A.5.1 Démonstration de Trussell et Rodriguez (1990) 438
A.5.2 Méthode des sœurs développée par Graham et al. (1989) . 439
A.5.3 Coefficients utilisés pour dater les estimations indirectes basées sur la survie des frères et sœurs 441
A.6 Annexes au chapitre 6 442
A.6.1 Données manquantes sur la survie des parents 442
A.6.2 Notes sur le traitement des réponses manquantes relatives à la survie des parents 443
A.6.3 Données manquantes sur les frères et sœurs 447
A.7 Annexes au chapitre 7 457
A.7.1 Bénin 457
A.7.2 Burkina Faso 458
A.7.3 Centrafique 459
A.7.4 Comores 460
A.7.5 Congo 461
A.7.6 Congo RDC 462
A.7.7 Côte d'Ivoire 463
A.7.8 Ethiopie 464
A.7.9 Gabon 465
A.7.10 Ghana 466
A.7.11 Guinée 467
A.7.12 Kenya 468
A.7.13 Lesotho 469
A.7.14 Liberia 470
A.7.15 Mozambique 471
A.7.16 Namibie 472

A.7.17 Nigeria 473
A.7.18 Sierra Leone 474
A.7.19 Swaziland 475
A.7.20 Tanzanie 476
A.7.21 Togo 477

Acronymes **478**

Introduction générale

Les trois dernières décennies ont marqué un tournant important en Afrique subsaharienne : la région a dû faire face à une double crise, économique et sanitaire.

D'une part, le sous-continent connaît une *récession économique* importante, surtout depuis 1985. Cette récession a de nombreuses manifestations, et elle s'inscrit dans un faisceau de causes déjà bien présentes avant même les indépendances [1] (Bayart 2006, Brunel 2003). C'est surtout à partir de 1985 que la situation s'est globalement détériorée, en raison de ce que Coussy et Vallin (1995) ont appelé une *crise du mode d'insertion des pays africains dans l'espace international*. Les flux extérieurs ont connu une évolution défavorable : les parts de marché et les prix des produits de rente ont diminué, et la charge de la dette s'est considérablement alourdie. De nombreux pays ont enregistré une désindustrialisation ainsi qu'un développement rapide et non maîtrisé du secteur informel. La baisse de l'aide internationale est venue pénaliser davantage ce continent "sous perfusion", tout comme les plans d'ajustements structurels. Ces plans d'ajustements, basés sur l'austérité et la viabilité économique de court terme, se sont traduits par une réduction importante des budgets consacrés à l'éducation et à la santé. Les pays africains sont devenus de plus en plus dépendants des aléas extérieurs, comme l'ont rappelé les récentes émeutes de la faim qui ont suivi l'envolée des prix des denrées alimentaires (en 2008 dans de nombreux pays africains [2] ainsi qu'en 2010 au Mozambique).

Cette crise économique s'est combiné à partir de la fin des années 1980 à *la propagation du VIH*, que le PNUD a appelé « la plus grande récession du développement humain ». L'Afrique subsaharienne concentre 67% des 33 millions de personnes infectées dans le monde (UNAIDS 2008). Bien que tous les pays africains n'aient pas été affectés avec la même intensité, il s'agit d'un choc majeur du point de vue démographique, sanitaire et socio-économique. C'est sans conteste l'un des événements contemporains les plus importants survenus dans le champ de la démographie et de la santé.

Toutes les composantes de la dynamique des populations sont concernées. C'est bien entendu les risques de décès qui sont les plus perturbés : le VIH-sida provoque de fortes reprises de mortalité (tant chez les enfants que chez les adultes), il redéfinit les schémas de mortalité par âge, il bouleverse

1. Faible productivité agricole, faible développement du salariat, inadaptation des investissements dans les infrastructures, fuite des cerveaux et des capitaux, captation des ressources par la violence et la corruption, etc.

2. Nigeria, Cameroun, Côte d'Ivoire, Mozambique, Sénégal, Burkina Faso, etc.

l'importance relative des différentes causes de décès. Globalement, le sida est même devenu la première cause de mortalité des adultes à l'échelle du sous-continent (WHO 2008b). Il a également relancé la tuberculose, dont l'incidence diminuait pourtant dans de nombreux pays sub-sahariens depuis les années 1960 (Cauthen et al. 2002). La tuberculose est aujourd'hui classée septième dans l'ordre des causes de charge morbide en Afrique, et elle est devenue la première des complications secondaires à l'infection du VIH.[3] Avec près d'un tiers de la population subsaharienne infectée de façon latente par la tuberculose (Dye et al. 1999), on estime son taux d'incidence à 363 pour 100 000, contre 139 en moyenne mondiale (WHO 2008a). La région australe est particulièrement touchée ; tous ses pays présentent des taux d'incidence de plus de 500 pour 100 000[4].

L'infection au VIH entraîne également une diminution de la fécondité, surtout dans la période symptomatique, et cette baisse reste significative même après contrôle de l'utilisation de la contraception (Lewis et al. 2004). La combinaison des fortes mortalités précoces et de cette baisse de la fécondité se traduit par un déficit de naissances, ainsi que par des bouleversements dans les structures par âge. Dans de nombreux pays, les mouvements migratoires ont également été redéfinis, notamment avec un afflux des personnes infectées vers le village (ce que Clark et al. (2007) appellent le phénomène *returning home to die*). Enfin, la crise du VIH-sida a perturbé l'organisation des ménages et les systèmes de soutien informel, notamment en laissant de nombreux orphelins exposés à des risques de mortalité plus élevés (Zaba et al. 2005), à la pauvreté ainsi qu'à la déscolarisation (Monasch et Boerma 2004). On estime qu'autour de l'an 2000, un peu moins d'un enfant sur dix a perdu au moins un des ses deux parents avant l'âge de 15 ans. En conséquence, dans plus d'une dizaine de pays, un cinquième des ménages avec enfants contient au moins un orphelin.

Au delà de la récession économique, de l'expansion du VIH ou de la reprise de la tuberculose, de nombreux autres indicateurs dessinent les contours d'une crise sanitaire sans précédent. Son intensité est bien sûr variable d'une région à l'autre, d'un pays à l'autre, et au sein même des différents pays. Mais globalement, l'impression qui se dégage de ce tableau contrasté est celle d'*une situation sanitaire moins bonne aujourd'hui que celle qui prévalait il y a quinze ans* (Jamison et al. 2006). Par exemple, certaines endémies connaissent une recrudescence, comme la trypanosomiase ou le choléra. A lui seul, ce maintien du choléra, apparu au début des années 1970, est d'ailleurs très emblématique de l'inadéquation des systèmes de soins africains, tant son traitement est simple et bon marché. Sur les quinze dernières années, le choléra a quasiment disparu de l'Amérique latine, où 42% des cas y étaient encore rapportés durant la décennie 1990 (Gaffga et al. 2007). En Afrique subsaharienne, sa létalité a diminué, mais son incidence n'a pas significativement baissé.

3. L'immunodépression provoquée par le VIH augmente les risques de développer une tuberculose active (soit par réinfection, soit par réactivation). La propagation de la tuberculose et celle du VIH s'alimentent donc mutuellement.

4. En raison d'une proportion importante des séropositifs parmi les personnes atteintes de la tuberculose (autour de 30%), sa létalité est également nettement plus élevée en Afrique que dans les autres régions du monde : elle s'élève à 34% contre 23% en moyenne mondiale (WHO 2008a).

Emblématique également, le maintien, et même la hausse de la mortalité palustre, en raison notamment de la diffusion de la chloroquinorésistance (Korenromp et al. 2003, Pison 2007). Mais nous pourrions également citer les faibles avancées en matière de couverture vaccinale durant les dernières années, les niveaux élévés de mortalité maternelle, ou encore l'importance de la malnutrition ...

Conséquence du maintien des maladies de pauvreté et de l'avancement timide du continent sur la voie de la transition sanitaire, l'Afrique subsaharienne fait aujourd'hui face à un "double fardeau" (Duthé et Pison 2008, Reniers, Masquelier, et Gerland forthcoming). Cela est évident en matière de malnutrition ; d'un côté, celle-ci concerne encore 212 millions de personnes en Afrique (FAO 2008), avec une prévalence estimée à 30% environ, soit un niveau bien supérieur à celui de l'Asie du Sud (20%). Dans le même temps, on constate l'apparition de surcharge pondérale dans les grandes villes ; Dakar, Kinshasa, Accra notamment (Salem et al. 2000). Ces cas de "double fardeau" contredisent l'hypothèse trop simpliste d'un passage linéaire des maladies transmissibles aux maladies intransmissibles (Omran 1971). Le développement des maladies chroniques coexiste avec la résurgence de certaines maladies infectieuses. Plus globalement, le sous-développement sanitaire de l'Afrique subsaharienne contredit l'idée, qui prévalait il y a quelques décennies, d'une convergence des niveaux d'espérance de vie à l'échelle mondiale (Moser et al. 2005, Wilson 2001). L'écart se creuse de plus en plus avec les autres régions en développement.

Cette détérioration de la situation sanitaire se traduit d'abord par un *tassement des espérances de vie*. Celles-ci connaissaient pourtant, au niveau régional, une hausse continue depuis les indépendances. Par endroits, la tendance s'est même inversée et la mortalité a connu de fortes reprises. A partir des années 1980, les progrès en matière d'espérance de vie sont stoppés en Afrique de l'Ouest, une régression s'amorce en Afrique centrale et en Afrique de l'Est, et l'Afrique australe voit son espérance de vie chuter de près de quinze ans en l'espace de dix ans (UN 2009c).

Ces hausses de mortalité concernent à la fois les enfants et les adultes, mais les adultes sont de loin les plus affectés. *Selon les estimations des Nations Unies, la probabilité de décéder entre 15 et 60 ans (${}_{45}q_{15}$), a connu une hausse sur les quinze dernières années dans toutes les régions de l'Afrique subsaharienne, sauf en Afrique de l'Ouest.* Ceci s'observe sur le graphe 1 (p.4) [5].

Ce tableau alarmiste n'est toutefois pas univoque. Si la crise est particulièrement prononcée dans les pays où la prévalence du VIH est élevée (Zimbabwe, Zambie, etc.), une série de pays enregistrent des progrès (Madagascar, ou l'Éthiopie, par exemple, comme nous le verrons dans le chapitre 7). A un autre niveau, si la mortalité *adulte* augmente dans de nombreux pays, certains d'entre eux voient leur mortalité *infantile et*

5. Source : Reniers, Masquelier, et Gerland (forthcoming), à partir des données des Nations Unies (UN 2009c), en ce compris des estimations non publiées. La classification régionale retenue, que nous reprenons pour l'ensemble des estimations fournies dans ce document, est celle de l'UNECA (United Nations Economic Commission for Africa). Elle est détaillée en annexe A.1.1, page 380.

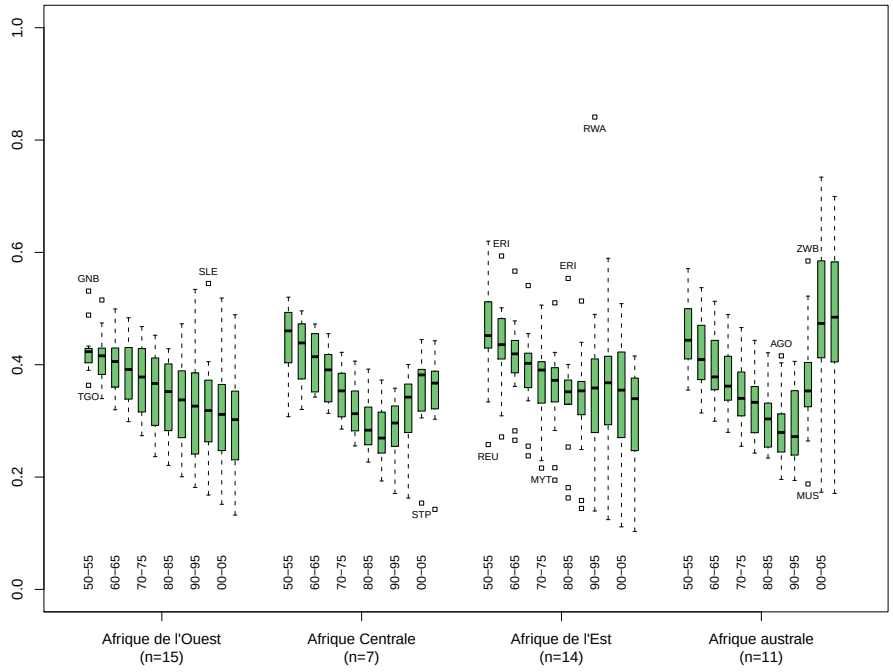

FIGURE 1 – *Probabilités de décès entre 15 et 60 ans - Sc : Reniers, Masquelier, et Gerland (forthcoming)*

LECTURE : Pour chaque région et chaque intervalle quinquennal, une barre de couleur comprend la moitié des pays. Un quart des pays présentent des valeurs de ${}_{45}q_{15}$ plus élevées que la limite supérieure de la barre de couleur (c-à-d. le dernier quartile), un autre quart présente des valeurs plus faibles que la limitée inférieure de cette barre. La médiane est notée par un trait noir. Les droites en pointillés s'étendent jusqu'aux points qui ne sont pas séparés de la boîte de plus de 1.5 fois l'intervalle interquartile. Les pays situés au-delà sont notés par un carré.

juvénile augmenter moins vite, voire, dans certains cas, continuer à baisser. Par exemple, selon les statistiques onusiennes, l'Ouganda a perdu 5 ans d'espérance de vie entre 1978 et 1998, tout en voyant sa mortalité infantile baisser de 105 à 90 $^0/_{00}$ sur la même période (UN 2009c).

On assiste donc à une diversification croissante des niveaux et des schémas par âge de la mortalité. Cette diversité des situations nationales, déjà apparente au niveau des espérances de vie, s'accroît encore lorsque la mortalité est analysée uniquement aux âges adultes. Aujourd'hui, sur le continent africain, les risques de mortalité entre 15 et 60 ans peuvent varier du simple au double, voire parfois du simple au triple. Cette grande diversité s'accompagne de profondes inégalités au sein même des différents pays, selon des clivages ethniques, géographiques et socio-économiques.

Dans ce contexte, il est crucial de disposer d'estimations précises sur les niveaux, les tendances et les différentiels de la mortalité. Pour les enfants, cela est relativement faisable, grâce à la multiplication des enquêtes rétrospectives de fécondité. Ces dernières permettent de collecter des informations sur l'histoire génésique des femmes en âge de reproduction et sur la survie de leurs enfants. Grâce à elles notamment, la plupart des pays de la région disposent de nombreuses sources de données et des tendances peuvent être reconstituées (Hill et Amouzou 2006). Par contre, ce travail est nettement plus difficile pour la mortalité adulte.

En plus d'un manque d'intérêt scientifique et politique, le manque de données et la faible qualité des informations qui sont collectées constituent les principaux freins à l'établissement de statistiques fiables (*cf.* chapitres 1 et 6). L'Organisation mondiale de la santé estime que la couverture de l'enregistrement des décès est inférieure à 25% dans quasiment tous les pays de l'Afrique subsaharienne. Seuls quelques pays font exception : le Zimbabwe [6], l'Afrique du Sud [7], et certaines îles (comme l'Île Maurice) [8].

Durant les décennies 1970-1980, un grand nombre de méthodes indirectes ont donc été développées pour pallier à ce manque de données (ou à leur mauvaise qualité). En ce qui concerne les adultes, les principales méthodes reposent sur :

1. la *comparaison des structures par âge entre deux recensements* : les rapports de survie calculés pour la période intercensitaire pour chaque cohorte sont convertis en indices de mortalité (Preston et Bennett 1983). Les estimations obtenues s'avèrent extrêmement sensibles aux variations dans la complétude de l'enregistrement, aux erreurs d'âges systématiques, ainsi qu'aux migrations. Cette méthode est donc quasiment inutilisable en Afrique.
2. l'*ajustement d'enregistrements incomplets des décès* : il s'agit des *méthodes de distribution des décès*. Elles consistent à rehausser les estimations de mortalité selon l'importance de la sous-déclaration des décès. D'une

6. Entre 25 et 50% des décès enregistrés en 2002.
7. Entre 75 et 90% décès en 2005.
8. WHO Statistical Information Systems (WHOSIS), http ://www.who.int/whosis/en/

part, la *generalized growth balance method* (GGB) compare ces décès aux taux de croissance inter-censitaires par âge (Hill 1987). D'autre part, la *synthetic extinct generation method* (SEG), développée par Bennett et Horiuchi (1984), repose sur une formalisation mathématique complexe, mais sur une idée assez simple ; si l'enregistrement des décès et de la population étaient parfaits, le nombre d'individus à un âge donné observé dans un recensement devrait être égal au nombre de décès qui vont avoir lieu au sein de cette cohorte [9].

3. enfin, troisième grande catégorie de méthodes, l'*exploitation de données relatives à la survie des proches* ; c'est-à-dire la survie des parents (Brass et Hill 1973), des frères et soeurs (Hill et Trussell 1977), et des premiers conjoints (Hill 1977). C'est ce troisième ensemble de méthodes qui fait l'objet de cette thèse.

Il faut rappeler ici que les deux principales organisations internationales en charge de la production d'estimations démographiques (les Nations Unies et l'OMS) ne se basent pas directement sur ces trois grandes catégories de méthodes d'estimation. En effet, elles s'appuient toutes deux principalement sur la *survie des enfants* pour estimer celle des adultes. Dans la pratique, pour la majorité des pays, des schémas standards de mortalité par âge ou des modèles relationnels (de type logits [10]) sont utilisés en combinaison avec des niveaux de mortalité observés chez les enfants pour inférer des tables de mortalité.

Par exemple, pour quatre pays africains sur cinq, la Division de la Population des Nations Unies combine la mortalité des enfants aux tables de Princeton, développées par Coale et Demeny (1983) pour estimer un niveau de mortalité "sans sida" (nous reviendrons sur cette démarche dans l'encadré 4, p.66). Dans la plupart des cas, les risques de décès sont ensuite revus à la hausse à partir de la prévalence et de l'incidence fournies par l'ONUSIDA (ainsi que des modèles relatifs aux probabilités d'infection par âge et sexe et aux durées de survie des personnes infectées). Les estimations de mortalité sont également ajustées en cas de périodes d'instabilité et de guerres (Burundi, Congo RDC, Liberia, Rwanda, Somalie, etc.). Nous qualifierons les niveaux de mortalité obtenus de la sorte d'estimations issues de la *modélisation*, pour bien les distinguer de celles qui seraient issues de l'*observation* des risques de décès à partir des trois grandes catégories de méthodes présentées plus haut.

Cette distinction est importante, car il faut bien comprendre que pour de nombreux pays, la démarche de l'OMS et de l'ONU consiste à ne faire intervenir qu'*à titre de validation de leurs estimations* (issues de la modélisation) les niveaux de mortalité plus "empiriques" dégagés des données relatives aux adultes. Généralement, plusieurs révisions et itérations sont nécessaires pour atteindre une relative cohérence, à la fois entre les différentes composantes de la dynamique démographique d'un pays, ainsi qu'entre les estimations issues de ce travail de modélisation et celles issues

9. En supposant que les décès à un âge donné vont croître au même rythme que la population du même âge, on peut alors estimer ces décès à venir à partir des décès survenus dans le passé. Ceci procure une estimation du sous-enregistrement des décès. Voir également page 15, ainsi que les tableaux de synthèse présentés en conclusion (page 347).

10. Voir encadré 6, page 112.

de l'observation (UN 2005) [11].

A côté des estimations de l'ONU et de l'OMS issues de la modélisation, il est donc crucial de disposer d'autres estimations, issues de l'*observation des risques de décès*. Partir de la survie des enfants - et plus largement modéliser la mortalité adulte - pose en effet problème à plusieurs égards :

- Le premier problème, qui est bien connu, est que les tables types de mortalité utilisées doivent nécessairement se baser des données fiables. Elles reposent donc surtout sur l'expérience historique des pays occidentaux.

- Deuxièmement, il est de plus en plus délicat d'intégrer l'*impact du VIH-sida* à des projections "sans sida", à partir d'estimations de prévalence (traduites en incidences) et de modèles types d'infection et de survie par âge. Dans certains pays, les schémas de mortalité ont été si bousculés - et plus généralement, la dynamique démographique est devenue si étroitement liée à l'évolution de l'épidémie - que le fait même de faire des projections contrefactuelles (sans sida) a perdu de son sens (Heuveline 2003). D'autant qu'il est délicat de distinguer la mortalité "due au sida" et la mortalité "sans sida" à partir des données empiriques dont on dispose aujourd'hui. Cela est vrai pour la mortalité des adulte, mais également pour celle des enfants. Or celle-ci sert à fixer le niveau de mortalité "sans sida" de départ. De façon plus déterminante, dans les contextes affectés par le VIH-sida, la mortalité des adultes a connu des évolutions partiellement déconnectées de celles que connaît la mortalité des enfants, et il devient donc risqué d'estimer la première à partir de la seconde. La mortalité des enfants et celle des adultes sont *moins corrélées* dans la pratique que ne le suppose une application systématique aux différentes situations africaines d'un ou deux schémas de mortalité théoriques.

- La troisième raison qui plaide en faveur d'une plus grande utilisation des estimations empiriques tient au fait que *les estimations relatives au VIH-sida sont elles-mêmes entourées d'incertitudes*. C'est le cas des estimations de prévalence ou d'incidence, ainsi que d'autres paramètres clés comme la distribution par sexe des infections. Ces incertitudes vont se répercuter sur les estimations de mortalité. Par exemple, les prévalences fournies par l'ONUSIDA ont été sensiblement revues à la baisse ces dernières années. Ceci a conduit à une ré-estimation assez conséquente des niveaux de mortalité. Autre exemple, la survie après l'infection reste estimée à partir de quelques données prospectives, et modélisée au moyen d'hypothèses parfois restrictives, comme c'est le cas quand on suppose que la survie après l'infection est invariante par âge (*cf.* chapitre 2).

11. Il faut toutefois préciser que plus récemment, et uniquement pour quelques pays d'Afrique de l'Ouest, les Nations Unies se sont directement basées sur des niveaux de mortalité adulte (${}_{45}q_{15}$) issus de la survie des parents (7 pays) et celle des frères et soeurs (6 pays). Il s'agit sans doute d'une tendance qui va se prolonger.

– Enfin, même si l'on admet que les schémas standards de mortalité peuvent au moins approcher la mortalité réelle, *la sélection de la famille de tables standards* pose question. Par exemple, la majorité des estimations nationales élaborées par les Nations Unies (UN 2009c) sont construites à partir de deux standards : le modèle Nord et le modèle Sud de Coale et Demeny (1983)[12]. Le premier est caractérisé par une mortalité adulte élevée par rapport à la mortalité infanto-juvénile. Il couvre une majorité des pays d'Afrique centrale et d'Afrique de l'Est. Le modèle Sud est caractérisé plutôt par une moindre mortalité adulte (par rapport à la mortalité infanto-juvénile). Il concerne essentiellement les pays d'Afrique de l'Ouest. Ce choix d'un standard fait l'objet d'un examen au cas par cas par l'ONU. Mais il faut relever ici qu'*il a un impact important sur les estimations* ; la répartition par âge de la mortalité varie sensiblement d'un modèle à l'autre. Ainsi, chez les femmes, pour une espérance de vie fixée à 50 ans, le modèle Nord prévoit un quotient ${}_5q_0$ de 0.183. Un même niveau de mortalité infanto-juvénile renvoie à une espérance de vie de 54 ans, dans le cadre du modèle Sud. On a donc là un même niveau de mortalité infanto-juvénile pour deux niveaux d'espérances de vie séparés de 4 ans. Et cela correspond à deux niveaux très différents de mortalité adulte : un quotient ${}_{45}q_{15}$ de 0.337 avec le modèle Nord, contre un quotient inférieur de 30% avec le modèle Sud (0.243).

Pour ces différentes raisons, *le recours à des tables types, combinées à la mortalité des enfants, est susceptible de produire des estimations incertaines au niveau national, même en tenant compte de spécificités régionales*. C'est d'ailleurs pourquoi les estimations de l'ONU doivent être révisées par itérations. Lorsque les écarts avec les estimations empiriques[13] sont trop importants, de nouvelles estimations peuvent être dégagées en révisant le niveau de mortalité "sans sida" (c'est-à-dire essentiellement la mortalité des enfants qui sert de point de départ). C'est aussi pourquoi les estimations de l'ONU et de l'OMS peuvent parfois sensiblement différer pour un même pays, alors même que les deux institutions partent des mêmes niveaux de mortalité des enfants (*cf.* chapitre 7).

Tout ceci invite donc à réexaminer le potentiel des données dont on dispose sur la survie des adultes, et à revisiter les méthodes d'analyse. Un regain d'attention leur a d'ailleurs récemment été accordé dans la littérature. Par exemple, les techniques basées sur l'ajustement d'enregistrements incomplets des décès ont été analysées à l'aide de simulations afin d'en identifier les principales limites (Hill et al. 2009, Dorrington et al. 2006c, Murray et al. 2010). Les données sur la survie des frères et soeurs, longtemps considérées comme peu fiables, ont également été davantage exploitées (Gakidou et al. 2004, Obermeyer et al. 2010, Rajaratnam et al. 2010, Reniers et al. forthcoming). Ces données sont collectées grâce à

12. Le modèle Ouest est également utilisé pour quelques pays d'Afrique australe et d'Afrique de l'Est.

13. Qu'elles soient déduites de la survie des proches, des décès observés dans les ménages ou d'états civils partiels.

quelques questions simples, adaptées à différents types de collecte (recensements, enquêtes, etc.) :

- vos parents sont-ils toujours en vie ?
- Combien de frères et soeurs avez-vous eu ?
- Combien sont décédés ? A quel âge ?
- Quels âges ont les survivants ?

A partir de là, les décès qui ont eu lieu dans le cercle familial peuvent être "récupérés". En dépit du regain d'intérêt que nous venons de noter, ces données restent relativement peu exploitées. Les travaux qui les exploitent de façon systématique restent rares, et ils sont généralement menés par un petit nombre de chercheurs. Quant aux rapports des enquêtes qui génèrent ces données, ils se limitent souvent à une brève description des niveaux de mortalité observés à partir des frères et soeurs, et ils ne font que mentionner la proportion d'orphelins. De la même façon, les rapports de recensements mentionnent parfois les proportions d'orphelins, mais souvent sans distinction par âge et sans présenter les niveaux de mortalité correspondants.

Cette faible exploitation des données sur la survie des proches s'explique pour partie par une certaine *réserve à l'égard de la qualité des données*. Les chercheurs qui les ont analysé en détail ont rapidement invité à la prudence (Stanton et al. 2000, Gakidou et al. 2004). Plusieurs sources d'erreurs sont régulièrement mentionnées : c'est le cas par exemple du "biais d'adoption", qui revient à omettre le décès des parents lorsque l'information collectée se réfère à tort aux parents adoptifs plutôt qu'aux parents biologiques. Quand elles concernent les frères et soeurs, les informations semblent également entachées d'erreurs sur les âges, de sous-déclaration des décès, d'erreurs sur la datation des décès, etc.

En plus de ces problèmes de qualité des données, de nombreux *biais* affectent les méthodes. Ils sont relevés dans la littérature, mais rarement quantifiés et il est donc difficile d'en mesurer les effets sur les estimations. Notamment, le développement de la pandémie du VIH-sida a mis à mal de nombreuses hypothèses, pour au moins trois raisons (Heuveline 2003) : (1) le VIH-sida a modifié radicalement les schémas par âge de mortalité, (2) il a sensiblement augmenté les risques de décès, (3) il a renforcé l'interdépendance des risques de mortalité au sein d'un même réseau de proches parents.

Face à ces différents enjeux, cette thèse poursuit donc trois objectifs :

1. revisiter les techniques d'estimation de la mortalité des adultes à partir de la survie des frères et soeurs et des parents, en combinant les enquêtes démographiques et de santé (EDS), des données de recensements ainsi que des microsimulations,
2. présenter les principaux biais de sélection et les problèmes de qualité qui caractérisent ces données,
3. fournir des estimations empiriques récentes sur les risques de décès entre 15 et 60 ans.

Sur le plan méthodologique, cette recherche s'appuie sur trois types de ressources :

1. la *démographie formelle*, et notamment la démographie des populations stables : elle est à la base du développement des méthodes indirectes analysées ici. Elle n'est pas toujours très intuitive, elle peut poser des problèmes quand il s'agit d'intégrer des expressions mathématiques complexes, mais elle n'est pas très demandeuse en terme de ressources informatiques. Nous l'utiliserons notamment pour détailler les méthodes indirectes et examiner certaines relations théoriques entre des proportions de proches survivants et des probabilités de survie.
2. la *macro-simulation* : il s'agit essentiellement de méthodes matricielles, avec différentes cellules et un ensemble de transitions entre différents états. Nous l'utiliserons pour examiner comment les proportions d'orphelins vont être affectées par le VIH-sida.
3. la *microsimulation*, détaillée dans le chapitre 2 : il s'agit de soumettre une population fictive à une série de risques prédéfinis (décès, mariage, grossesse, etc.). Nous utiliserons le programme SOCSIM, développé à l'Université de Berkeley, en retenant des populations et des taux proches de ceux observés dans les populations africaines. Ce programme présente l'avantage d'enregistrer les liens de parentés entre les individus simulés. Par rapport à la démographie formelle et à la macro-simulation, la microsimulation est beaucoup plus intuitive que les deux approches précédentes, mais elle est également beaucoup plus demandeuse en ressources informatiques (et en temps). Nous lui préfèrerons donc les deux premières approches quand cela est possible. Par contre, la microsimulation devient très intéressante quand il s'agit d'aller au delà des moyennes pour analyser la dispersion des phénomènes (qui a encore beaucoup de frères et soeurs survivants et qui en a peu ?), ou quand il s'agit de travailler sur des sous-groupes de populations (quelle est la proportion de frères survivants et âgés de 15 à 49 ans parmi les femmes aujourd'hui âgées de 15 à 49 ans ?).

L'ensemble du travail se décline en sept chapitres.

1. Le premier vise à fournir un aperçu de la *disponibilité des données* par pays sur la survie des frères et soeurs ainsi que sur la survie des parents.
2. Le deuxième présente une série de *microsimulations* construites pour être représentatives de la diversité démographique du continent africain. Ces microsimulations serviront ensuite à l'examen des différents biais de sélection.
3. Le troisième chapitre vise à revisiter les différentes *techniques d'estimation* basées sur la survie des parents. Il confronte ces méthodes indirectes aux résultats des microsimulations. Il accorde une attention particulière aux hypothèses sur lesquelles reposent la méthode des orphelins.
4. Le quatrième chapitre se penche en particulier sur la question des *biais introduits par le VIH-sida* dans l'estimation de la mortalité à partir de la survie des parents.

5. Le cinquième chapitre est consacré aux méthodes basées sur la survie des *frères et soeurs*. Là encore, les différentes approches sont confrontées aux résultats de la microsimulation.
6. Le sixième chapitre est centré sur la *qualité des données et leur ajustement*. Où se situent les principaux problèmes dans les déclarations faites par les proches ? Comment peut-on y remédier ? Quels sont leurs conséquences sur les estimations ?
7. Le septième chapitre présente un large panorama des *niveaux de mortalité adulte* qu'il est possible d'obtenir en combinant les données sur la survie des parents et celle des frères et soeurs, pour de nombreux pays d'Afrique subsaharienne. Ces niveaux sont confrontés aux estimations issues de la modélisation, telles qu'elles sont produites par les Nations Unies et l'OMS.

A de nombreuses reprises au cours de cette recherche, nous nous sommes appuyés sur les travaux de Ian Timaeus, qui a consacré à la mortalité adulte des pays en développement plus de 20 ans de recherches. Ce sont ses travaux qui servent aujourd'hui de référence pour la méthode des orphelins (Timaeus 1986; 1991b;a; 1992). Il est également un des seuls auteurs à avoir cherché à adapter la méthode des orphelins aux contextes perturbés par le VIH-sida (Timaeus et Nunn 1997). Il a par ailleurs développé une méthode indirecte permettant d'exploiter les données sur les frères et soeurs (Timaeus et al. 2001), ainsi qu'un modèle de régression pour l'analyse de ces données sous forme biographique (Timaeus et Jasseh 2004). Il faut encore ajouter à cela la méthode utilisée par l'UNICEF pour estimer le nombre d'orphelins dus au sida (Grassly et Timaeus 2005), ainsi que les nombreux articles proposant une vue d'ensemble de la mortalité adulte africaine (Timaeus 1999, Bradshaw et Timaeus 2003). Ses travaux balisent donc le champ parcouru ici, et nous espérons que nos recherches en constituent des prolongements.

1 Données disponibles sur la survie des parents et des frères et sœurs

Sommaire

1.1 Collecte des données relative à la survie des parents . 22
1.2 Collecte des données relative à la survie des frères et sœurs 29
Conclusion 42

Pour l'estimation de la mortalité adulte, l'Afrique au sud du Sahara reste un véritable défi. Elle compte 50 pays, plus de 850 millions d'individus âgés de plus de quinze ans - soit un adulte sur dix dans le monde - mais très peu de statistiques fiables quant à leurs risques de décès. Combien d'années ces adultes peuvent-ils espérer vivre ? Quels sont les principaux risques sanitaires auxquels ils doivent faire face ? Comment leurs risques de décès évoluent-ils par âge ? Dans quelle mesure les adultes ont-ils été inégalement exposés aux récentes reprises de mortalité ? A ces questions, les démographes ne peuvent répondre qu'au moyen d'estimations entourées de larges incertitudes. Et sur bien des points, la crise sanitaire qui frappe aujourd'hui la région a prolongé ces zones d'incertitudes.

Comme on l'a relevé en introduction, la rareté des données empiriques constitue l'un des principaux freins à l'estimation des risques de décès adultes. Afin de présenter un panorama des données disponibles, nous avons rassemblé différentes sources dans le tableau 1.1, repris en page 14. Celui-ci détaille le nombre de pays pour lesquels des données empiriques sont *exploitables* à l'échelle nationale, selon la source de données et la période. Il s'agit donc d'opérations de collecte où (1) les questions nécessaires ont été posées, et (2) soit les données sont accessibles, soit les tabulations nécessaires ont été faites et publiées [1].

1. Nous ne nous limitons pas ici aux données sur les proches pour fournir un aperçu plus général des différentes sources disponibles.

	1950-54	1955-59	1960-64	1965-69	1970-74	1975-79	1980-84	1985-89	1990-94	1995-99	2000-04	2005-10[a]
Etat civil[b]	3	7	8	9	9	2	4	5	5	3	3	3
Recensements[c]												
Décès (12 mois)	0	0	0	2	4	9	10	9	9	9	14	6
Survie des parents	0	0	0	2	5	4	6	11	8	8	11	4
Enquêtes[d] (EDS, MICS, LSMS, etc.)												
Survie des parents	0	0	0	0	0	3	1	2	2	9	23	16
Survie des frères/sœurs	0	0	0	0	0	0	0	0	8	18	29	25
Décès (12 mois)	0	0	2	0	1	3	1	0	0	0	1	0
Au moins une source	3	7	8	13	16	15	19	21	19	20	38	20

[a] Le nombre de données exploitables pour la période 2005-2010 est sous-estimé ici, car nous ne disposons pas d'informations sur les recensements les plus récents (Burundi 2008, Cameroun 2005, Tchad 2009, Congo 2007, Gabon 2003, Kenya 2009, Liberia 2008, Mali 2009, Mauritanie 2000, Nigeria 2006, Soudan 2008, Djibouti 2009).

[b] Les données d'état civil sont disponibles dans différentes éditions du *Demographic Yearbook* (UN 2009a; 2008a; 2007; 1997), ainsi que dans Feeney (2001) et Dorrington et al. (2001). Mathers et al. (2005) fournissent également une liste des années pour lesquelles l'OMS dispose de données. Nous avons exclu Mayotte, les Seychelles, Saint Hélène et Sao Tomé. L'Île Maurice, la Réunion et Djibouti sont comptabilisés.

[c] Les dates de recensements sont issues de la Division statistique des Nations Unies (http ://unstats.un.org) et du US Census Bureau (http ://www.census.gov). Les questionnaires de recensements ont été consultés sur le site de l'ACAP (African Census Analysis Project) (http ://www.acap.upenn.edu/), ou sur les sites des institutions nationales. D'autres informations sur les recensements sont issues de Tabutin (1984), de Timaeus (1999) et des Nations Unies (UN 2006). Nous n'avons pas d'informations sur le recensement mené aux Comores en 2003. Par ailleurs, il semble que le recensement ougandais de 1980 ait intégré ces questions mais que les tabulations ne soient pas disponibles.

[d] Les questionnaires et les données d'enquêtes ont été téléchargés à partir des sites suivants : http ://www.measuredhs.com pour les données EDS, http ://www.childinfo.org/ pour les enquêtes MICS, http ://www.worldbank.org/lsms/ pour les enquêtes LSMS, et http ://www.who.int/healthinfo/survey/en/index.html pour les enquêtes WHS.

TABLE 1.1 – *Nombre de pays sub-sahariens pour lesquels au moins une source de données sur la survie des adultes est disponible - selon la période et le type de sources (n*

47 pays interviennent dans ce tableau, d'où les plus petits territoires insulaires sont exclus [2].

En 2000-2004, *pour 8 pays africains sur 10*, au moins une source de données sur les adultes était disponible, et ces données étaient issues essentiellement de questions sur la survie des frères et sœurs ou sur la survie des parents intégrées aux enquêtes nationales. Sur la période plus ancienne, de 1980 à 2000, on dispose de données par intervalle quinquennal pour *un peu plus de 4 pays sur 10* [3]. Ce sont alors davantage les recensements, et plus tardivement, la survie des frères et sœurs, d'où proviennent les estimations. Plus loin encore dans le passé, entre 1965 et 1980, des données sur la survie des adultes sont disponibles au niveau national seulement *pour un pays sur trois environ*, et principalement par le biais des recensements et des systèmes d'état civil. Cette proportion chute ensuite lorsque l'on remonte avant 1965.

Mentionnons rapidement les avantages et inconvénients des différents types de données, pour relever l'intérêt des données sur les proches.

Décès déclarés à l'état civil et dans les recensements

Les données d'état civil sont sans conteste les plus appropriées pour analyser la mortalité. Mais en Afrique subsaharienne, quand elles ne sont pas tout simplement inexistantes, elles sont rarement utilisables (Mathers et al. 2005). Dans la majorité des pays où un système est mis en place, les données sont encore encodées manuellement, et dans certains cas, elles sont collectées au niveau *régional* mais elles ne sont pas analysées au niveau *national* (p.ex. au Mozambique, au Malawi, etc. [4]). En prenant une large fenêtre temporelle, de 1950 à 2010, on ne comptabilise qu'une douzaine de pays ayant fourni au moins une année d'observation ; notamment l'Angola, la Guinée Equatoriale, le Kenya, le Mozambique, le Zimbabwe et l'Afrique du Sud. Parfois les systèmes d'état civil fonctionnent suffisamment dans les grandes villes pour que quelques estimations puissent en être dégagées - par exemple à Abidjan (Garenne et al. 1995).

Au niveau national, à l'exception des territoires insulaires qui parviennent à maintenir une bonne couverture de l'enregistrement des décès (Maurice, Cap Vert, Seychelles), seuls deux systèmes d'état civil ont permis de dégager des estimations *récentes* et *plausibles* ; le Zimbabwe [5] et l'Afrique du Sud [6]. Mais dans les deux cas, les décès sont encore loin d'être enregistrés de façon complète. Au Zimbabwe, Feeney (2001) estime que la couverture de l'enregistrement des décès adultes déclarés à l'état

2. Sont exclus : Sao Tomé et Principe, les Seychelles, Mayotte, Saint Helène. Sont retenus parmi les 47 pays : Djibouti, Maurice, la Réunion.

3. Mais ce n'est pas nécessairement les mêmes pays qui fournissent les données d'une période à l'autre.

4. Voir UN (2008c), Report on the United Nations Workshop on the improvement of civil registration and vital statistics in SADC region, 1-5 december 2008, Blantyre, Malawi.

5. Pour lequel des estimations sont disponibles pour 1982, 1986, 1990-1992 et 1995 - (Feeney 2001).

6. Les données sont disponibles de 1985 à 2000 (Dorrington et al. 2001), ainsi que pour 2005 (UN 2008a) et 2006 (UN 2009a).

civil s'établissait sur la période 1982-1992 à 44% pour les femmes et à 63% pour les hommes. L'enregistrement a ensuite connu une nette amélioration, jusqu'à atteindre en 1995 84.9% pour les décès masculins et 59.1% pour les décès féminins. Depuis 1995, l'état civil fonctionne encore mais les données ne sont plus disponibles. En Afrique du Sud, sur la période 1989-2000, la complétude de l'enregistrement des décès adultes a varié de 54% en 1990 à 89% en 2000 (Dorrington et al. 2001)[7].

Ces données d'état civil sont à rapprocher des *décès déclarées dans les recensements* (ceux survenus dans les douze derniers mois), car leur exploitation est similaire. Dans les deux cas, les méthodes de distribution des décès sont utilisées pour corriger les estimations brutes.

Pour les 47 pays considérés ici, nous avons comptabilisé 71 recensements africains ayant intégré une question sur les décès survenus dans les 12 derniers mois, et 106 recensements ne l'ayant pas intégré[8]. A partir des recensements dont nous avons pu établir si cette question avait été posée ou non, on peut relever une *collecte de plus en plus systématique de ces informations*. En effet, cette question était posée dans un peu plus d'un recensement sur trois dans les années 1970, et cette proportion s'éleve à 0.53 dans les années 1990, puis à 0.76 dans les années 2000. Sur la période allant de 1985 à 2010, cette proportion a également varié par région : elle est la plus faible en Afrique de l'Est (0.39), et la plus élevée en Afrique australe (0.7), l'Afrique centrale et l'Afrique de l'Ouest occupant des positions intermédiaires (respectivement 0.44 et 0.67).

A partir des décès déclarés dans les douze derniers mois précédant les recensements, ou de ceux enregistrés à partir de l'état civil, la *generalized growth balance method* (GGB) et la *synthetic extinct generation method* (SEG), peuvent être utilisées. Ces deux méthodes sont appliquées sur base (1) de la distribution par âge de la population, généralement observée dans deux recensements, et (2) de la distribution des décès par âge (idéalement ceux survenus entre ces deux recensements).

Les deux méthodes procurent une estimation de la complétude de l'enregistrement des décès relative aux deux recensements[9]. Elles reposent sur des hypothèses assez lourdes : une population fermée aux migrations[10], une complétude de l'enregistrement de la population et des décès invariante par âge, une absence d'erreurs d'âges systématiques, une constance des taux de natalité et de mortalité dans le temps[11]. Ces deux méthodes ont été longtemps dépréciées car les décès sont généralement sensiblement

7. Des estimations récentes sont disponibles pour 2005 et 2006 (UN 2008a; 2009a).

8. Pour 15 recensements (les plus récents, menés entre 2003 aux Comores et 2009 au Mali) nous n'avons pas pu obtenir cette information. Au Burundi en 1990 et au Nigeria en 1991, les questions semblent avoir été posées mais les résultats ne sont pas accessibles, soit parce que les données n'ont pas été encodées, analysées ou publiées.

9. La CGB procure également une estimation de la complétude de l'enregistrement de la population lors du second recensement (par rapport au premier). Voir les tableaux de synthèse présentés en conclusion, page 347, et les travaux de Hill et al. (2009) ou Murray et al. (2010).

10. Sauf si des données sont disponibles pour procéder à des ajustements.

11. La méthode SEG suppose en outre un enregistrement des décès et de la population invariant dans le temps.

sous-déclarés dans les recensements, ce qui accroît l'incertitude autour des estimations.

Récemment, Hill et al. (2009) ont évalué par simulation la performance de ces méthodes, et ils ont estimé que leur combinaison procurait des résultats satisfaisants [12]. Cette approche a été appliqué notamment en Afrique australe (Dorrington et al. 2006c). Par contre, lorsque les hypothèses de base de ces méthodes ne sont pas rencontrées, notamment en matière de migration, les marges d'erreurs peuvent s'avérer importantes. En combinant des simulations et des données empiriques, Murray et al. (2010) estiment que *ces marges d'incertitudes autour des taux d'enregistrement des décès estimés s'élèvent au moins à 20%* [13]. Il va sans dire que ces résultats sont peu encourageants, quand on connaît l'ampleur de la sous-déclaration des décès. Par exemple, Timaeus (1993) estime que seuls la moitié des décès adultes ont été enregistrés lors du recensement malien de 1987, et qu'au Congo-Brazza en 1984, seul un décès masculin sur trois et un décès féminin sur cinq ont été enregistrés. Autre exemple, si l'on reprend les estimations de Feeney (2001) pour le Zimbabwe, et que l'on entoure d'une marge d'incertitude de 20% la couverture de l'enregistrement des décès estimée à 44% pour les femmes et à 63% pour les hommes (sur la période 1982-92), il devient très délicat d'utiliser les données corrigées pour fournir autre chose qu'un ordre de grandeur.

De plus, plusieurs éléments laissent penser que le VIH-sida peut poser des problèmes à l'application de ces méthodes. D'une part, dans les contextes fortement perturbés par le VIH, le décès d'un adulte est souvent suivi de la dissolution et de la reformation des ménages, surtout s'il s'agissait du chef de ménage (Urassa et al. 2001). Ceci va accentuer le sous-enregistrement des décès. Par ailleurs, les fortes probabilités de transmission du VIH entre partenaires impliquent que le décès de l'un des deux partenaires est susceptible d'être suivi du décès du second, ce qui réduit les chances de voir ces deux décès renseignés dans les enquêtes ou les recensements. Ces deux sources de sous-déclaration des décès vont introduire des distorsions au niveau des schémas par âge de la mortalité estimés, car les décès des jeunes adultes seront disproportionnellement

12. "In populations not greatly affected by migration, our results suggest that an optimal strategy would be to apply GGB to estimate census coverage change, adjust for it and then apply SEG ; in populations affected by migration, applying both GGB and SEG, fitting both to the age range 30+ to 65+, and averaging the results appears best." Hill et al. (2009), p.235.

13. L'analyse de Murray et al. (2010) pointe d'autres éléments intéressants d'un point de vue méthodologique. Ils rappellent notamment que dans la pratique, seul un nombre limité de groupes d'âges sont retenus pour estimer le degré de complétude de l'enregistrement des décès. Or la *sélection* de ces groupes d'âge a un impact significatif sur la performance des méthodes SEG et GGB ainsi que sur leur combinaison. Ils discutent également des méthodes classiques utilisées pour juger de la performance des méthodes. Par exemple, lorsque les taux de mortalité observés directement sont présentés en abscisse et les taux corrigés par le biais de la GGB en ordonnées, plus les points s'inscrivent sur une ligne droite, plus la méthode est susceptible de procurer des résultats satisfaisants. De même, le degré de complétude par âge obtenu par la SEG devrait s'inscrire sur une ligne droite de pente nulle si les hypothèses de base de la méthode sont rencontrées. Or Murray et al. (2010) montrent que ces techniques ne procurent en fait qu'une vague idée de la performance des méthodes. Ils indiquent également que les erreurs d'âge systématiques jouent un rôle important.

sous-déclarés. Enfin, troisième difficulté, l'estimation de la complétude de l'enregistrement des décès suppose que les tendances n'ont pas un impact majeur sur la distribution par âge (Timaeus et Nunn 1997). Or c'est peu le cas dans les pays où de fortes hausses de mortalité sont enregistrées aux premiers âges adultes .

Dans les cas où les questions sur les décès survenus dans le ménage ne sont pas prévues, il faut alors recourir à d'autres techniques pour exploiter les données de recensements. Il est par exemple possible d'estimer la mortalité à partir de l'évolution de la structure de la population entre deux recensements (Preston et Bennett 1983). Mais là encore, les biais sont nombreux ; ils peuvent être causés par la migration, les erreurs d'âge systématiques, ou les changements dans la couverture des recensements. Cette méthode est généralement considérée comme inapplicable aux données de recensements africaines (Timaeus 1993).

Données sanitaires ou administratives

Les informations issues des établissements sanitaires sont fragmentaires et biaisées, notamment parce qu'une grande proportion des décès surviennent au domicile. En outre, les données sont peu représentatives, car les services de soins concentrent des populations et des pathologies spécifiques. Certaines estimations sont également dégagées des morgues et des cimetières, par exemple à Addis Abeba (Araya et al. 2004). Mais il y a là aussi des effets de sélection importants. Des registres paroissiaux ont par ailleurs été utilisés en Namibie (Notkola et al. 2004), mais il s'agit d'analyses locales difficilement généralisables.

Sites d'observation démographique

D'autres données sont générées par les sites d'observation démographique. Ces sites permettent de détailler plus précisément les schémas par âge, d'analyser les tendances, et de spécifier parfois les causes de décès grâce aux autopsies verbales (INDEPTH 2002; 2004). L'Afrique subsaharienne en compte aujourd'hui plus d'une vingtaine, qui sont pour la plupart intégrés dans le réseau INDEPTH [14]. Ils couvrent surtout l'Afrique australe et l'Afrique de l'Est, et dans une moindre mesure l'Afrique de l'Ouest [15].

Le coût financier et humain d'une collecte régulière, la complexité des données générées et les problèmes de confidentialité empêchent encore une large diffusion de ces données et limitent leur exploitation. La représentativité des sites pose également question. Par exemple, la sélection de ces sites repose sur un examen des infrastructures disponibles. Par ailleurs, ces sites servent souvent de terrains d'expérimentations épidémiologiques. En dépit de cela, les sites d'observation sont sans doute la

14. International Network of field sites with continuous Demographic Evaluation of Populations and Their Health in developing countries : http ://indepth-network.org/

15. L'Afrique de l'Ouest compte toutefois les seuls sites africains mis en place autour de 1985 et toujours en activité (au Sénégal, en Gambie, et en Guinée-Bissau).

meilleure source d'information dont on dispose aujourd'hui pour la mortalité adulte africaine. Le fait que les données du réseau INDEPTH soient de plus en plus analysées de façon comparative constitue également une avancée majeure, notamment pour dégager des standards de mortalité "africains" (INDEPTH 2004).

Survie des parents et des frères et sœurs

A ce rapide tour d'horizon des données, il faut encore ajouter deux types d'informations, qui font l'objet de cette recherche ; celles qui sont relatives à la survie des parents, et celles qui concernent la survie des frères et sœurs.

D'une part, nous avons comptabilisé *59 recensements africains menés entre 1969 et 2010 au cours desquels les individus ont été interrogés sur la survie de leur parents* (au moins de leur mère). 82 recensements menés sur la même période n'ont pas prévu cette question [16]. Ces informations sont également collectées dans différents programmes d'enquêtes rétrospectives, le plus souvent en interrogeant les enfants de moins de 15 ans. *Plus de 140 enquêtes à l'échelle nationale ont intégré ces questions*, et nous y reviendrons plus loin.

D'autre part, certains programmes d'enquêtes ont prévu des modules de questionnaires relatifs à la *survie des frères et sœurs*. Ces données ont surtout été collectées durant les deux dernières décennies, lorsque la crise du VIH-sida, ainsi que la persistance d'une forte mortalité maternelle ont relancé l'intérêt sur la santé sexuelle et reproductive. Ces deux facteurs ont permis d'orienter un peu plus les regards vers la santé des adultes. Interroger les enquêtés sur la survie de leurs sœurs, ce fut au départ une tentative expérimentale menée à petite échelle au sein d'un site d'observation continue en Gambie (Graham et al. 1989). C'est une approche qui s'est ensuite rapidement imposée comme le mode de collecte le plus fréquent pour enregistrer les décès maternels, et qui s'est même étendu à l'analyse de la mortalité adulte. Récemment, la collecte des données sur la mortalité maternelle a été quelque peu repensée et on tente aujourd'hui d'intégrer dans les questionnaires de recensements des informations sur la survie des femmes enceintes (Stanton et al. 2001) (nous y reviendrons dans l'encadré 2, p.30).

Pour être tout à fait complet, il faudrait également présenter ici les données collectées sur le *décès du premier conjoint*. L'idée derrière la collecte de telles données consiste à convertir les proportions de premiers conjoints survivants en probabilités de survie (Hill et Trussell 1977). Mais nous n'aborderons pas cette méthode dans notre travail de thèse, pour les raisons évoqués dans l'encadré 1, repris en page suivante.

Il faut préciser également que ces données sur les proches nous renseignent surtout sur les niveaux de la mortalité "*toutes causes*". Dans le cas de la survie des parents, nous ignorons tout à fait les causes de décès.

16. Pour 14 recensements récents, nous n'avons pas encore cette information. En Ouganda, un recensement semble avoir intégré cette question mais les données ne sont pas disponibles (1980).

Dans le cas de la survie des frères et sœurs, les enquêtés spécifient généralement s'il s'agit de décès maternels, et occasionnellement s'il s'agit de décès accidentels ou liés à la violence. Les enquêtes mondiales de santé de l'OMS prévoient par ailleurs quelques questions de précisions sur les circonstances du décès des frères et sœurs.

Précisons enfin que les données sur la survie des proches peuvent prendre deux formes ;

- soit les âges à l'enquête et les âges au décès sont collectés, ce qui permet de calculer la durée d'exposition et autorise une exploitation *directe* des données,
- soit les questions portent uniquement sur la survie ou non des proches, sans précisions sur les âges ou les dates, et il faut alors opter pour des méthodes *indirectes*.

Les données sur les frères et sœurs sont souvent collectées de façon à ce que les deux types d'exploitation soient possibles. Par contre, il est rare de pouvoir exploiter directement les données sur les parents, car leurs âges ou les dates de décès sont rarement collectés.

Encadré 1 Données collectées sur la survie du premier conjoint

Théoriquement, la méthode basée sur la survie du premier conjoint est intéressante pour plusieurs raisons (Blacker 1984, Hill 2005) ;

1. les estimations obtenues font référence à une période récente,
2. il ne peut y avoir de double compte,
3. il n'y a pas d'"effet d'adoption" comparable à ce qui est observé pour la survie des parents (au sens où les informations recueillies font référence aux parents adoptifs et pas aux parents biologiques),
4. les premiers mariages se concentrent sur un intervalle d'âge restreint, ce qui réduit les chances de s'écarter sensiblement des modèles théoriques qui sont à la base de la méthode.

Cette méthode n'a toutefois pas connu d'application large et systématique, et l'expérience manque pour se prononcer sur la plausibilité des estimations obtenues dans les contextes sub-sahariens. A notre connaissance, six enquêtes nationales menées en Afrique subsaharienne ont intégré une question sur la survie du premier conjoint ; notamment deux enquêtes WFS [17], et l'enquête EDS menée au Burundi en 1987 (Makinson 1993) [18].

17. Il s'agit de l'enquête soudanaise menée en 1978-79, et de l'enquête conduite au Lesotho en 1977. Au Lesotho, les estimations féminines sont plausibles mais les estimations masculines ne cadrent pas avec les autres estimations disponibles. Au Soudan, c'est plutôt l'inverse ; la survie masculine est plausible mais les tendances de la mortalité féminine sont improbables (Timaeus 1987). Notons que ces deux enquêtes souffrent d'un biais de sélection, puisque les répondants n'ont été interrogés explicitement sur la survie de leur premier époux que lorsqu'ils avaient déjà été mariés *plus d'une fois*. Cela signifie que la survie des conjoints des individus divorcés ou séparés n'ayant été mariés qu'une fois est inconnue.

18. Les niveaux de mortalité issus de cette enquête burundaise à partir des premiers conjoints sont jugées beaucoup trop bas par Makinson (1993). Notons que Blacker (1977) cite également une enquête démographique menée en Tanzanie en 1973 et deux enquêtes nationales menées au Kenya en 1977 et 1983.

Plus encore que les autres données sur les proches, on soupçonne que ces informations soient de mauvaise qualité, surtout parce qu'il s'agit de collecter des données sur les *premiers* conjoints. Or les hommes qui répondent à la place de leurs femmes pourraient avoir tendance à sous-déclarer les mariages précédents de celles-ci (Blacker et Mukiza-Gapere 1988). Dans les ménages où les femmes se sont remariées, l'enquêteur pourrait également passer les questions relatives à la survie du conjoint, et considérer que le partenaire aperçu est le premier conjoint (UN 1983b). De même, il semble que certains enquêtés, et tout particulièrement les plus jeunes, aient tendance à dissimuler le décès de leur partenaire (Malaker 1986). A toutes ces erreurs de déclaration s'ajoute le fait que les enquêtés sont susceptibles de préférer faire part d'un décès plutôt qu'un divorce lorsque celui-ci est peu reconnu socialement. Ce "biais de désirabilité" explique sans doute pourquoi on observe généralement que les veufs se déclarent célibataires et les divorcés veufs (Timaeus et Nunn 1997).

Ces différentes imprécisions seront d'autant plus fréquentes que le divorce et le remariage seront fréquents eux aussi. Elles seront également plus répandues si le mariage s'inscrit dans une temporalité longue. En effet, plus sa temporalité est étendue, plus l'enquêté dispose d'opportunités pour redéfinir son union dans des termes socialement acceptables (Timaeus et Nunn 1997). Or, dans les contextes africains, le mariage relève plus d'un processus que d'un événement bien délimité dans le temps. A ces imprécisions dans les informations collectées, il faut encore ajouter un biais supplémentaire dans les pays perturbés par le VIH-sida, du fait qu'il y a un risque accru de corrélation entre les probabilités de survie des deux conjoints.

Pour toutes ces raisons, la méthode du premier conjoint a été rarement utilisée. Quand elle l'a été, les résultats obtenus ont été peu encourageants. Dans le cadre de cette recherche, nous ne nous intéressons donc uniquement à la survie des parents ainsi qu'à celle des frères et sœurs, en laissant de côté la survie des premiers conjoints.

1.1 Collecte des données relative à la survie des parents

La "méthode des orphelins" consiste à estimer les risques de décès des adultes en interrogeant les descendants. Comme la durée d'exposition n'est généralement pas connue, ce sont les *proportions d'enquêtés qui déclarent que leur père ou leur mère vit toujours*, distribuées par âge, qui forment la matière première des estimations. Au niveau de la collecte des données, seules deux informations sont nécessaires : (1) l'âge des individus interrogés et (2) la survie de chacun de leurs parents. Comme on le verra plus loin, on se situe donc bien dans le cadre d'une exploitation indirecte des données : la durée d'exposition sera estimée à partir de l'âge des enfants qui procurent des informations. Des modèles seront ensuite utilisés pour convertir la proportion de survivants en probabilité de survie.

Au niveau national, la survie des parents est collectée dans deux types d'opérations : les enquêtes à larges échantillons et les recensements.

1.1.1 Enquêtes nationales

Historiquement, c'est dans le courant des années 1960 que les questions sur la survie des parents ont été intégrées pour la première fois dans les questionnaires d'enquêtes. Il s'agissait d'enquêtes rétrospectives menées au Tchad, au Cameroun et en Mauritanie en 1964-1965 (Blacker 1984). Par la suite, ces questions ont été reprises par les programmes internationaux d'enquêtes à large échantillons, qui se sont multipliées dès le début des années 80. On compte notamment parmi ces programmes ; les enquêtes démographiques et de santé (EDS), les enquêtes par grappes à indicateurs multiples (MICS), les enquêtes sur les niveaux de vie (LSMS) et plus récemment les enquêtes mondiales de la santé (WHS). Notons qu'avant le lancement des EDS, un module était consacré à la mortalité dans le programme d'enquêtes WFS (World fertility Surveys). Il recueillait notamment des données sur la survie des parents et des premiers conjoints, ainsi que sur les décès survenus dans les ménages [19].

Enquêtes EDS

Les enquêtes démographiques et de santé sont conduites dans la plupart des pays africains à l'échelle nationale sur de larges échantillons (généralement entre 6000 et 1000 ménages) depuis 1984 [20]. Elles visent avant tout à collecter des données sur la fécondité, la planification familiale, la santé maternelle et infantile, la mortalité des enfants, le VIH-sida ainsi que le paludisme et les problèmes de malnutrition. Depuis la fin des années 80, certaines enquêtes ont géo-codé les grappes d'échantillonnage au moyen d'un système GPS. Récemment, cet encodage des coordonnées GPS est même devenu systématique. Pour davantage de détails sur le

19. Ce module a été intégré aux enquêtes menées au Lesotho en 1977, ainsi qu'en Mauritanie en 1981, au Soudan en 1978/79 et au Cameroun en 1975. Ces données ne seront pas utilisées ici, car elles sont difficilement accessibles. Certains fichiers des enquêtes WFS peuvent être téléchargés sur http ://opr.princeton.edu/archive/wfs/, mais cela ne concerne pas les modules sur la mortalité.

20. http ://www.measuredhs.com/aboutdhs/history.cfm

programme, nous renvoyons au site http ://www.measuredhs.com.

Dans le premier cycle des EDS, une série de questions sur la survie des parents a été introduite *dans le questionnaire individuel* de certaines enquêtes [21]. Ces questions n'étaient pas destinées à l'étude de la mortalité ; elles visaient plutôt à établir les influences familiales sur les comportements de fécondité [22]. Les femmes *en union* répondaient aux questions suivantes :

1. Votre père et votre mère sont-ils encore en vie ?
2. Le père et la mère de votre partenaire sont-ils encore en vie ?
3. Votre parent décédé était-il en vie quand vous avez commencé à vivre avec votre premier mari/ partenaire ?
4. Depuis le jour où vous avez commencé à vivre avec votre premier mari/partenaire, avez-vous habité, vous et votre mari, avec un de ces parents dans le même ménage pendant au moins six mois ?
5. Combien d'années avez-vous vécu avec un parent ?

Ces questions permettent d'analyser la *survie des parents depuis le premier mariage*, ce qui présente plusieurs avantages (*cf.* section 3.4.2, p.159). Elles ont été supprimées par la suite, et durant les années 1980, certaines enquêtes n'ont prévu aucune question sur la survie des parents (Liberia 1986, Mali 1987, Togo 1988 ...). Avec l'enquête menée au Cameroun en 1991, une série de questions réapparaît. Elles sont cette fois intégrées au questionnaire ménage, et limitées aux enfants de *moins de 15 ans* (parfois 18 [23]) :

1. Est-ce que la mère qui a mis au monde (...) est toujours en vie ? (*"Is (NAME)'s natural mother alive ?"*) - Oui, non ou Ne sais pas
2. Est-ce que la mère qui a mis au monde (...) vit toujours dans ce ménage ? Si oui, quel est son nom ?
3. Est-ce que le père qui a mis au monde (...) est toujours en vie ?
4. Est-ce que le père qui a mis au monde (...) vit toujours dans ce ménage ? Si oui, quel est son nom ?

Depuis 1991, la collecte de ces informations est devenue quasiment systématique [24]. Elle s'accompagne de questions additionnelles sur les

21. Sénégal 1986, Burundi 1987, Ghana et Ouganda 1988. Dans l'enquête burundaise de 1987, des questions additionnelles étaient prévues sur le nombre d'années écoulées depuis le décès de la mère ou du père de l'enquêtée (Makinson 1993).

22. "The mean reason why the questions were included in the questionnaire was a filter for further questions, that establish whether respondents lived with their parents(-in-law) after marriage, so that the impact of residential patterns on contraceptive use and fertility could be investigated." Timaeus (1991a), p. 456

23. Dans un peu plus d'une enquête sur 10.

24. Après 1991, les enquêtes suivantes n'ont pas collecté ces informations : MIS Angola 2007, Kenya 2009, Burkina Faso 1999, Sénégal 1997, Côte d'Ivoire 1999.

enfants observés dans les ménages. [25]

La carte 1.1 indique le nombre d'enquêtes EDS ayant intégré des questions sur la survie des parents. Comme ces questions ont été posées quasiment systématiquement (sauf autour de 1988-1991), cette distribution recoupe bien sûr celle de la disponibilité des EDS. Globalement, l'ensemble de l'Afrique subsaharienne est couverte. L'Afrique australe et l'Afrique de l'Est (Tanzanie, Ouganda, etc.) sont particulièrement riches en données. Mais quelques pays font exception (Bostwana, Angola). Les pays les moins fournis sont généralement ceux qui ont traversé des phases de conflits ; la Somalie et le Soudan, notamment.

Depuis les années 1990, *seuls les jeunes enfants sont donc concernés par les questions sur la survie des parents*. Cette restriction aux enfants limite beaucoup l'intérêt des données EDS pour la méthode des orphelins, pour au moins trois raisons sur lesquelles nous reviendrons :

1. Quand les données sont disponibles pour différents groupes d'âge, la méthode procure des estimations de mortalité pour différentes périodes dans le temps. Or ici, les probabilités de survie ne peuvent être calculées que sur deux intervalles d'âge (5-9 ans et 10-14 ans) [26], ce qui fournit des estimations récentes uniquement.
2. Le fait que seuls les enfants soient interrogés réduit également la fiabilité des données. Par rapport aux réponses fournies pour les adultes, les déclarations qui concernent les enfants sont davantage sujettes à caution, en raison de l'"effet d'adoption" (*cf.* section 6.1.1, p.278).
3. Les biais induits par le VIH-sida sont par ailleurs particulièrement prononcés chez les plus jeunes enquêtés (*cf.* chapitre 4).

On peut donc espérer que les prochains rounds d'enquête étendent les questions sur la survie des parents *à tous les membres du ménage, quel que soit leur âge*.

Enquêtes MICS

Les enquêtes par grappes à indicateurs multiples (Multiple Indicator Cluster Survey) sont développées par l'UNICEF de façon assez comparable aux EDS. Elles sont axées sur la scolarisation, la santé et le travail des enfants, la prévention contre le paludisme, le support accordé aux orphelins, les comportements sexuels et la planification familiale. Un accent

25. Notamment sur leur scolarisation. Dans certains cas, des questions sont posées sur la prise en charge des enfants de 3 à 5 ans (école maternelle, garderie, etc.). Certaines enquêtes collectent également des données sur le travail des enfants de 5 à 14 ans (travail fait pour quelqu'un qui n'est pas membre du ménage, ainsi que pour le ménage), ou sur l'enregistrement à l'état civil des naissances des moins de cinq ans. D'autres informations détaillées (santé, nutrition, etc.) sont disponibles *sur les enfants des mères survivantes à partir du questionnaire individuel*, mais ceci exclut les orphelins. Depuis quelques années, les enfants sont mesurés et pesés au niveau du ménage, ce qui signifie que les orphelins sont compris également (Monasch et Boerma 2004).

26. Les données relatives aux enfants de 0 à 4 ans ne sont pas prises en compte par les méthodes de calcul.

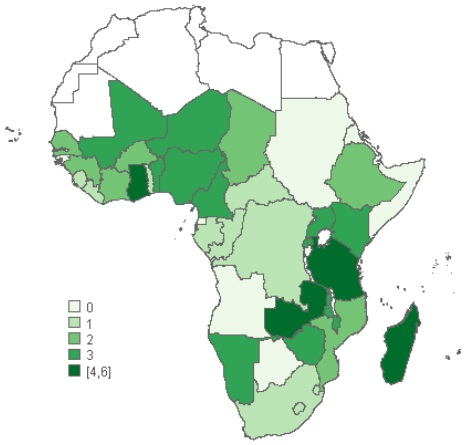

FIGURE 1.1 – *Nombre d'enquête EDS ayant intégré des questions sur la survie des parents*

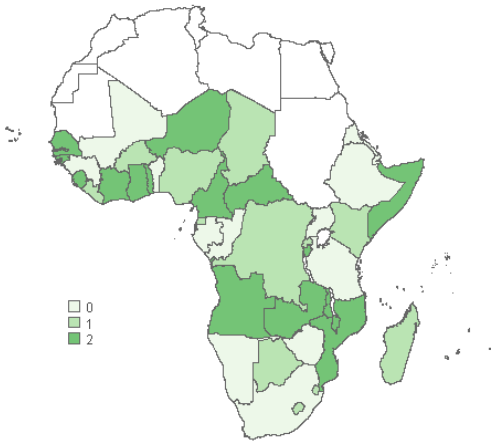

FIGURE 1.2 – *Nombre d'enquête MICS ayant intégré des questions sur la survie des parents*

est mis également sur l'enregistrement des naissances des enfants à l'état civil. Trois vagues d'enquête ont eu lieu ; les enquêtes MICS 1 ont été lancées dès 1995, mais elles sont assez sommaires et les fichiers de données ne sont pas facilement accessibles. Les enquêtes MICS 2 ont été menées autour de 2000 et construites sur le mode des EDS [27]. Quant aux enquêtes MICS 3, lancées en 2005 et progressivement distribuées dans le domaine public, elles sont surtout destinées à documenter les progrès faits en matière de programmes internationaux, comme les Objectifs du Millénaire, le programme *World Fit for Children*, ainsi que les objectifs de l'UNGASS [28] relatifs au VIH/sida. Les données et les détails du programme sont accessibles sur le site http ://www.childinfo.org/.

En ce qui concerne la mortalité adulte, les enquêtes MICS prévoient les mêmes questions sur la survie des parents que celles des enquêtes EDS. Elles sont également réservées aux 0-17 ans et intégrées au questionnaire ménage.

Comme on l'observe sur la carte 1.2, nous disposons de moins de données sur la survie des parents à partir des enquêtes MICS, parce que celles-ci sont moins nombreuses que les EDS. Cela dit, ces enquêtes présentent l'avantage de couvrir des pays qui n'étaient pas couverts par le programme EDS : c'est le cas notamment de la Somalie, de la Gambie, de la Guinée Bissau et de la Guinée équatoriale. Notons aussi qu'une nouvelle vague d'enquêtes est en cours en 2010, et ces enquêtes ne sont pas reprises ici : République centrafricaine, Tchad, Congo RDC, etc.

Enquêtes LSMS

Les enquêtes sur les niveaux de vie (Living Standards Measurement Study) sont conduites sous l'égide de la Banque Mondiale depuis 1980.

27. Les fichiers de données sont accessibles sur demande.
28. L'UNGASS est la déclaration onusienne faite en 2001 durant laquelle les pays signataires se sont engagés à relancer la lutte contre le vih/sida (http ://www.ua2010.org/)

Elles sont centrées sur la pauvreté, le chômage et les inégalités dans les pays en développement. Elles collectent des informations sur les caractéristiques des ménages, l'activité économique et agricole des individus, les dépenses en nourriture et en biens durables, l'épargne, les crédits, etc. Pour davantage de détails, voir http ://www.worldbank.org/lsms/.

Certaines enquêtes disposent de questions sur la survie des parents [29] A la différence des EDS, ces données permettent de mener des *analyses différentielles* directement sur les caractéristiques des parents, car quelques questions additionnelles sont prévues : on interroge notamment les individus sur la scolarisation de leurs parents et leur occupation... [30] Idéalement, de telles questions devrait être intégrées également aux enquêtes EDS [31].

***

Le tableau repris en annexe A.1.2 détaille, pour ces trois programmes de collecte, l'ensemble des enquêtes qui disposent de données sur la survie des parents.

1.1.2 Recensements

Les premières enquêtes nationales comprenant des questions sur la survie des parents ont été menées peu avant 1965. A la fin des années 1960 et au début des années 1970, *plusieurs recensements les ont intégrées à leurs questionnaires*. Dans le feuillet ménage, et pour chaque individu identifié, l'agent recenseur doit noter si chacun des deux parents a survécu jusqu'au recensement ou non. Il s'agit donc de questions qui ne représentent pas un surcroît de charge de travail conséquent. Pour autant, ces questions n'ont pas été intégrées dans tous les recensements. Sur les 142 recensements africains organisés depuis 1969 pour lesquels nous avons pu établir si la survie des parents avait été intégrée ou non aux questionnaires, près de 6 recensements sur 10 ne l'ont pas intégré [32].

29. C'est le cas des enquêtes menées en Côte d'Ivoire (de 1985 à 1988) et au Ghana (1987/8, 1988/9, 1991/2, 1998/9), qui présentent l'intérêt de prévoir plusieurs passages successifs sur une partie de l'échantillon.

30. Chaque membre du ménage déclare si son père est toujours vivant ou s'il est décédé, s'il a fréquenté l'école, quel a été son diplôme obtenu le plus élevé, et quel a été le genre de travail exercé pendant la majeure partie de sa vie. Les mêmes questions sont posées pour la mère. Les membres du ménage de plus de dix ans doivent également déclarer s'ils habitaient avec leurs parents lorsqu'ils avaient dix ans.

31. Outre le Ghana et la Côte d'Ivoire, une enquête menée en Afrique du Sud en 1993 a également collecté des données sur la survie des parents. Le quatrième pays africain a avoir organisé une enquête LSMS est le Malawi, et l'enquête de 2004 comprend également des questions sur l'âge des répondants au décès de leurs parents ainsi que sur les décès survenus dans le ménage. A ces enquêtes nationales, il faut ajouter un projet conduit en Tanzanie, dans la région de Kagera (*Kagera Health and Development Survey* 1991-1994 & 2004), où les individus ont été suivis d'une enquête à l'autre entre 1991 et 1994 (avec 912 ménages intégrés dans l'enquête), puis retrouvés en 2004. Ce projet a collecté des données détaillées sur la survie des membres des ménages entre les différents passages, mais également sur la survie des parents (niveau d'éducation, date de leur décès, etc.).

32. Notons que toutes les données qui ont pu être collectées n'ont pas nécessairement été exploitées - ni publiées ; certains recensements ont été perdus en raison de conflits ou de mauvaises conditions d'archivage, d'autres ont été jugés politiquement inacceptables ou trop peu fiables.

Les tableaux 1.2 et 1.3 reprennent les recensements africains menés depuis 1970 et indiquent (en gras) ceux qui ont interrogé les individus sur la survie de leurs parents (au moins des mères). Les recensements dont nous n'avons pas pu établir si cette question avait été intégrée aux questionnaires sont repris entre crochets [33].

TABLE 1.2 – *Recensements en Afrique de l'Ouest et en Afrique Centrale*

Afrique de l'Ouest	1970	1980	1990	2000
Bénin	1979		1992	**2002**
Burkina F.	1975	1985	**1996**	2006
Cap-Vert	1970	1980	1990	2000
Côte d'Iv.	1975	1988	**1998**	
Gambie	**1973**	**1983**	**1993**	**2003**
Ghana	1970	1984		2000
Guinée		1983	1996	
Guinée-Bissau	1970 1979		1991	2009
Liberia	1974	**1984**		[2008]
Mali	1976	**1987**	**1998**	[2009]
Niger	1977	**1988**		**2001**
Nigeria	1973		1991	[2006]
Sénégal	1976	**1988**		2002
Sierra Leone	**1974**	**1985**		**2004**
Togo	1970	1981		

Afrique centrale	1970	1980	1990	2000
Cameroun	1976	**1987**		[2005]
Centrafrique	1975	**1988**		2003
Congo	**1974**	1984	1996	[2007]
Gabon	1970	1980	1993	[2003]
Guinée équ.		1983	1994	2002
Tchad		**1989**	1993	2009

Comme c'est le cas des questions sur les décès des 12 derniers mois, on note au fil du temps une tendance à l'*intégration des questions sur la survie des parents dans les recensements* ; de 0.24 dans les années 1970, la proportion de recensements ayant collecté des données sur la survie des parents grimpe à 0.41 dans les années 1980, et atteint 0.58 dans les années 2000.

L'intégration de ces questions tient à l'histoire statistique de chaque pays. Certains d'entre eux n'ont jamais collecté ces données (Guinée, Zambie, Ethiopie...), tandis que d'autres pays l'ont fait systématiquement (Kenya, Gambie, Sierra Leone...). En prenant une fenêtre temporelle de 1980 à 2010, on constate que l'Afrique de l'Est est la région où cette question a le plus été posée (avec une proportion qui s'établit à 0.54), alors qu'elle était celle où les décès dans les ménages étaient le plus rarement collectés. L'Afrique centrale n'a intégré la survie des parents que dans un questionnaire sur quatre, l'Afrique australe dans un ques-

33. On peut y ajouter la Mauritanie (1977, 1988, [2000]) et le Soudan (**1973**,**1983**,**1993** et [2008]), qui sont classés en Afrique du Nord dans la classification UNECA.

TABLE 1.3 – *Recensements en Afrique de l'Est et en Afrique australe*

Afrique australe	1970	1980	1990	2000
Afr. du Sud			**1996**	**2001 2007**
Angola	1970			
Botswana	**1971**	1981	1991	**2001**
Lesotho	1976	**1986**	**1996**	**2006**
Malawi	**1977**	1987	**1998**	**2008**
Maurice	1972	1983	1990	2000
Mozambique	1970	1980	1997	**2007**
Namibie	1970	1981	1991	**2001**
Swaziland	**1976**	**1986**	**1997**	**2007**
Zambie		1980	1990	2000
Zimbabwe		**1982**	**1992**	**2002**

Afrique de l'Est	1970	1980	1990	2000
Burundi	1979		**1990**	[2008]
Comores		**1980** 1985	**1991**	[2003]
Congo RDC		**1984**		
Djibouti		1983		[2009]
Erythrée		1984		
Ethiopie		1984	1994	2007
Kenya	**1979**	**1989**	**1999**	[2009]
Madagascar	1975		**1993**	
Ouganda		1980	**1991**	**2002**
Réunion	1974	1982	1990 1999	2006
Rwanda	1978		**1991**	**2002**
Somalie	1975	1987		
Tanzanie	**1978**	**1988**		**2002**

tionnaire sur deux, l'Afrique de l'Ouest dans un peu moins d'un sur deux.

L'intérêt de ces données de recensement est que généralement, tous les individus sont concernés par les questions sur la survie des parents, *quelque soit leur âge*. De plus, ces données permettent de mener des analyses à des *niveaux géographiques plus fins*. Leur accès reste toutefois assez difficile. Les rapports des recensements ne comprennent pas systématiquement les tabulations nécessaires pour procéder aux calculs. Lorsque les proportions d'orphelins sont publiées, c'est souvent avec des classes d'âge trop larges, en privilégiant une distribution régionale, plutôt que par âge. Il faut alors pouvoir accéder aux données elles-mêmes pour procéder à ces tabulations. A ce stade, deux projets visent à faciliter la diffusion des données censitaires au-delà des offices statistiques nationaux : le premier, l'*African Census Analysis Project*, constitue une base de données unique, dans la mesure où les données individuelles d'un grand nombre de recensements africains y sont stockées. Elles restent toutefois distribuées sur base de conventions avec les pays et elles sont très difficiles d'accès. Le second projet, l'*IPUMS-International Census Data*, est la plus grosse opération de distribution des données censitaires au monde[34]. Les variables sont aussi standardisées que possible et la politique de distribution vise à favoriser au maximum l'accès aux données. Seule une fraction des données est distribuée (représentative de l'ensemble du pays). A ce jour, 18 fichiers de recensements sont disponibles pour l'Afrique subsaharienne dans 9 pays différents (Ghana, Guinée, Kenya, Mali, Rwanda, Sénégal, Afrique du Sud, Ouganda et Tanzanie). Quatorze d'entre eux contiennent des informations sur la survie des parents[35], et nous les utiliserons dans les chapitres 3 et 7.

1.2 Collecte des données relative à la survie des frères et sœurs

La collecte des données relatives à la survie des frères et sœurs démarre au début des années 1980. Les premières tentatives d'analyse ne sont pas très concluantes (Hill 1984), et les choses auraient pu en rester là, un peu comme c'est le cas des données sur les premiers conjoints.

C'est en fait la recherche de nouveaux moyens d'estimer la *mortalité maternelle* qui a relancé la collecte de données sur les sœurs, et celle-ci s'est ensuite étendue aux frères. L'intérêt de ces données du point de vue de l'estimation de la mortalité maternelle est détaillé dans l'encadré 2.

Rappelons que l'Afrique sub-saharienne concentre à elle seule 57% des décès maternels qui surviennent dans le monde. Selon les plus récentes statistiques, le nombre de décès maternels pour 100000 naissances vivantes se chiffre à 640 en Afrique subsaharienne, loin devant l'Asie du Sud (280), l'Océanie (230) et l'Asie du Sud-Est (160) (WHO, UNICEF, UNFPA and The World Bank 2010). La probabilité qu'une femme de plus de 15 ans décède d'une cause maternelle est de une chance sur 31 en

34. https ://international.ipums.org/international/

35. Mali 1987 et 1998, Sénégal 1988, Tanzanie 1988 et 2002, Kenya 1989 et 1999, Rwanda 1991 et 2002, Ouganda 1991 et 2002, Afrique du Sud 1996, 2001 et 2007.

Afrique subsaharienne. En 2008, le nombre de décès maternel variait de 180 pour 100 000 naissances vivantes en Namibie [36] à près de dix fois plus, soit 1 200 décès maternels pour 100 000 naissances vivantes au Tchad et en Somalie. A titre de comparaison, ce rapport est inférieur à 15 décès pour 100 000 naissances vivantes dans les pays industrialisés. On note également d'importantes fluctuations au sein des régions, parfois même entre pays limitrophes : 260 décès au Gabon contre 600 au Cameroun, par exemple, ou 180 décès en Namibie contre 610 en Angola. Mais de tous les indicateurs de santé publique, les estimations de mortalité maternelle sont certainement ceux qu'il faut considérer avec la plus grande prudence.

Encadré 2 **Estimation de la mortalité maternelle et importance des données sur les sœurs**

Dans les pays où les estimations de mortalité maternelle sont jugées relativement fiables, les institutions internationales se contentent de corriger les biais associés à la sous-déclaration (et aux problèmes de classifications). Mais dans les autres pays, et notamment dans la totalité des pays africains, d'autres méthodes doivent être utilisées. Les estimations les plus récentes des institutions internationales sont basées sur un modèle multi-niveaux faisant intervenir différentes sources de données (WHO, UNICEF, UNFPA and The World Bank 2010) [37].

Ce modèle multi-niveaux consiste à estimer la *proportion de décès maternels* parmi les décès survenus chez les femmes de 15 à 49 ans pour chacun des pays (PMDF). La proportion de décès maternels est ensuite combinée aux estimations de mortalité disponible (en l'occurence celles de l'OMS) et aux estimations relatives aux naissances (celles des Nations Unies) pour produire des rapports de mortalité maternels. Cette procédure est parfois appelée *envelope adjustment* parce qu'elle garantit que les niveaux de mortalité maternelle obtenus sont cohérents avec les niveaux de mortalité toutes causes. En fait, si le recours à la proportion de décès maternels est jugée plus fiable que le calcul direct des niveaux de mortalité maternels, c'est en raison d'une sous-estimation de la mortalité lorsque celle-ci est estimée de façon empirique (plutôt que d'être modélisée comme le font les Nations Unies et l'OMS). Pour estimer cette proportion de décès maternels pour chacun des pays, le modèle intègre une série de variables discriminantes au niveau macro : la fréquence de l'assistance à l'accouchement par du personnel qualifié, l'intensité de la fécondité (ISF), et le produit national brut (ainsi

36. Et même moins dans les territoires insulaires : 36 décès pour 100 000 naissances vivantes à l'Île Maurice, 94 au Cap Vert.

37. Voir www.who.int/reproductivehealth/publications/monitoring/9789241500265. En 2008, les mêmes institutions avaient recours à trois approches pour produire les statistiques africaines, dont un modèle de régression pour les pays où les données sur la mortalité maternelle étaient inexistantes ou jugées inutilisables (WHO, UNICEF, UNFPA and The World Bank 2008). Pour ces pays, le modèle de régression générait des estimations hors échantillon (*out-of-sample predictions*) de la proportion de décès maternels parmi les décès adultes féminins à partir de variables au niveau macro (PNB, ISF, etc.). C'est ce modèle qui a été modifié, étendu dans un cadre multi-niveau, et adopté pour tous les pays à statistiques déficientes pour la révision de 2010. Il ne s'agit plus cette fois de prédiction hors-échantillons. Une attention spéciale est accordée aux décès maternels associés au VIH. Voir également Graham et al. (2008), et Hogan et al. (2010), qui ont récemment proposé une méthodologie alternative.

que des indicatrices par pays, et par région).

Il nous paraît utile de mentionner ici les deux principales sources de données qui fournissent des estimations relatives à la proportion de décès maternels en Afrique, afin de situer la place occupée par les données sur les frères et sœurs [38].

1. La première source de données ne concerne encore qu'une minorité de pays africains ; il s'agit *des données de recensements*. Les Nations Unies recommandent d'intégrer aux recensements des questions sur les décès maternels [39], ce qui a été fait dans plusieurs pays : Afrique du Sud (2001, 2007), Burkina Faso (2006), Lesotho (1986, 1996, 2006), Bénin (1992, 2002), Zimbabwe (1992, 2002), Madagascar (1993), ou Namibie (1996, 2001), notamment. Par rapport à une collecte classique des décès des 12 derniers mois, l'ajout de questions sur les décès maternels ne nécessite pas un surcroît de travail important, puisque seuls les ménages où un décès de femme adulte a été observé sont concernés par une courte série de questions de précision (Hill et al. 2001) [40]. L'intérêt de ces données de recensement est également que des estimations peuvent être dégagées pour les niveaux sous-nationaux. Par contre, une plus grande sous-estimation des décès est possible (par rapport aux données basées sur les sœurs) du fait que les ménages se déstructurent souvent après le décès d'un adulte.

   Plutôt que de ne retenir que la proportion de décès maternels parmi les décès adultes - comme le font l'OMS, l'UNICEF, le FNUAP et la Banque Mondiale (2010) - , on pourrait alternativement travailler directement à partir des décès et de naissances déclarées dans les recensements. Mais il faut alors souvent ajuster à la fois les données de mortalité et de fécondité (en raison de leur sous-estimation fréquente). On utilisera pour cela les méthodes classiques de correction ; *general growth balance method*, *synthetic extinct generation method*, etc. Ces méthodes invitent souvent à procéder à des ajustements substantiels. A titre d'illustration, mentionnons le cas des recensements béninois (1992) et malgaches (1993). Des facteurs de corrections qui s'élèvent respectivement à 2.7 et à 2.5 devraient être appliqués aux nombres bruts de décès de femmes adultes pour approcher le nombre estimé de décès 'réels' (Stanton et al. 2001). De telles corrections sont délicates, d'autant que ces méthodes reposent sur des hypothèses qui ne sont parfois pas rencontrées, et que les facteurs de correction sont entourés de larges intervalles de confiance (*cf.* p. 15).

2. Le deuxième type de données utilisé en Afrique se base sur la *survie des sœurs*. Il ne s'agit donc plus d'enregistrer les décès maternels dans les *mé-*

38. Il faudrait ajouter à ces deux sources de données les études RAMOS - Reproductive Age Mortality Surveys, qui consistent à recourir à diverses sources d'information (enquêtes ménages, données administratives, registres d'état civil, systèmes de surveillance des maladies...). Ces enquêtes sont rares à l'échelle nationale en Afrique, en raison de leur coût prohibitif.

39. Voir les *Principes et recommandations des Nations Unies pour les recensements* (UN 2008b) (§2.196).

40. On outre, on peut facilement concevoir qu'un suivi soit organisé dans les ménages concernés pour procéder par exemple à des autopsies verbales. Ceci est fait régulièrement dans les observatoires de population, mais nous n'avons pas connaissance d'un recensement africain ou ceci ait été organisé à plus large échelle.

nages, mais dans les *fratries*. Ces données sont collectées régulièrement depuis que Graham et al. (1989) ont élaboré, à la fin des années 1980, la "méthode des sœurs".

Interroger les individus sur la survie de leurs sœurs s'avère être une approche combinant au moins trois avantages :

(a) Tout d'abord, comme les décès maternels restent des événements rares, leur analyse requiert de mener des enquêtes sur de très larges échantillons, ce qui est souvent hors de portée (notamment pour des raisons budgétaires). Poser des questions sur l'ensemble des sœurs permet de résoudre partiellement le problème ; le nombre de personnes soumises au risque de décéder augmente proportionnellement à la taille des fratries. Les intervalles de confiance autour des estimations restent cependant importants, ce qui empêche de suivre les tendances ou de travailler au niveau sous-national.

(b) Interroger les sœurs permet également d'obtenir des informations qui sont jugées relativement fiables, notamment parce que les femmes enceintes sont parfois assistées par leurs propres sœurs pendant la période de grossesse, lors de la naissance ou durant les semaines qui suivent (Graham et al. 1989).

(c) Cette collecte permet aussi de contourner le problème classique lié au fait que les ménages se désagrègent souvent à la suite d'un décès adulte. Les données ne sont pas non plus affectées par la migration.

1.2.1 Enquêtes nationales

Pour estimer la mortalité maternelle, plusieurs enquêtes ont adopté une approche indirecte basée sur les sœurs, sur laquelle nous reviendrons dans le chapitre 5. Cette méthode est souvent appliquée en Afrique subsaharienne (David et al. 1991, Walraven et al. 1994, Danel et al. 1996, Font et al. 2000, Lech et Zwane 2002, Mbaruku et al. 2003), en dépit de problèmes méthodologiques importants (Garenne et Friedberg 1997). Son principal avantage est qu'elle nécessite que soient posées une série de questions assez simples :

1. Combien de sœurs nées de la même mère avez-vous eu ?
2. Combien d'entre elles se sont mariées ?
3. Parmi ces sœurs mariées,
   - combien sont toujours vivantes ?
   - combien sont décédées ?
4. Parmi ces sœurs mariées,
   - combien l'ont été lors de la grossesse ?
   - combien l'ont été lors de l'accouchement ?
   - combien l'ont été dans les six semaines qui ont suivi ?

Notons que ces enquêtes récoltent des informations sur *les décès associés à la grossesse,* c'est-à-dire les décès survenus durant la grossesse ou les

2 mois qui ont suivi, quelle que soit leur cause. Il ne s'agit pas des *décès maternels* proprement dits (liés à une cause déterminée ou aggravée par la grossesse) [41]. L'hypothèse est qu'il y a une certaine compensation entre l'inclusion dans les calculs de décès non maternels et la sous-déclaration des décès maternels, liés par exemple aux avortements provoqués (Stanton et al. 1997). Par ailleurs, certaines enquêtes s'écartent de la définition établie par l'OMS, en considérant une période de référence soit de six semaines, soit de deux mois - tandis que l'OMS ne considère pas les décès survenus plus de 42 jours après la naissance comme des décès maternels [42].

A partir de ces questions, on utilisera des méthodes indirectes, parce qu'aucune information sur l'âge au décès n'est collectée : comme pour la survie des parents, la période d'exposition est estimée à partir de *l'âge des individus qui procurent les informations*. Ces questions concernent uniquement les sœurs ayant survécu jusqu'à leur quinzième anniversaire, ou alternativement les sœurs mariées (lorsqu'il est clair que l'exposition au risque de grossesse débute majoritairement au mariage).

Au départ, seules les informations sur les *sœurs* étaient collectées et la méthode ne concernait que la mortalité maternelle. Mais elles permettaient toutefois d'estimer également la mortalité féminine générale. Par la suite, certaines enquêtes ont ajouté des questions sur les *frères*. Etendre les questions aux frères visait avant tout à réduire les omissions et à pouvoir tester la fiabilité des données (Rutenberg et Sullivan 1991). Les enquêtées semblent en effet procurer des informations de meilleure qualité lorsqu'elles ont la possibilité de lister toute leur fratrie (on imagine aisément que l'omission d'une sœur est moins fréquente lorsque les enquêtées ont l'occasion de classer tous les frères et sœurs par ordre de naissance).

Dans certains cas, au prix d'une durée moyenne d'entretien allongée de dix minutes environ (Rutenberg et Sullivan 1991), certaines enquêtes ont également prévu des questions sur l'âge des frères et sœurs survivants, et/ou sur leur âge au décès et le nombre d'années écoulées depuis lors (c'est le cas des enquêtes EDS et WHS). Dès lors, la durée d'exposition peut être calculée directement et des taux de mortalité générale peuvent être obtenus simplement en rapportant le nombre de décès à la durée totale d'exposition. Enfin, plus récemment, certaines enquêtes ont posé ces questions aux hommes également, et plus seulement aux femmes enquêtées.

Trois grands programmes d'enquête collectent aujourd'hui de façon routinière de telles données (enquêtes EDS, MICS, et WHS).

41. L'OMS définit le décès maternel comme le "décès d'une femme survenu au cours de la grossesse ou dans un délai de 42 jours après sa terminaison, quelle qu'en soit 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" (10e révision CIM).

42. On parlera alors plutôt de mortalité maternelle tardive, ce qui inclut les décès survenus entre 6 semaines et un an après l'accouchement. Allonger la période de référence tend à diminuer les taux de mortalité, dans la mesure où une large majorité des décès maternels surviennent dans les tous premiers jours qui suivent la naissance.

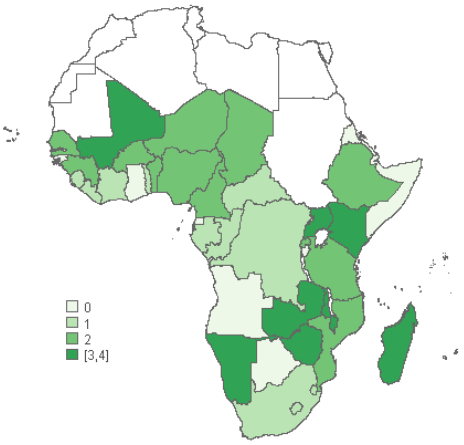

FIGURE 1.3 – *Nombre d'enquête EDS ayant intégré des questions sur la survie des frères et sœurs*

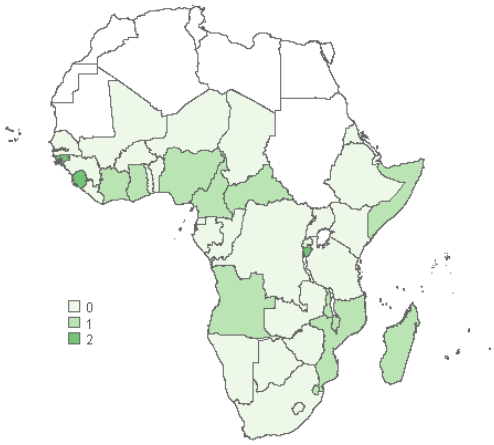

FIGURE 1.4 – *Nombre d'enquête MICS ayant intégré des questions sur la survie des frères et sœurs*

Enquêtes EDS

Les enquêtes EDS sont de loin la première source d'information sur la survie des frères et sœurs. La carte 1.3 présente la répartition des 60 enquêtes ayant intégré un module collectant de telles données[43]. Le tableau repris en annexe A.1.3 (p.385) en fait le détail.

De façon générale, la procédure de collecte a peu variée depuis la première enquête du programme qui a étendu sa portée à la fratrie (Egypte 1988). Dès 1989, les questions sur les *frères* ont été ajoutées, ainsi que les *âges actuels ou au décès*, et la durée écoulée depuis chaque décès (Bolivie 1989). Ces questions ont été intégrées au questionnaire individuel, dans un nouveau module intitulé "mortalité maternelle".

En Afrique, la taille des échantillons féminins varie de 4 849 personnes (Malawi 1992) à 33 380 personnes (Nigeria 2008) pour les enquêtes disposant du module, avec une médiane à 7 988 personnes. La moitié des enquêtes comptent entre 6 690 et 10 480 enquêtées.

Depuis l'enquête malienne de 1987, les hommes ont également été interrogés dans de nombreuses enquêtes EDS africaines[44]. Dans une dizaine d'enquêtes, les questionnaires "hommes" ont intégré le module sur la mortalité maternelle (Cameroun 2004, Congo 2005, Malawi 1992...). Tous les hommes éligibles ont alors fourni des informations sur leurs frères et sœurs. Les plages d'âge d'éligibilité varient d'un pays à l'autre, de 15 à 64 ans[45]. Les échantillons d'hommes sont également générale-

43. A la date du 10 novembre 2010.

44. Dans 6 enquêtes (Ghana 1988, Cameroun 1991, etc.) seuls les maris des femmes éligibles ont été interrogés. Mais cela ne concerne pas les enquêtes où les hommes ont été interrogés sur la survie de leurs frères et sœurs. Dans 9 enquêtes les hommes n'ont pas du tout été interrogés (Afrique du Sud 1998, Madagascar 1992, etc.).

45. Par exemple : 20-54 (Malawi 1992), 15-59 (Cameroun 2004), 15-54 (Zimbabwe 2006)

ment plus petits que ceux des femmes ; sur l'ensemble des enquêtes EDS collectant des données sur les hommes, les échantillons masculins sont en moyenne 3 fois plus petits que ceux des femmes. En ce qui concerne la survie des frères et sœurs, les échantillons varient de 1 151 (Malawi 1992) à 15 490 hommes (Nigeria 2008), avec une médiane qui s'établit à 2 913 personnes.

Quel que soit le sexe des enquêtés, le module prend la forme d'une *biographie génésique de la mère de l'enquêté(e)*. Chacun de ses enfants - c'est-à-dire chacun des frères et sœurs de l'enquêté(e)- est listé par ordre de naissance et l'enquêtée doit répondre aux questions suivantes :

- Quel nom a été donné à votre frère ou sœur le/la plus agé(e) (puis suivant) ?
- Est-il de sexe masculin ou féminin ?
- Est-ce qu'il/elle est toujours en vie ?
- Quel âge a-t-il/elle ?
- Combien y a-t-il d'années qu'il/elle est décédé(e) ?
- Quel âge avait il/elle lorsqu'il/elle est décédé(e) ?
  - Dans certains cas, en cas de non-réponse : est-ce qu'il/elle est mort(e) avant l'âge de douze ans ? (si oui : code spécifique)
  - Si non, dans certains cas : est-ce qu'il/elle est mort(e) avant de se marier ?
- Etait-elle enceinte quand elle est décédée ?
- Est-ce qu'elle est décédée au cours d'un accouchement ?
- Est-ce qu'elle est décédée dans les deux mois suivant la fin d'une grossesse ou d'un accouchement ?

Certains enquêtes ont prévu des questions additionnelles [46]. Dans la première enquête EDS intégrant des questions sur la mortalité maternelle, un filtre était prévu pour n'*autoriser qu'une sœur par ménage à répondre à ces questions*. Quelques mois plus tard, l'EDS menée au Soudan en 1989-1990 a demandé à toutes les femmes observées dans le ménage de répondre aux questions. Les enquêtes qui ont suivi ont fait de même. On verra que cela donne des estimations moins biaisées que lorsqu'une seule sœur est enquêtée. Par contre, cela signifie que le fichier compte un certain nombre de fratries dupliquées, puisqu'une certaine proportion des enquêté(e)s vivent dans leur ménage avec une ou plusieurs de leurs frères/sœurs.

ou 15-64 (Nigeria 1999). Sur les autres enquêtes hommes n'ayant pas posé des questions sur les frères et sœurs, on trouve d'autres plages d'âge ; 15-60 (Tanzanie 1992), 20-64 (Benin 1996), 18 et plus (Burkina Faso 1993), 20 et plus (Sénégal 1993 et 1997) et même 15-49 (Swaziland 2006). 15-59 est la plage d'âge retenue dans 6 enquêtes sur 10, 15-54 dans une enquête sur 6.

46. L'enquêteur demande parfois le nombre total de frères et sœurs nés de la même mère. Dans certains cas, l'enquêté doit également se placer dans l'ordre de naissance des frères et sœurs. Une question a aussi été intégrée dans certaines enquêtes pour identifier si les décès étaient associés à des complications durant la grossesse ou l'accouchement, une autre vise à comptabiliser le nombre d'enfants nés vivants des sœurs de l'enquêté (ou le nombre de grossesses), et certaines enquêtes collectent également des données sur le statut matrimonial des sœurs ou sur le lieu du décès (hôpital, centre de santé, domicile. . .).

Dans un premier temps, une question était posée dans les enquêtes EDS sur l'*existence d'une sœur de 15 à 49 ans déjà interrogée dans le ménage*. Cette question permettait donc d'identifier ces fratries dupliquées dans le fichier. Ce fût le cas des enquêtes menées en 1992 au Sénégal et au Niger. Mais cette question n'a pas permis de collecter des réponses fiables, et par la suite, elle a été supprimée des questionnaires, rendant impossible l'identification complète des fratries dupliquées dans les fichiers. Nous reviendrons sur ce point dans l'encadré 3 (p. 38).

Enquêtes MICS

A partir de la seconde vague d'enquêtes MICS, l'UNICEF a opté pour une collecte des données sur la survie des frères et sœurs à partir des questions types utilisées par la "méthode des sœurs" (c'est-à-dire les sept questions types sans information sur la durée d'exposition)[47]. Lors de la troisième vague, quelques enquêtes ont également opté pour cette approche indirecte, en ne collectant que des informations sur les sœurs de tous les adultes de plus de 15 ans[48]. Comme pour les EDS, aucune question relative aux sœurs déjà interrogées dans le ménage n'était prévue dans les questionnaires MICS. La carte 1.4 (p. 34) présente la distribution des 19 enquêtes disponibles par pays. Le round 4, mené en 2009-2011, ne semble pas collecter d'informations sur la survie des adultes.

Enquêtes WHS

Une autre vaste vague d'enquêtes a été lancée par l'OMS en 2001. Une quinzaine d'enquêtes ont notamment été menées récemment en Afrique subsaharienne, sur des échantillons comprenant entre cinq à dix milles personnes de plus de 18 ans.[49] Elles visent la comparabilité internationale, et sont centrées principalement sur les enjeux ciblés par les Objectifs du Millénaire pour le Développement. Les enquêtes comprennent plusieurs modules, comme l'auto-évaluation de l'état de santé, l'exposition aux facteurs de risques sanitaires, ainsi que les différents types de traitement reçus, et même les dépenses de santé. Les questionnaires prévoient également un module sur la survie des frères et sœurs. Les détails du programme sont accessibles sur http ://www.who.int/healthinfo/survey/en/index.html.

A la différence des autres programmes d'enquêtes, *un seul adulte* (âgé de plus de 18 ans) répond au questionnaire individuel. Il est sélectionné au hasard à partir d'une table Kish[50]. Par rapport aux enquêtes EDS, ce module contient également de nombreuses questions de précision assez

47. Cette collecte a caractérisé les enquêtes menées en 2000 au Burundi, au Comores, au Swaziland, au Sierra Leone, à Madagascar, ainsi qu'en Guinée Bissau et en Gambie (en 2001).

48. C'est le cas des enquêtes Burundi 2006, Sierra Leone 2005, Somalie 2006, Guinée Bissau 2006. L'enquête menée au Malawi en 2006 a opté pour une approche directe de type EDS.

49. Au total, septante deux pays ont participé à la série d'enquêtes de 2002-2003, aussi bien parmi les pays occidentaux que les pays en développement.

50. La technique consiste à lister toutes les personnes éligibles par âge puis à sélectionner l'une de ces personnes à partir d'un nombre défini dans des tables (distribuées aux

originales. Deux d'entre elles spécifient d'abord le nombre de frères et sœurs nés avant et après l'enquêté(e). Une question porte ensuite sur le temps écoulé depuis le dernier contact avec les différents frères et sœurs. Leur type de résidence est également enregistré [51]. Quelques questions classiques portent sur l'éventualité d'un décès maternel, et une courte autopsie verbale est également prévue : le décès était-il lié à des violences ? Quelles étaient les circonstances et les symptômes qui l'ont entouré ? Où a-t-il eu lieu ? Quelles étaient les conditions de morbidité sous-jacentes ? Il est donc possible d'amorcer une analyse par cause, ce qui reste assez rare et très problématique dans les pays en développement. Cela dit, cette analyse reste sommaire et permet tout au plus d'identifier les morts maternelles et les morts violentes. Dans un article à la fois très cité et très critiqué, Obermeyer et al. (2008) ont utilisé ces données pour analyser la mortalité liées aux guerres dans 13 pays. La qualité des données des enquêtes WHS nous a toutefois semblé nettement moins bonne que celle des autres programmes d'enquête. Nous avons donc fait le choix de ne pas les utiliser dans le cadre de cette recherche.

Le Tableau repris en annexe A.1.3 (p. 385) détaille l'ensemble des enquêtes des différents programmes (EDS, MICS et WHS) ayant intégré des questions sur la survie des frères et sœurs.

* * *

De ce qui précède ressort une différence importante entre les enquêtes EDS et MICS par rapport aux enquêtes WHS : les premières interrogent *tous* les adultes éligibles présents dans le ménage, tandis que les enquêtes WHS se limitent à *un enquêté par ménage*. On touche là à l'une des principales difficultés de la méthode des frères et sœurs : celle de saisir en quoi les choix faits en matière d'échantillonnage renforcent ou réduisent les biais de sélection. Faut-il ou non permettre à plusieurs frères et sœurs de répondre à l'enquête [52] ? Nous reviendrons sur cette question dans le chapitre 5, mais précisons d'emblée que c'est bien le cas : il est important que *tous les membres d'une fratrie repris dans le plan d'échantillonnage soient effectivement enquêtés*. Il n'y a donc pas lieu de chercher à ne retenir qu'une seule déclaration par fratrie. Par contre, il faut reconnaître que certaines fratries seront de ce fait mentionnées à plusieurs reprises dans les fichiers d'enquête ; deux sœurs ou deux frères enquêtés vont pouvoir fournir les mêmes informations sur leur fratrie. Cette question est discutée dans l'encadré 3.

enquêteurs). Le système est conçu de telle façon qu'il garantit que chaque membre du ménage a la même chance d'être sélectionné s'il fait partie de la population d'intérêt.

51. Habitation privée, institution militaire, etc.

52. Nous avons vu que cette question s'est posée dès les débuts. Dans l'enquête EDS menée en Bolivie en 1989, on a évité que plusieurs sœurs procurent des informations au sein d'un même ménage pour ne conserver qu'une seule déclaration par fratrie. Par la suite, tous les femmes éligibles (âgées de 15 à 49 ans) ont été interrogées sur leur fratrie.

Encadré 3 **Répétition de certaines fratries dans les fichiers d'enquêtes**

Dans le chapitre 5, nous traiterons des biais de sélection qui caractérisent les données sur les frères et sœurs. Nous verrons notamment que la répétition de certaines fratries dans les fichiers (quand différents membres d'une même fratrie sont enquêtés) est plutôt bénéfique : elle joue un rôle important pour annuler certains de ces biais de sélection. Répétons donc qu'il n'y a pas lieu de ne retenir qu'une seule déclaration par fratrie pour estimer la mortalité adulte.

Il peut toutefois être utile d'estimer la fréquence avec laquelle certaines fratries sont dupliquées dans les fichiers. Cette question a son intérêt pour au moins deux raisons :

- D'une part, la répétition de certaines fratries dans les fichiers a un impact sur les caractéristiques des fratries observées, en termes de *distribution par sexe et par âge*.
  Prenons l'exemple du rapport de masculinité des frères et sœurs. Imaginons que l'on parvienne à n'interroger qu'une seule personne par fratrie. Si l'on calculait un rapport de masculinité sur les frères et sœurs observés (en excluant les enquêtés), on devrait dans ce cas retrouver une valeur proche du rapport de masculinité de la population dont ils sont issus. Par contre, dans de nombreuses enquêtes EDS, on trouvera nettement plus de sœurs que de frères (cfr. p.295). Cela tient au fait que de nombreuses enquêtes n'interrogent que les femmes sur leurs fratries, et que les fratries peuvent être répétées dans les fichiers. Or, dans cette configuration, les fratries *plus féminines* auront plus de chances d'être mentionnées à plusieurs reprises dans les fichiers. On voit donc ici que la répétition de certaines fratries dans les fichiers peut avoir un impact sur la composition par sexe des fratries observées. Nous reviendrons sur ce point par la suite.

- D'autre part, le fait que certaines fratries sont comptabilisées plusieurs fois dans les fichiers contredit l'hypothèse d'indépendance des observations. En effet, cette répétition des fratries est une source d'"hétérogénéité non observée". La variation *entre* les fratries est accentuée par rapport à la variation *au sein* des fratries. Si nous utilisons des modèles de régression, les erreurs standards vont donc être sous-estimées, et les intervalles de confiance réduits. Idéalement, il faudrait dès lors pouvoir tenir compte de cette hétérogénéité non observée. Nous reviendrons sur ce point brièvement.

Dans cet encadré, nous présentons les premiers résultats d'une procédure que nous avons développée en vue d'*apparier les frères et sœurs adultes vivant sous un même toit*. Cette procédure, qui s'applique aux fichiers d'enquêtes EDS, est détaillée dans l'annexe A.1.4 (p.388).

Précisons d'abord que la probabilité qu'une fratrie soit dupliquée dans le fichier d'enquête va dépendre d'un grand nombre de facteurs : sa taille, la survie des frères et sœurs qui la composent, les caractéristiques de cette fratrie (en terme de composition par âge et par sexe), ou encore la fréquence de la cohabitation entre frères et sœurs aux âges adultes. Notons par ailleurs que des informations

peuvent être collectées à plusieurs reprises sur une fratrie donnée *même quand une seule femme est enquêtée par ménage* : en effet, deux sœurs peuvent vivre dans deux ménages différents, mais être toutes deux enquêtées[53]. La répétition de certaines fratries dans les fichiers d'enquêtes dépend donc également de la taille des clusters enquêtés ainsi que de la proximité des frères et sœurs qui ont quitté le domicile parental (Trussell et Rodriguez 1990).

Mais peut-on au moins identifier les fratries qui sont dupliquées dans les fichiers parce que plusieurs de leurs membres ont été interrogés *dans un même ménage* ? Peut-on les apparier ? Il faut bien sûr éviter un appariement manuel de ces fratries, ce qui serait trop laborieux. L'approche que nous proposons consiste donc à combiner un appariement *déterministe* (notamment à partir des liens de parenté entre chaque individu et le chef de ménage), et un appariement *probabiliste* (basé sur la ressemblance des différentes déclarations faites sur les fratries). Cette procédure s'applique uniquement aux adultes *éligibles au questionnaire individuel*. Une reconstitution complète de tous les liens de fratrie qui unissent les différents membres du ménage serait impossible, en raison de la relative pauvreté des informations collectées sur les liens de parenté (Tichit et Robette 2006)[54]

Le graphe et les cartes qui suivent résument les résultats de cet appariement.

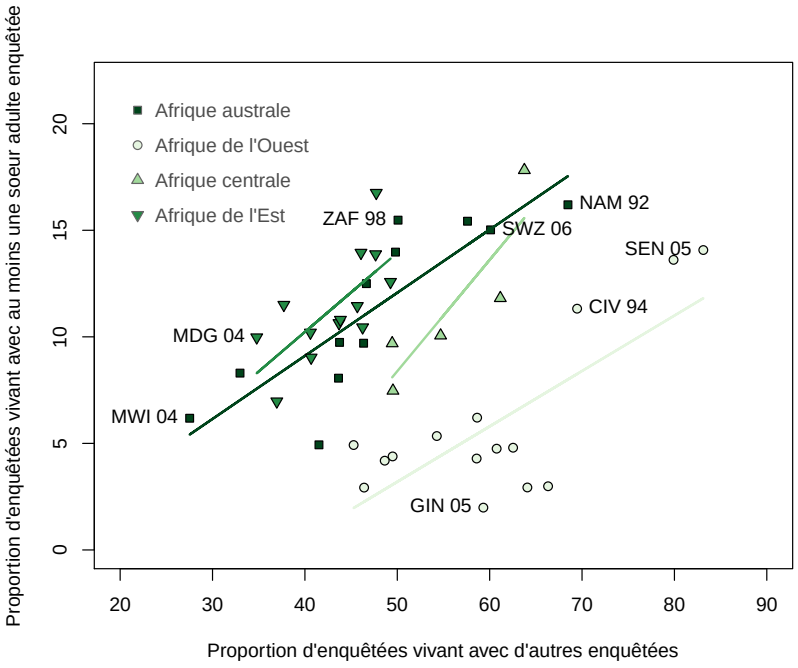

FIGURE 1.5 – *Proportion d'enquêtées vivant avec au moins une sœur adulte enquêtée selon la proportion d'enquêtées vivant avec d'autres enquêtées (n = 45 EDS)*
Note : La liste des codes utilisés pour repérer les pays et la classification régionale retenue sont repris en page 380.

Nous nous intéressons ici aux *proportions de femmes enquêtées qui co-*

53. En parcourant les fichiers d'enquêtes, on peut d'ailleurs retrouver des femmes qui habitent dans des *ménages distincts* (au sein d'un même cluster), mais qui procurent en fait les mêmes informations sur leur fratrie respective.

54. Par contre, l'appariement détaillé en annexe A.1.4 intègre les adultes éligibles masculins lorsqu'ils ont également été interrogés sur leur fratrie.

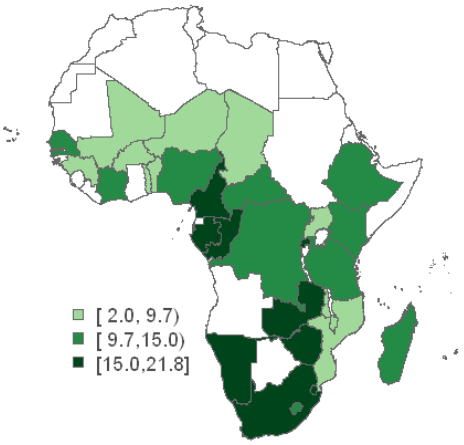

FIGURE 1.6 – *Proportion de répondants co-habitant avec (au moins) un frère ou une sœur (dernières enquêtes EDS)*

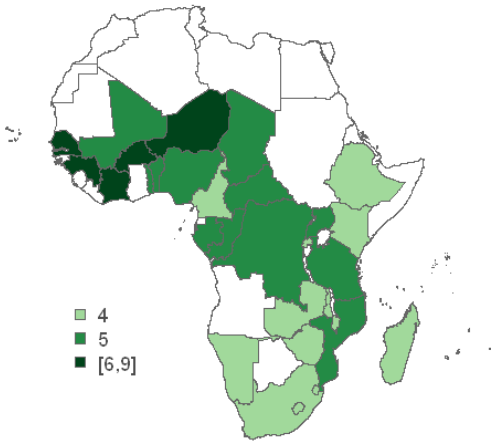

FIGURE 1.7 – *Taille moyenne des ménages (dernières enquêtes EDS)*

habitent avec au moins une sœur enquêtée elle-aussi[55]. Cette proportion dépend bien sûr de l'intensité de la cohabitation entre femmes adultes (qu'elles soient sœurs ou non). Plus les ménages comptent de femmes éligibles au questionnaire individuel, plus la probabilité qu'une enquêtée co-habite avec une sœur qui a été enquêtée elle aussi est élevée. Le graphe 1.5 présente donc en ordonnées les proportions de *femmes enquêtées cohabitant avec (au moins) une sœur adulte enquêtée*, selon la proportion de *femmes enquêtées cohabitant avec d'autres femmes enquêtées* (en abscisses). On constate que la cohabitation avec au moins une sœur adulte (de 15 à 49 ans) varie sensiblement d'une enquête à l'autre (entre 2 et 17%). Comme attendu, elle est positivement associée à la proportion d'enquêtées cohabitant avec d'autres femmes enquêtées (qui varie globalement entre 27 et 85%). On note également d'importantes différences régionales ; les proportions de sœurs adultes qui cohabitent entre elles sont les plus élevées en *Afrique australe* et en *Afrique de l'Est*, et les plus faibles en *Afrique de l'Ouest*.

On retrouve ce découpage régional sur la carte 1.6, construite à partir des enquêtes EDS les plus récentes. Cette carte est à comparer avec la carte 1.7, qui présente la taille moyenne des ménages pour les mêmes enquêtes. Cette comparaison indique que *la taille des ménages n'est pas un bon prédicteur de la proportion de fratries dupliquées dans les fichiers*. En effet, les ménages d'Afrique de l'Ouest sont les plus nombreux, mais la cohabitation entre sœurs adultes y est proportionnellement moins importante qu'ailleurs. La corrélation entre la taille moyenne des ménages enquêtés et la proportion d'enquêtées qui vivent avec au moins une sœur enquêtée s'avère négative (-0.18) et non significative.

Par contre, ces variations régionales recoupent celles observées en termes de *précocité du mariage*. Cela est assez intuitif : plus le mariage est tardif, plus on devrait trouver de sœurs adultes (15-49) qui cohabitent ensemble. C'est ce que nous observons ici, puisque le mariage féminin est traditionnellement plus

55. Pour des raisons de comparabilité, nous ne présentons ici que les résultats pour les enquêtes n'ayant interrogé que les femmes.

tardif en Afrique australe et en Afrique de l'Est, et c'est là que nous retrouvons les proportions les plus élevées de sœurs adultes cohabitantes. Si nous calculons l'âge au premier mariage à partir des proportions de célibataires déclarées dans les enquêtes [56], nous obtenons une corrélation positive (0.72) et significative (<0.001) entre cet âge et la proportion d'enquêtées qui co-habitent avec au moins une sœur enquêtée.

En définitive, si, en moyenne, *10% des enquêtées de 15 à 49 ans cohabitent avec au moins une sœur enquêtée elle aussi, une proportion importante de fratries sont dupliquées dans les fichiers*. Sur l'ensemble des enquêtes EDS, cette proportion se chiffre à 5.6%, mais elle varie de 1 à 12% d'une enquête à l'autre.

56. Voir annexe A.2.7, et en particulier la carte A.13, page 416.

Conclusion

Il est devenu classique de répéter que le manque de données est le principal obstacle à l'élaboration de statistiques fiables de mortalité en Afrique, surtout pour les âges adultes. Il est vrai que la santé des adultes figure encore rarement parmi les priorités des programmes de santé ou des opérations de collecte. Souvent, ces priorités restent axées sur la santé maternelle, sur la santé sexuelle et reproductive, et sur la survie des enfants.

Ce chapitre a toutefois mis en avant le fait qu'il existe de *nombreuses séries de données disponibles sur la survie des proches* (parents et frères et sœurs), qui viennent compléter les données d'état civil et celles relatives aux décès survenus dans les ménages.

Au total, en combinant les différents programmes d'enquêtes, plus de 135 enquêtes nationales ont collecté des données sur la survie des parents depuis les années 1980. Un peu moins d'une centaine d'enquêtes disposent également d'informations sur la survie des frères et sœurs, et la plupart des pays sub-sahariens sont couverts par au moins l'une d'entre elles. A cela s'ajoutent également les recensements, qui, quoique moins accessibles, collectent souvent des données sur la survie des parents sans restriction d'âge, ce qui n'est pas le cas des enquêtes.

Le principal intérêt de ces données sur les proches est qu'elles sont collectées à intervalle régulier, la plupart du temps de façon comparable d'une enquête à l'autre et d'un pays à l'autre. En l'absence de données d'état civil, ce sont également les seules données qui permettent de reconstruire des tendances de la mortalité à partir d'une seule opération de collecte.

L'existence de ces données dans un contexte où les informations démographiques restent rares plaide en faveur d'un examen plus systématique de leur potentiel et de leurs faiblesses.

La microsimulation : un outil pour exploiter les données imparfaites

2

Sommaire

2.1 Avantages de la microsimulation 46
2.2 SOCSIM en bref 50
2.3 Paramètres démographiques introduits 55
2.4 Intégration du VIH-sida 75
Conclusion 96

Dans les recensements et les enquêtes démographiques, la qualité des informations qui concernent les âges, les dates, les décès ou encore les relations de parenté reste assez faible et extrêmement variable en Afrique subsaharienne (Pullum 2006, Ewbank 1981). L'estimation de la mortalité adulte à partir de la *survie des proches* combine toutes ces imprécisions.A ces erreurs de déclaration s'ajoutent en outre de nombreuses autres inconnues. Par exemple, les schémas par âge de la mortalité ne sont pas connus avec précision. Les estimations déjà disponibles ne constituent souvent que des ordres de grandeur entourés d'incertitudes. L'ampleur de la variabilité aléatoire autour de ces estimations est également difficile à mesurer - quand elle n'est pas tout simplement incalculable - tout comme l'ampleur des biais inhérents au caractère rétrospectif des données. Par ailleurs, les hypothèses sur lesquelles reposent les méthodes d'estimation ne sont pas nécessairement rencontrées dans la pratique.

Sans pouvoir distinguer ce qui tient aux *spécificités des méthodes* et ce qui a plutôt trait aux *caractéristiques des données*, il est donc difficile de comparer les différentes méthodes d'estimation ou de se prononcer sur la plausibilité des niveaux de mortalité qu'elles procurent dans les contextes africains.

Pour pallier à cela, la simulation peut s'avérer utile. Elle permet de contrôler les différents éléments en jeu, et donc de distinguer l'effet des différents biais et des erreurs de déclaration. Elle peut servir de guide,

indiquant où certaines avancées méthodologiques pourraient être tentées.

Comme nous traitons de la survie des proches, les modèles de simulation doivent être adaptés au suivi des *relations de parenté*. Bongaarts (1983) suggère de classifier les modèles de ce type en trois catégories : les modèles *analytiques*, les modèles de *macrosimulation*, et les modèles de *microsimulation*.

1. Les *modèles analytiques* utilisent des relations mathématiques, souvent exprimées sous forme continue, pour rendre compte de l'évolution du nombre de proches parents - sous forme d'agrégats.
2. La *microsimulation* soumet plutôt une population fictive d'individus à des risques compétitifs de décès, de mariage et de grossesse, et fait transiter ces individus entre différents sous-groupes (parités atteintes, stades de la maladie, etc).
3. Les *modèles de macrosimulation* sont une sorte de compromis entre les modèles analytiques et la microsimulation (Ruggles 1987). Plutôt que de travailler au niveau individuel, les processus démographiques sont modélisés et simulés au niveau agrégé pour différentes sous-populations (par âge, sexe, statut matrimonial, etc.). Comme en microsimulation, ces modèles recourent à un ensemble de transitions entre différents états. Ils reposent généralement sur la méthode des composantes (ou méthode matricielle de projection).

Dans les chapitres suivants, nous recourrons à ces trois types de simulation, en préférant la macrosimulation et les modèles analytiques quand cela est possible. Ces derniers sont en effet nettement moins demandeurs en termes de ressources informatiques. Dans le chapitre 3 notamment, nous détaillerons et nous appliquerons une méthode développée par Grassly et Timaeus (2005) pour estimer le nombre d'orphelins dûs au sida. Cette méthode repose sur la combinaison d'une séries de matrices dont les différentes dimensions renvoient à des âges, des années, et différentes durées. Dans d'autres cas, nous utiliserons la microsimulation, car elle permet de raffiner les calculs. A cet effet, nous avons développé une série de microsimulations exécutables à partir du programme SOCSIM, et basées sur les paramètres démographiques estimés par les Nations Unies pour une quarantaine de pays africains. Cette série de microsimulations est détaillée dans la section 2.3.

Avant d'entrer dans les détails techniques, la section 2.1 présente les avantages de la microsimulation. La section 2.2 fournit ensuite un bref aperçu du programme SOCSIM.

Précisons d'emblée que lorsque la microsimulation est utilisée, l'objectif ne consiste pas à fournir de nouvelles estimations des paramètres démograghiques de base. En effet, les paramètres démographiques qui interviennent sont pour la plupart connus d'avance. Wolf (2001) définit d'ailleurs la microsimulation en ces termes : "microsimulation consists of drawing a sample of realizations of a prespecified stochastic process". La dynamique démographique est décomposée en une série de processus stochastiques qui s'appliquent aux individus repris dans une population

fictive. Les différents événements dépendent de leurs caractéristiques propres (leur âge, leur groupe d'appartenance, leur statut matrimonial, etc.), fixes ou variables dans le temps. La probabilité d'occurrence de chacun de ces événements est délimitée par une série de risques prédéfinis. Le prochain événement qu'un individu va subir est sélectionné par un processus de compétition entre événements qui dépend du temps d'attente avant chacun d'eux. L'idée n'est donc pas d'estimer des taux de mortalité ou de fécondité, qui sont connus, mais plutôt de travailler sur une population fictive issue d'un jeu complexes d'interaction (entre phénomènes démographiques, ainsi qu'entre individus).

Où est l'intérêt d'une telle approche pour notre propos, l'analyse de la mortalité adulte basée sur la survie des parents ou des frères et soeurs ? Simplement, elle permet d'*observer directement les proportions de survivants* au sein des populations finales - plutôt que de les exprimer analytiquement (comme on le ferait par exemple au moyen d'équations qui découlent de la théorie des populations stables). Si l'on accepte de considérer que les microsimulations cadrent bien avec les phénomènes démographiques réellement à l'oeuvre, elles vont procurer des populations finales très utiles. En effet, la cohérence d'ensemble au niveau agrégé est garantie (p.ex. les taux de mortalité simulés sont ceux introduits) mais ces populations finales fournissent des informations sur la variabilité des trajectoires démographiques au niveau individuel. Comme les microsimulations s'appuient sur des *expériences aléatoires répétées* plutôt que sur des proportions moyennes (Van Imhoff et Post 1997), nous pouvons analyser la *dispersion* des phénomènes, ce qui est beaucoup plus difficile en macrosimulation puisque les individus en sont absents.

Par exemple, si la macrosimulation ou les calculs analytiques peuvent répondre à des questions du type : "quelle est la proportion de frères et soeurs décédés par âge dans tel contexte démographique", la microsimulation est bien plus à même de répondre à des questions du type "quelle est la proportion d'individus qui n'ont plus qu'un frère survivant entre 15 et 49 ans dans tel contexte démographique ?". Le nombre moyen de membres du réseau de parenté d'un individu (nombre de cousins, nombre de frères et soeurs, etc.) est facilement calculable, mais le nombre d'individus qui n'ont plus aucun frère ni aucune soeur qui survit l'est nettement moins.

En somme, avec la microsimulation, il devient aisé de rentrer dans des détails pour lesquels il faudrait, si l'on voulait travailler de façon analytique, de subtiles expressions mathématiques et des procédures d'approximations numériques complexes. En ce sens, les microsimulations autorisent davantage de souplesse, et donc davantage de développements. Par exemple, il est facile de lever les hypothèses sur lesquelles reposent les méthodes indirectes : celle d'une indépendance entre les risques de mortalité au sein des familles, ou celle d'une stabilité de la mortalité dans le temps...

Dans la section suivante, nous précisons de façon plus systématique les différents avantages de la microsimulation.

2.1 Avantages de la microsimulation

La distinction entre les modèles analytiques et les deux autres types de modélisations des structures de parenté (la macrosimulation et la microsimulation) est évidente. Les premiers se basent uniquement sur des expressions mathématiques, qui n'ont souvent pas de forme close et qui sont résolues par approximation numérique. Les macro- et microsimulations procèdent différemment : elles simulent de façon dynamique le processus évolutif que connaît une population qui subit des événements démographiques [1].

Dans les deux cas, les événements (décès, naissances, etc.) sont des variables aléatoires associées à une certaine probabilité d'occurrence, qui peut être synthétisée dans une *espérance mathématique*. Mais celle-ci n'a pas le même statut entre macro- et microsimulation. C'est précisément ce qui permet de faire la distinction entre les deux approches :

> "Un modèle macro suppose que la taille de la population est si grande que le nombre projeté d'événements peut être posé comme égal à son espérance mathématique ; un modèle micro suppose que le nombre de répétitions de l'expérience aléatoire dans l'échantillon est si grand que le nombre projeté d'événements attendus sera approximativement égal à son espérance mathématique"(Van Imhoff et Post 1997).

1. On voit là le premier intérêt des méthodes de microsimulation : tout comme les modèles de macrosimulation, il s'agit de modèles probabilistes. Mais *la nature aléatoire du processus y est plus évidente* :

   > "Bien sûr, il serait, en principe, possible de calculer des écarts-types dans un modèle macro, mais cela n'est pratiquement jamais fait, essentiellement parce que les calculs nécessaires sont extrêmement compliqués. Par contraste, la nature aléatoire du processus est explicitement modélisée en microsimulation" (Van Imhoff et Post 1997).

   Puisqu'il se base sur des proportions moyennes, un modèle macro va toujours fournir les mêmes résultats s'il part des mêmes paramètres. Tandis que deux séries de microsimulations ne vont jamais reproduire les mêmes trajectoires démographiques, puisqu'elles se basent sur des expériences aléatoires répétées [2].

   Le caractère aléatoire des microsimulations se situe en fait dans la concurrence qui se joue entre les différents risques afin de décider de la nature du prochain événement (décès, naissance, etc.) [3]. Les événements susceptibles d'arriver à chaque individu sont mutuellement exclusifs. Les taux sont traduits en temps d'attente et l'occurrence du prochain événement est définie par le temps d'attente le plus court. Par exemple, une femme se mariera avant de mettre au monde un nouvel enfant si le temps d'attente avant son mariage est plus

1. Leurs similarités et différences sont bien détaillées par Van Imhoff et Post (1997).
2. A moins qu'elles génèrent des nombres pseudo-aléatoires de la même façon
3. On parlera d'*événement* ou de *transition* pour faire référence aux décès, aux naissances, aux mariages, aux passages d'un groupe à l'autre.

court que le temps d'attente avant sa prochaine naissance. Mais ce mariage va en retour augmenter ses chances d'avoir un enfant, et réduire le temps d'attente avant la prochaine grossesse. En d'autres termes, chaque fois qu'un individu est affecté par un événement (qu'il survienne chez lui - une naissance - ou chez un autre individu - le décès de son partenaire), les nouveaux temps d'attente sont recalculés et un nouvel événement est programmé. Dans ce passage des taux aux temps d'attente, un *nombre aléatoire* doit intervenir.[4] De ce fait, une variation aléatoire - qui peut parfois s'avérer importante - est introduite.[5] Multiplier les simulations en retenant différents nombres aléatoires permet de calculer des marges d'erreur encadrant les estimations. C'est ce que nous ferons notamment lorsque nous examinerons les biais de sélection qui caractérisent les données sur les frères et soeurs.

2. Durant les microsimulations, les différentes variables qui se réfèrent à un individu (sa parité, son statut matrimonial, etc.) varient au cours du temps, et elles sont mises à jour chaque fois qu'un événement affecte cet individu. C'est là que réside le second avantage de la microsimulation : *le nombre de variables qui influent sur les probabilités de décès, de grossesse ou de transition peut être beaucoup plus élevé en microsimulation*, par rapport à la macrosimulation. Pennec (1996) l'explique en ces termes :

> "En simulation matricielle, plus on augmente le nombre d'événements et de conditionnantes, plus les matrices de transition atteignent une taille importante et deviennent difficilement manipulables. En effet, la matrice contient l'ensemble des variables qui conditionnent chaque événement simulé, alors que dans la microsimulation, chaque événement est simulé par une matrice particulière. Cette séparation par événement rend plus aisé l'ajout de conditionnantes supplémentaires puisqu'il n'est nécessaire de recalculer qu'une petite matrice et non l'ensemble des transitions" (p.32).

Parmi les différentes variables qui se réfèrent à un individu, il est possible de retenir les *relations de parenté* : les structures familiales peuvent donc être entièrement reconstruites. Concrètement, quand nous parlons de maintenir les liens de parenté dans les simulations, cela signifie *a minima* que l'identifiant de la mère et du père de chaque enfant est encodé comme une variable des fichiers de résultats. Il est possible de voyager à travers tous les liens de parenté, et même de suivre la séquence des événements qui surviennent au sein d'une même famille.

4. Si pour cette même femme, à un moment donné de son existence, la compétition est forte entre le décès et une grossesse, un certain nombre aléatoire pourra lui permettre de donner naissance avant de décéder, tandis qu'un autre nombre aléatoire ne lui aurait pas permis.

5. En effet, nous nous situons à l'échelle individuelle, et chaque nouvel individu qui naît va potentiellement pouvoir lui-même donner lieu à d'autres naissances, etc.

Il nous faut préciser ici que l'enregistrement des liens de parenté suppose que le modèle soit *fermé* :

- Dans ce cas, aucun individu n'entre dans la population autrement que par la naissance.[6] En cas de mariage, une recherche de partenaires est conduite parmi la population déjà simulée et le premier individu trouvé qui remplit certaines conditions est retenu comme conjoint. Les partenaires appariés conservent chacun leurs référents familiaux. Cela autorise le programme à stocker les liens de parenté de tous les individus introduits dans la simulation, et à les actualiser en cours de route. C'est le cas du programme SOCSIM, que nous détaillons plus loin.
- A l'inverse, dans les modèles de microsimulation *ouverts*, des partenaires doivent être créés en cas de mariage. Mais il faut alors définir des liens de parenté pour ces individus nouvellement arrivés (et créés à partir de rien), ce qui va poser problème.

3. Le troisième avantage de la microsimulation tient au fait que *les interactions entre les processus démographiques* ainsi qu'*entre les individus* peuvent être beaucoup plus facilement pris en compte (Van Imhoff et Post 1997). D'une part, au niveau des processus démographiques, on pourra par exemple introduire des taux de fécondité plus élevées chez les femmes mariées, ou une association positive entre la mortalité et la fécondité. D'autre part, au niveau des individus, la microsimulation permet facilement d'imaginer que des comportements démographiques se "diffusent" d'une génération à l'autre : par exemple, une mère dont la fécondité est élevée peut donner naissance à des filles qui auront elles-mêmes une fécondité élevée. De même, le modèle peut prévoir que des risques de mortalité plus élevés se transmettent d'une génération à l'autre, ou que des risques d'infection au VIH se transmettent verticalement.
4. Enfin, les données générées maintiennent pour chaque individu les *dates d'occurrence des événements démographiques*. Le fichier final prend donc la forme d'un fichier de données longitudinales, sur lequel il est possible d'appliquer des méthodes démographiques classiques. Il est par exemple possible de confronter les résultats obtenus de façon biographique à ceux obtenus de façon indirecte. Les procédures visant à dater les estimations basées sur la survie des proches peuvent également être ré-évaluées. Par exemple, nous verrons que la datation des estimations de mortalité calculées à partir des orphelins suppose une évolution linéaire de la mortalité dans le temps (Brass et Bamgboye 1981, Palloni et Heligman 1985). Quel est l'ampleur du biais introduit par des changements non linéaires ? Enfin, signalons qu'il est possible également de répliquer une situation d'enquête, en tirant au hasard un certain nombre d'enquêtés selon leurs caractéristiques.

Après avoir relevé les différents avantages de la microsimulation, il faut insister sur ses deux principales limites.

6. Sauf les individus de la population de départ, bien entendu...

1. Tout d'abord, la microsimulation recquiert *une large quantité de données*, qui, bien souvent, sont inexistantes ou inaccessibles. Par exemple, mener des simulations sur de longues périodes avec des taux de fécondité variables par statut matrimonial, par âge et par parité, c'est infaisable à partir d'estimations empiriques dans la majorité des situations, et certainement pas envisageable pour les contextes africains. En ce sens, la microsimulation doit s'appuyer sur *la modélisation* des comportements démographiques. En l'absence de tables de mortalité empiriques, par exemple, les taux doivent nécessairement être dégagés de modèles pour lesquels il faut sélectionner une série de paramètres. De même, pour prévoir une mortalité imputable au VIH, il faut s'appuyer sur des modèles relatifs à l'âge à l'infection, à la durée d'incubation, etc. La portée des conclusions dégagées des microsimulations est donc limitée par la qualité de la modélisation des événements démographiques.
2. Une autre limite de la microsimulation est relevée par Ruggles (1993; 1987). Cet auteur critique le fait que les modèles de microsimulation imposent aux individus des *risques démographiques indépendants des risques auxquels font face les membres de leur réseau de parenté*. En d'autres termes, la plupart des modèles supposent que la formation des réseaux de parenté dépend uniquement de la variation aléatoire et des taux introduits[7]. Or, dans les populations réelles, on observe des similarités dans les comportements démographiques des membres d'un même réseau de parenté. Pour le dire autrement, les taux de fécondité ou de mortalité sont corrélés au sein d'une même famille. Dès lors, dans certains groupes, les réseaux de parentés seront systématiquement plus étendus, dans d'autres groupes, systématiquement plus retreints que ne laisserait supposer la seule variation aléatoire. Les simulations vont donc parvenir à restituer correctement les réseaux de parenté *en moyenne* (nombre moyen de petits enfants, etc.), mais *la variation des réseaux de parenté sera sous-estimée par ces simulations* : par exemple, il n'y aura pas assez de grands-parents avec très peu (ou beaucoup) de petits enfants, et il y aura trop de grand-parents avec un nombre moyen de petits enfants. Ceci concerne la plupart des relations de parenté, et notamment la survie des parents[8], si une fraction des risques de décès sont transmis d'une génération à l'autre. Cela concerne également les fratries, comme on le verra plus loin.

   Pour répondre à cette critique, certains programmes de microsimulations introduisent une *hétérogénéité des taux de fécondité et de mortalité*. Il faut de plus prévoir une certaine transmission de cette hétérogénéité au sein du réseau de parenté (*cf.* infra). Ces choix rendent toutefois la calibration du modèle plus délicate. Par

7. Ruggles parle de *Whopper Assumption* pour faire référence au fait que cette hypothèse n'est clairement pas tenable.

8. Contraitement à ce qu'écrit Ruggles (1987) : "The error should not affect the proportion of the population with a living mother, since it is only possible to have one mother and her chance of being alive depends entirely on her own characteristics." C'est vrai pour la mère, mais pas pour l'enfant.

exemple, SOCSIM peut allouer à chaque femme un multiplicateur individuel qui s'applique à ses taux de fécondité. De cette façon, certaines femmes ont une fécondité plus élevée, certaines femmes ont une moindre fécondité, mais la procédure garantit que le niveau moyen de la fécondité est préservé. La difficulté introduite est que la transmission de cette hétérogénéité de la fécondité aux enfants a un impact sur la croissance de la population. En effet, les femmes plus fécondes, qui ont davantage de filles, vont transmettre une partie de leur "sur-fécondité". Leurs filles vont donc elles-mêmes avoir davantage de filles, et transmettre en retour cette plus grande fécondité à leurs enfants. De cette façon, la valeur moyenne de la fécondité s'élève graduellement de générations en générations (Murphy 2006). Il est alors nécessaire d'ajuster à la baisse la fécondité introduite. Ceci complexifie beaucoup le calcul des taux à introduire dans les simulations [9].

Après cette présentation générale de la microsimulation, la section suivante est consacrée au détail du fonctionnement du programme SOCSIM.

2.2 SOCSIM en bref

Initiée dès les années 1960, la microsimulation a connu une expansion importante dans les dernières décennies, notamment grâce à l'accroissement des capacités informatiques. Elle est aujourd'hui utilisée couramment dans les pays occidentaux - généralement pour estimer les effets des politiques publiques, mais les démographes y recourent également [10]. Parmi les différents programmes de microsimulation, certains enregistrent les relations de parenté. C'est le cas des anciens programmes KINSIM (développé au NIDI) et CAMSIM (développé à Cambridge - Smith (1987)), ainsi que de SOCSIM. Ce dernier a été écrit au début des années 1970 à Berkeley par Peter Laslett, Eugene Hammel et Kenneth Wachter, au départ pour analyser l'impact des changements démographiques sur les réseaux de parenté étendus (dans une perspective historique). Il a été constamment retravaillé depuis [11].

Dans le cadre de cette thèse, nous utiliserons SOCSIM, pour trois raisons : (1) il s'agit d'un programme *open source* accessible dans le domaine public, (2) la structure du programme est très intuitive et bien documen-

9. Il y a également un problème sérieux lié aux hypothèses à retenir : comment modéliser l'hétérogénéité des taux de mortalité, par exemple ? Qui transmet (mère, père, partenaire ?), sous quelle forme et à quelles conditions ?Pour répondre à ces questions, il faudrait pouvoir disposer d'éléments empiriques très détaillés sur les réseaux de parenté et la corrélation des risques démographiques au sein de ces réseaux. Or, si ces données existaient, la microsimulation perdrait son intérêt...

10. L'*International Microsimulation Association* liste pas moins de 17 modèles de microsimulations démographiques dynamiques. O'Donoghue (2001) fournit une description de certains d'entre eux, tout comme Pennec et Bacon (2007).

11. On trouvera des descriptions et différentes applications de SOCSIM dans les travaux suivants : Hammel et al. (1976), McDaniel et Hammel (1984), Wachter (1980), Ruggles (1987), Wachter (1997), Wachter et al. (1997; 2002), Jiang (1995), ou Murphy (2004; 2006). Zhao (2006) compare les programmes CAMSIM et SOCSIM.

tée, (3) le code peut être modifié pour rencontrer les objectifs de chaque recherche [12].

Les détails techniques relatifs au fonctionnement de SOCSIM sont expliqués par Carl Mason (2010) [13]. Nous résumons ici dans les grandes lignes comment les microsimulations sont menées par le programme [14]. Tout part d'une population fictive d'individus, et de taux de transitions entre différents états. A partir de ces taux, un événement est programmé pour chaque individu. Comme on l'a indiqué plus haut, les taux sont traduits en temps d'attente et l'occurrence du prochain événement est définie par le temps d'attente le plus court [15]. A tout moment, chaque individu de la simulation a un événement en attente. Ces événements sont exécutés mois après mois, et les uns à la suite de l'autre. Il est nécessaire d'exécuter les événements successivement pour éviter les incompatibilités [16]. Notons que si *Ego* a toujours un événement programmé en attente, les différents événements qui affectent *les autres individus* peuvent avoir un impact sur les taux de transition d'*Ego*. Par exemple, si la mortalité des veufs est différente de celle des mariées, quand le partenaire d'*Ego* décède, un nouvel événement peut lui être reprogrammé.

Pour détailler le fonctionnement de SOCSIM, le plus facile est de présenter les différents fichiers qui interviennent (soit en amont, soit en aval des simulations).

2.2.1 Le fichier de base

Chaque simulation débute à partir d'un *fichier de base* (appelé *supervisory file*). Il indique le nom des fichiers où sont enregistrés la population et les différents taux - et où stocker les résultats. Ce fichier de base définit également les principaux paramètres utilisés par SOCSIM. Parmi ces paramètres, on trouve notamment le nombre et la durée des différents *segments*. Les simulations peuvent en effet être découpées en segments temporels, ce qui permet de mener des simulations dans des contextes

12. Nous avons reçu à plusieurs reprises l'appui technique de Carl Mason, de l'University de Berkeley, qui est le programmeur actuel derrière SOCSIM.

13. Une documentation plus ancienne est également disponible sur www.demog.berkeley.edu/ socsim.

14. Ce dernier se présente sous la forme d'une série de fichiers de code, qui peuvent être compilés pour former un exécutable. SOCSIM peut tourner sous différents systèmes d'exploitation (Unix, mais aussi Windows).

15. "Each event for which the individual is at risk (e.g., men rarely give birth) can be modeled as a piecewise exponential distribution. A random number is used to generate a waiting time until this event occurs (which is bounded by the individual's maximum possible age at death). The individual's next event is the one with the shortest randomly generated waiting time. The event competition thus follows a competing risk framework wherein the probability of each event is independent of all others" - Mason (2010), p.27

16. Par exemple, lorsqu'un individu se marie, il cherche un partenaire dans une liste de partenaires potentiels (*cf.* infra). S'il trouve un partenaire qui lui convient et qu'il l'épouse, ce partenaire est retiré de la liste des partenaires potentiels, et il n'est donc plus accessible pour le prochain mariage programmé.

démographiques changeants. Ce fichier définit également la durée de la simulation, exprimée en mois [17].

2.2.2 Les populations de départ

Les *populations de départ* sont enregistrées dans un fichier qui reprend les différents individus organisés en ligne, avec à minima, leur date de naissance et leur sexe. SOCSIM inclut également l'appartenance à différents *groupes*, pour lesquels des taux spécifiques peuvent être détaillés (des taux de mortalité, de fécondité et de mariage). L'appartenance à ces groupes est transmis aux enfants par l'un des parents (par défaut par la mère). Mais les individus peuvent changer de groupe au cours de leur vie [18]. Les simulations vont alors débuter par une longue période de stabilité, afin d'obtenir des réseaux de parenté plausibles et la structure par âge souhaitée. Ensuite, les simulations sont menées d'un bloc ou par segments. Bien entendu, les populations issues des simulations pouvent être utilisées comme point de départ pour des simulations ultérieures [19].

2.2.3 Les fichiers de taux

Les *fichiers de taux* définissent l'intensité des transitions d'un état à l'autre : de l'état de célibataire à marié, de la survie au décès, d'une parité à l'autre, etc. Cette intensité peut varier en fonction :

- de différents status matrimoniaux (par ex. la fécondité maritale et pré-maritale)
- de l'âge
- du sexe
- de la parité atteinte (dans le cas de la fécondité)
- de l'appartenance à différents groupes (par ex. une mortalité plus forte dans un sous-groupe de population)
- de la durée passée dans un état donné (par exemple, selon la durée de mariage dans le cas d'une transition vers le divorce, ou selon la durée passée au sein d'un groupe dans le cas d'une transition vers un autre groupe - migrations, stades de la maladie, etc.).

Les quatre sections suivantes précisent comment SOCSIM modélise les différentes composantes de la dynamique démographique.

17. Il peut contenir également différents paramètres que nous détaillons plus loin dans le texte : un intervalle minimal entre naissances, une option introduisant une hétérogénéité dans les taux de fécondité, une série de paramètres définissant les préférences guidant le choix des partenaires, etc.

18. Dans le fichier de départ, un paramètre définit également l'intensité de l'endogamie, au cas où le mariage entre groupes est permis. SOCSIM a également besoin d'un fichier reprenant les différents mariages, dont la structure est comparable au fichier d'individus. Généralement, ce fichier est vide au départ des simulations, car les individus sont célibataires et non apparentés entre eux.

19. Ceci permet par exemple de faire un recensement des survivants au terme de chaque segment, ou de redistribuer automatiquement des individus dans certaines catégories.

La fécondité

La fécondité peut varier par groupe, âge, statut matrimonial et parité atteinte. Elle est par contre fixe au sein d'un même segment. Elle est modélisée sous forme "démographique", et pas "biologique", pour reprendre une distinction citée par Pennec (1996). En d'autres termes, plutôt que de tenir compte du cycle biologique de la fécondité et/ou des actes sexuels en soi, le programme part de taux de fécondité.

Si le recours à des taux de fécondité par âge peut s'avérer satisfaisant en macrosimulation, il faut aller au-delà pour mener des microsimulations. En effet, à la différence des macrosimulations, *la plausibilité des microsimulations ne s'évalue pas uniquement au niveau agrégé*. Il faut maintenir une certaine cohérence entre, d'une part, les événements qui adviennent au niveau individuel et, d'autre part, ce qui en résulte au niveau de la population. C'est particulièrement le cas en matière de fécondité, car il s'agit d'un événement renouvelable. D'un côté, il faut faire en sorte que les simulations reproduisent *en moyenne* la fécondité qui correspond aux taux introduits : ISF, nombre total de naissances, nombre moyen d'enfants nés vivants, etc. De l'autre, il faut également que la *distribution* des événements entre les individus soit plausible ; distribution du nombre d'enfants nés vivants, intervalles entre naissances, etc. C'est la raison pour laquelle les seuls taux de fécondité par âge ne sont pas suffisants. Introduits comme tels dans les simulations, ces derniers vont générer des événements qui cadrent avec la fécondité *moyenne*, mais qui sous-estiment sa *dispersion*. Par exemple, le nombre moyen d'enfants nés vivants sera approximativement 'correct' [20] mais la dispersion autour de la moyenne sera sous-estimée. Par rapport à une situation réelle, il n'y aura pas assez de femmes sans enfants, pas assez de femmes avec beaucoup d'enfants, tandis qu'il y aura trop de femmes avec un nombre moyen d'enfants. Nous retrouvons ici l'observation faite par Ruggles (1987). Elle a de nombreuses implications, notamment lorsqu'il s'agit d'examiner la dispersion des phénomènes démographiques entre les individus, ou d'analyser l'évolution des réseaux de parenté. Deux cas de figure où, précisément, la microsimulation est souvent utilisée. Dans les deux cas, le recours aux taux de fécondité par âge est problématique :

> "For models designed to reveal the availability of kin (...), the use of age-specific rates of fertility produces unacceptable inaccuracy. We are not primarily concerned with the mean number of available kin of a given type ; rather, we are interested in the percentage of persons who had any kin of a given type available for coresidence. The frequency distribution of the number of kin available is therefore just as important as the central tendency" - Ruggles (1987), p. 164.

Il faut donc travailler de façon plus fine qu'avec les seuls taux de fécondité, pour par exemple distribuer les naissances en fonction de la parité atteinte, ou tenir compte de la "fécondabilité" des femmes. SOCSIM

20. Au sens où il va correspondre à celui qui pourrait être calculé à partir des taux introduits (Keyfitz et Caswell 2005).

offre différentes possibilités. Afin d'examiner ces différentes options, nous avons comparé cinq séries de simulations, toutes basées sur les mêmes taux de fécondité et de mortalité, mais en modifiant la façon avec laquelle la fécondité est introduite dans ces simulations. Nous renvoyons la discussion de ces différentes options à l'annexe A.2.1, page 394. La conclusion de cette discussion est qu'il apparaît essentiel de tenir compte de *la variation de la fécondité par parité*, surtout si les microsimulations doivent apporter des informations qui vont au-delà du nombre "moyen" d'enfants nés vivants, ou du nombre moyen de frères et soeurs. Nous reviendrons sur ce point dans la section 2.3.3.

La mortalité

La *mortalité* est variable par groupe, par âge, et par statut matrimonial. Contrairement aux autres phénomènes démographiques, son intensité peut varier au sein du segment car SOCSIM peut automatiquement générer des taux à partir des paramètres du modèle de Lee et Carter (1992). Ce modèle va résumer l'évolution de la mortalité en trois paramètres : $K(t)$, a_x et b_x. Ceci est détaillé en annexe A.2.2 (p. 406). Ces trois paramètres sont ensuite directement introduits dans SOCSIM, qui va recalculer les taux.

La nuptialité

Lorsqu'un *mariage* est programmé pour un individu, SOCSIM lance une recherche de partenaires parmi les individus de sexe opposé disponibles (le *marriage market*). Le mariage n'est conclu que lorsqu'un de ces "candidats" rencontre les conditions définies par une fonction de préférence. Par défaut, celle-ci interdit les écarts d'âge longs (plus de 10 ans entre les partenaires) et interdit le remariage de couples divorcés, par exemple. Un score d'assortiment peut être défini pour introduire des conditions supplémentaires. Le partenaire retenu est tiré au hasard parmi la liste des partenaires qui remplissent les conditions nécessaires. SOCSIM prévoit cinq catégories de statut matrimonial : célibataire, marié, divorcé, cohabitant, ou veuf[21].

Les transitions entre groupes

Comme on l'a indiqué, les simulations peuvent être menées en distinguant plusieurs sous-groupes. L'appartenance à ces groupes est acquise à la naissance et transmise par l'un des deux parents. Mais elle peut changer au cours de la vie, selon des taux de transitions. Ces taux varient

21. Pour la cohabitation, une fraction (variable avec l'âge) des unions formées est considérée comme donnant lieu à une cohabitation plutôt qu'à un mariage. Les couples en cohabitation sont sujets à la transformation de cette cohabitation en mariage ainsi qu'au risque de rupture de la cohabitation. Pour le divorce, l'événement qui survient chez les femmes doit survenir dans le même temps chez les hommes. Le divorce est donc considéré comme un événement qui concerne le mariage plutôt que les mariés. Chaque mariage est exposé deux fois au risque de divorce, une fois via le partenaire féminin, une fois via le partenaire masculin, et la probabilité de divorce est simplement divisée par deux. A ce stade, le divorce n'est pas intégré dans nos simulations.

soit selon l'âge, soit selon la durée passée au sein d'un groupe.

Bien entendu, il n'est pas nécessaire de spécifier des taux pour chaque sous-groupe/parité/sexe/état matrimonial, et il existe une série de règles qui attribuent des taux par défaut [22].

2.2.4 Les populations finales

Ces dernières contiennent tous les individus qui étaient présents dans la population de départ et ceux qui ont été générés au cours de la simulation, qu'ils soient décédés ou non. Elles reprennent pour ces individus : leur numéro d'identifiant, leur sexe, leur date de naissance, les numéros d'identification des membres de leur famille, le numéro de leur dernière union, leur statut matrimonial, ainsi que leur éventuelle date de décès. Pour chaque individu, SOCSIM enregistre toujours l'identifiant de sa mère, mais il ne retient l'identifiant de son père que quand la naissance de l'enfant est survenue au sein d'un mariage. C'est là une des limites principales du programme ; seuls les pères mariés sont enregistrés.

Les fichiers relatifs aux mariages ont la même structure que les fichiers reprenant les populations, à ceci près que les entitées listées sont les différentes unions (cohabitations ou mariages) [23].

L'historique des transitions que font les individus (quand ils passent d'un sous-groupe de population à l'autre) est également enregistrée dans un fichier spécifique [24].

***

Après avoir précisé comment fonctionne SOCSIM, nous détaillons dans la section suivante le modèle de microsimulations que nous avons construit pour reproduire la dynamique démographique d'un grand nombre de pays africains.

2.3 Paramètres démographiques introduits

Dans le cadre de cette recherche, nous cherchons à construire des populations fictives qui reflètent les caractéristiques démographiques des pays de l'Afrique subsaharienne. Les microsimulations sont paramétrisées de façon à ce qu'au niveau agrégé, elles reproduisent les estimations des Nations Unies (Révision 2008) [25] Nous élaborons différents scénarii, pour

22. Par exemple, SOCSIM utilise les taux de mortalité du groupe 7 si ces taux ne sont pas spécifiés pour le groupe 8.

23. Un même individu peut donc y apparaître plusieurs fois, s'il a connu plusieurs mariages, ou s'il a d'abord cohabité avant de formaliser son union.

24. Ce fichier reprend l'identifiant des individus, la date de chaque transition, le groupe de départ, le groupe d'arrivée, et l'ordre des transitions.

25. Le modèle de microsimulation présenté ici se base partiellement sur des données *non publiées* utilisées par la Division de la Population des Nations Unies pour établir leurs projections de population. Toutefois, les résultats repris dans le texte contiennent exclusivement des estimations distribuées dans le domaine public. C'est la raison pour laquelle

chacun des pays de la région. Plus précisément, quarante-et-un scénarii sont élaborés, car une série de pays ne sont pas retenus. Par exemple, le Rwanda et la Sierra Leone sont écartés car leurs dynamiques démographiques, fortement perturbés par les conflits, échappent aux tentatives de modélisation [26]. Pour chacun des 41 pays restants [27], une série de taux de mortalité, de fécondité, et de risques associés au VIH-sida sont définis. La migration n'est pas intégrée dans les simulations.

Dans tout exercice de microsimulation, un équilibre doit être trouvé entre d'une part, le degré de complexité des populations dont le modèle rend compte, et d'autre part, l'exigence de le paramétriser à partir d'estimations empiriques qui soient à la fois disponibles et plausibles (Clark 2006) [28]. Précisons à ce propos que notre objectif n'est pas de "répliquer" les trajectoires démographiques spécifiques des différents pays. Il s'agit uniquement de calibrer les microsimulations à partir d'une large gamme de paramètres plausibles, afin d'obtenir des populations virtuelles qui reflètent la diversité du continent.

Comme il s'agit de cadrer avec l'évolution des différentes populations africaines, il faut intégrer des risques d'infection au VIH-sida et tenir compte des multiples conséquences de l'épidémie sur la dynamique démographique. Les difficultés associées à la paramétrisation se concentrent surtout à ce niveau.

1. D'une part, utiliser SOCSIM pour simuler des contextes perturbés par le VIH-sida est en fait assez complexe, simplement parce que *le programme n'intègre pas, à la base, de risques de transmission et de décès spécifiques au VIH-sida*. Il doit donc être adapté, ce qui a été rarement fait. Nous présentons ici l'une des premières tentatives visant à mener des microsimulations avec SOCSIM dans des contextes perturbés par le sida [29].

certains indices de mortalité ou de fécondité par âge ne sont présentées qu'à partir de 1980 pour les pays affectés par le VIH-sida et à partir de 1995 pour les autres.

26. Les Seychelles, Sao Tome, la Réunion, l'Île Maurice et Mayotte sont éliminés, parce qu'ils comptent peu d'habitants et que leur histoire démographique les démarquent du continent. Le Soudan et la Mauritanie sont considérés ici comme appartenant à l'Afrique du Nord, puisque nous adoptons la classification UNECA. *Cf.* annexe A.1.1

27. Afrique du Sud, Angola, Bénin, Botswana, Burkina Faso, Burundi, Cameroun, Cap Vert, Centrafique, Comores, Congo, Congo RDC, Côte d'Ivoire, Djibouti, Erythrée, Ethiopie, Gabon, Gambie, Ghana , Guinée, Guinée-Bissau , Guinée Equ., Kenya, Lesotho, Liberia, Madagascar, Malawi, Mali, Mozambique, Namibie, Niger, Nigeria, Ouganda, Sénégal, Somalie, Swaziland, Tanzanie, Tchad, Togo, Zambie, Zimbabwe.

28. Plus le modèle intègre des spécificités au niveau individuel, plus il devient difficile de le paramétriser de sorte qu'au niveau agrégé, il reproduise les dynamiques démographiques observées. Par exemple, si les taux de fécondité sont distribués par rang et par statut matrimonial, rien ne garantit que les niveaux de fécondité recalculés à partir des populations finales vont cadrer avec l'ISF total estimé. Il faudra donc notamment tenir compte de la distribution des femmes par âge, rang et statut matrimonial pour calculer des taux qui permettront de retomber sur les taux de fécondité toutes parités confondues. De même, dans la recherche de cet équilibre, plus nous cherchons à raffiner le modèle, plus il devient difficile de se baser sur des estimations empiriques propres à chaque pays.

29. Wachter et al. (2002; 2003) ont déjà utilisé SOCSIM pour estimer l'impact de l'épidémie sur les structures de soutien en Thaïlande, mais leur prise en compte du VIH-sida restait statique, et partielle. Les risques de transmission verticale n'y étaient pas pris en

Dans la pratique, il faut travailler en distinguant de nombreux sous-groupes de population. Différents taux de transitions peuvent alors être spécifiés (transition du groupe des séropositifs aux séronégatifs, de l'infection au stade clinique, survie au sein du stade clinique, transmission des mères aux enfants, survie des enfants infectés, etc.). Dans certains groupes, les individus restent exposés à la mortalité sans sida, en plus de risques de décès spécifiques (les individus sous traitement, par exemple), dans d'autres cas, la survie est modélisée de façon à ce qu'elle intègre les décès toutes causes.

Ces difficultés techniques sont surmontables de façon relativement aisée grâce à la souplesse offerte par SOCSIM. Par exemple, SOCSIM peut faire varier les taux de transitions en fonction de l'âge des individus infectés, ou alternativement avec la durée depuis l'infection. La section 2.4 (pp. 75-96) détaille les stratégies adoptées pour intégrer ces différentes transitions au sein des microsimulations.

2. La difficulté principale de l'intégration du VIH-sida dans les simulations tient à la nécessité de disposer d'*estimations au niveau national, désagrégées par age et par sexe* :

   – Ceci est surtout problématique pour les données d'incidence. L'UNAIDS ne distribue que des estimations de prévalence[30]. Nous disposons des estimations d'incidence utilisées par la Division de la population des Nations Unies[31], mais sous une forme non répartie par age, ni par sexe. Il faut donc recourir à des schémas théoriques d'infection par âge et par sexe pour identifier les individus qui vont être infectés.

   – Une fois les risques d'infection introduits dans les simulations, il faut également modéliser la survie depuis l'infection, et pour cela, à nouveau, recourir à des modèles théoriques.

   – Nous souhaitons nous approcher au plus près des estimations des Nations Unies, pour garantir une cohérence d'ensemble, et faire en sorte qu'au niveau agrégé, nos microsimulations collent au plus près des projections des Nations Unies. Mais la Division de la Population n'a plus publié l'annexe méthodologique à leurs projections depuis la révision 2004. Entre-temps, la modélisation de l'épidémie a sensiblement évolué, et nous ignorons certaines des options méthodologiques prises récemment par la Division de la Population.

compte, par exemple. Plus récemment, E. Zagheni vient de terminer une thèse à l'université de Berkeley, où SOCSIM a été également aménagé pour intégrer le VIH-sida, dans une perspective proche de la nôtre (Zagheni 2010).

30. Les dernières estimations sont disponibles via http ://data.unaids.org. Jusqu'à la révision 2007, l'UNAIDS ne publiait pas des tendances mais uniquement la dernière estimation

31. Données non publiées - Source : United Nations, Department of Economic and Social Affairs, Population Division : World Population Prospects DEMOBASE extract. 2009.

– Il subsiste également de nombreuses inconnues, notamment autour de la distribution par sexe de l'incidence. Celle-ci n'est jamais directement observée, or elle joue un rôle clé dans la modélisation du VIH-sida. Les choix retenus vont par exemple avoir un impact important sur l'ampleur des hausses de mortalité et sur leur timing.

– Par ailleurs, les Nations Unies ne modélisent pas la survie des personnes infectées selon l'âge à l'infection. Celui-ci a pourtant un impact significatif sur la survie des séropositifs (*cf.* infra), et donc sur les proportions d'orphelins par âge, notamment.

Pour ces différentes raisons, dans les pays fortement affectés par le VIH-sida, nos estimations de mortalité adulte s'écartent parfois de celles des Nations Unies, surtout chez les hommes (*cf.* infra). Le timing et l'ampleur des hausses de mortalité qui sont simulées peuvent différer des estimations "avec sida" des Nations Unies. Mais quitte à sacrifier un peu la fidélité aux estimations onusiennes, nous avons privilégié le recours aux paramètres les plus récents, notamment en matière de survie des personnes infectées.

Les simulations sont menées entre 1850 et 2010. Bien sûr, les analyses des chapitres suivants ne seront menées sur les populations simulées qu'à partir des années 1980. La période 1850-1950 est en quelque sorte une période de *burn-in* ; elle ne sert qu'à lancer les simulations. Cette période permet de mettre les choses en place, notamment de constituer des réseaux de parenté pour tous les individus survivants, et de distribuer les femmes en différents niveaux de parité atteinte [32].

Pour cette période 1850-1950, nous devons faire l'hypothèse de la *stabilité de la population*. Il n'existe en effet aucune série de paramètres démographiques couvrant les différents pays africains pour cette période [33]. Concrètement, les paramètres démographiques de la période 1950-1955 (c'est-à-dire les premiers disponibles via les Nations Unies) sont retenus pour 1850-1950. Cette approximation est grossière, car l'on sait notamment que la croissance de la population a sans doute diminué durant les premiers temps de la colonisation (en raison de déplacements de population et de fortes mortalité). Certaines populations ont également connu des croissances plus rapides, autour des années 1920 notamment (1ers reculs de mortalité). Mais cette hypothèse de stabilité permet de garantir que *la structure par âge de la population simulée en 1950-1955 soit identique à*

32. Lorsque les simulations incluent des taux de nuptialité, cette période de *burn-in* a également son importance ; les différences d'âge aux mariage et les règles qui encadrent le choix des partenaires peuvent générer des schémas par âge de nuptialité implausibles dans les premières années (*cf.* annexe A.2.7, p. 422).

33. Les travaux de démographie historique africaine offrent rarement la possibilité de dégager des estimations de mortalité générale. Au mieux, nous disposons de quelques informations sous forme de quotients de mortalité des enfants, ou sous forme de taux bruts de mortalité (Brass et al. 1968, Coale 1966, Centre of African studies 1981). Au niveau de la fécondité, nous avons également très peu d'informations sur les tendances avant 1950 (Garenne 2008).

celle estimée par les Nations Unies. Ceci minimise les perturbations liées au fait que les taux de fécondité ou de mortalité de la première moitié du siècle pourraient se "réverbérer" dans la structure par âge au cours des générations suivantes.

Pour chacun des pays, 71 segments temporels sont spécifiés pour mener les simulations :

- Le premier segment dure 50 ans. Il couvre la période de 1850 à 1900. Durant cette période, la stabilité de la population prévaut. Comme nous l'avons précisé, les taux de mortalité et de fécondité sont ceux estimés par les Nations Unies pour la période 1950-1955. La fécondité ne varie pas par parité.
- De 1900 à 1950, les simulations sont menées sur des segments de 5 ans. A partir de là, la fécondité commence à varier par parité. Les taux de fécondité par parité sont établis selon un modèle construit à partir des enquêtes EDS (*cf.* infra).
- A partir de 1950, les différentes populations se conforment aux estimations de l'ONU et les taux de fécondité et de mortalité varient par année d'âge et par année calendrier.
- A partir de 1975, et jusqu'en 2010, le VIH est intégré explicitement dans les simulations, dès que l'incidence est différente de zéro. Au terme de chaque année, nous distribuons les risques d'infection par âge et par sexe en tenant compte de la structure par âge de la population non infectée observée dans les simulations.

Dans les sections suivantes, nous détaillons les paramètres démographiques utilisés pour construire ces différents scénarii. La section 2.3.1 présente les populations de départ et les évolutions de la structure par âge au cours des simulations. La section 2.3.2 est consacrée à la mortalité sans sida. Les taux de fécondité et de nuptialité sont détaillés dans les sections 2.3.3 et dans l'annexe A.2.7. Comme le programme SOCSIM n'est pas conçu à l'origine pour intégrer le VIH-sida dans les simulations, la section 2.4 consacre quelques pages à la stratégie adoptée ici pour incorporer les risques de transmission et l'impact de l'infection sur les trajectoires de vie.

2.3.1 Taille et structures des populations

Populations stables de départ

Les microsimulations débutent en 1850 avec une population de départ, formée d'individus dont on connaît uniquement les dates de naissance et le sexe. Les individus de cette population d'origine ne sont pas reliés entre eux par des liens de parenté.

Afin de trouver un compromis entre la rapidité du calcul et la variation aléatoire[34], les effectifs finaux en 2010 sont fixés à environ 30 000

34. Il faut partir d'une population de départ suffisamment conséquente. Mais certains régimes de mortalité faible combinés à une fécondité élevée peuvent provoquer une croissance rapide de la population. Ces populations qui augmentent très rapidement vont ralentir les calculs et compromettre la capacité du programme à réaliser les microsimula-

personnes. En fonction des taux de croissance estimés, la taille de la population de départ est donc spécifique à chaque pays. Pour chacun d'eux, les différents scénarii sont répliqués dix fois, et les différentes populations atteintes sont combinées en une population unique comptant 300 000 survivants (auxquels s'ajoutent tous les décédés).

Afin d'atteindre 30 000 personnes survivantes par simulation, les taux de croissance doivent être estimés entre 1850 et 2010. Jusqu'à 1950, comme les taux de fécondité et de mortalité sont connus (ce sont ceux de 1950-1955), *le taux de croissance intrinsèque* de la population stable correspondante peut être estimé en appliquant la démarche itérative suggérée par Coale (1957). L'idée consiste à affiner la valeur du taux r en différentes étapes, jusqu'à ce que l'équation de Lotka se vérifie (Preston et al. 2001) [35] :

$$\int_{\alpha}^{\beta} e^{-r(a)} \times_a p_0 \times m(a)da = 1 \tag{2.1}$$

Appliqué aux différents pays africains à partir des paramètres de 1950-1955, ce calcul procure des valeurs de 2.37% en moyenne, avec des valeurs de r variant de 0.56 % (au Gabon) à 3.4% (au Kenya). Ces taux de croissance sont combinés avec les schémas de mortalité de 1950-1955 pour calculer la *structure par âge stable correspondante*, c'est-à-dire la distribution par âge qui est utilisée pour générer la population au début des simulations [36].

Ces *taux de croissance intrinsèques* (pour 1850-1950) sont ensuite combinés aux *taux de croissance naturelle* estimés par les Nations Unies entre 1950 et 2010 afin de calculer la taille de la population de 1850 nécessaire pour obtenir, en 2010, un effectif d'environ 30 000 survivants [37]. Dans la grande majorité des cas, il y a une bonne continuité entre ces taux intrinsèques (1850-1950) et les premières valeurs des taux de croissance naturelle estimés par les Nations Unies (1950-1955). On pourra l'observer en examinant

tions. Avec les populations très nombreuses, l'exploitation des fichiers de résultats devient également assez laborieuse.

35. Où r correspond au taux de croissance intrinsèque, $m(a)$ au taux de fécondité à l'âge a et ${}_ap_0$ à la probabilité de survie de la naissance à l'âge a

36. Le taux brut de natalité se calcule comme ceci :

$$b = 1/\sum_{a=0.1}^{\omega} e^{-r(a+0.5)} \frac{{}_1L_a}{l_0}$$

La structure par année d'âge de la population stable équivalente est calculée comme ceci (Preston et al. 2001) :

$${}_1c_a = be^{-r(a+0.5)} \frac{{}_1L_a}{l_0}$$

Dans ces deux équations, ${}_1c_a$ correspond à la proportion d'individus d'âge a, b au taux brut de natalité, r au taux de croissance, ${}_1L_a$ au nombre d'années vécues par les tous les individus de la cohorte au cours de l'intervalle d'âge a à $a+1$, et l_0 à la racine de la table.

37. En supposant une croissance exponentielle et en calculant P_{1850} comme

$$P_{1850} = 30000/exp(r_{int} * 50 + \sum_{i=1953,5}^{2003} (r_i * 5) + r_{2008} * 2) \tag{2.2}$$

où r_{int} correspond au taux intrinsèque et $r_{1953,5}$ au taux de croissance naturelle estimée par les Nations Unies pour la période 1950-1955. (Il est noté $r_{1953,5}$ car il est centré sur le 1er janvier 1953 et il s'étend sur 5 ans.)

les valeurs reprises dans le tableau présenté en annexe A.2.3. Dans trois quarts des pays, la différence entre le taux intrinsèque et le taux estimé pour 1950-1955 s'élève à moins de 0.25% de taux de croissance [38]. Ceci indique que *la plupart des populations africaines peuvent bien être considérées comme stables à la veille de 1950* [39].

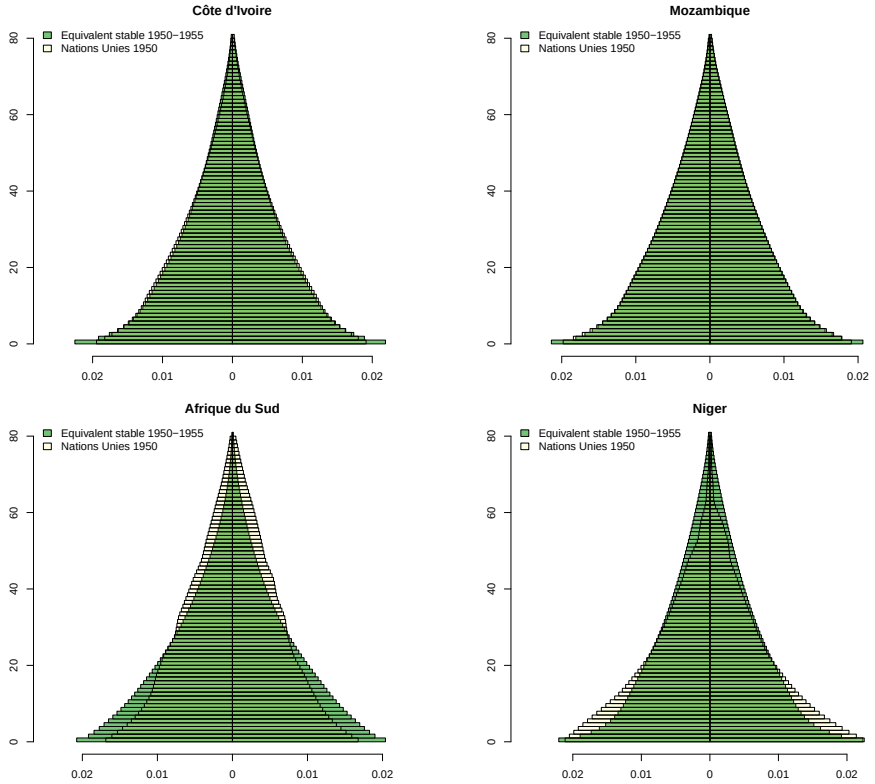

FIGURE 2.1 – *Populations stables équivalentes et populations estimées par les Nations Unies pour 1950 - Côte d'Ivoire, Afrique du Sud, Mozambique et Niger*

Pour trois pays sur quatre, les structures stables équivalentes sont donc proches de celles estimées par les Nations Unies pour 1950, comme on l'observe sur le graphe 2.1, avec le cas de la Côte d'Ivoire et du Mozambique. Dans ces cas là, les structures par âges des populations simulées entre 1850 et 1950 sont donc quasiment identiques à celles estimées par les Nations Unies en 1950.

Par contre, dans un pays sur quatre, les deux types de taux de croissance (ceux qui correspondent à la population stable et ceux qui sont estimés par l'ONU) diffèrent de plus de 0.25%. Dans ces cas là, l'hypothèse de la stabilité de la population en 1950 est donc peu recevable. Ceci se traduit par un décalage entre les différentes structures par âge. Les discontinuités les plus prononcées sont observées en Afrique du Sud et au Bénin (avec respectivement 0.51 et 0.53% de différence entre les taux de croissance). Ces différences vont rendre caduc le calcul de la taille de la population de départ en 1850 [40]. Dans ces quelques cas, les structures par

38. Qu'il s'agisse de différences positives ou négatives

39. Alternativement, on peut supposer que la reconstitution par les Nations Unies de la dynamique démographique des différents pays africains jusqu'à 1950, fait intervenir pour de nombreux d'entre eux une hypothèse de stabilité pour les premières décennies.

40. Car la croissance des populations va être influencée par la désadéquation entre les

âges des microsimulations dans le courant des premières décennies qui suivent 1950 seront légèrement différentes de celles estimés par les Nations Unies. On l'observe dans le graphe 2.1, où la structure par âge stable sud-africaine correspondante aux paramètres de 1950-55 a une base plus large que celle estimée par les Nations Unies pour 1950. C'est l'inverse au Niger : le taux de croissance de la population stable équivalente est plus faible que le taux estimé par les Nations Unies pour 1950-1955, et la base de la structure par âge estimée est plus large.

Le choix d'opter pour une stabilité sur la période 1850-1950 dans les microsimulations offre donc l'avantage de garantir la continuité des paramètres de fécondité et de mortalité au milieu du siècle. Mais ce choix induit également, dans une minorité de pays, des structures par âges légèrement différentes de celles estimées par les Nations Unies autour de 1950. Ceci ne devrait toutefois pas affecter les résultats présentés dans les chapitres suivants, dans la mesure où lorsque nous travaillerons sur les populations simulées, ce sera essentiellement à partir de 1980.

Populations simulées

Le graphe 2.2 présente, pour six pays parmi les 41 pays retenus, les taux de croissance recalculés par année à partir des simulations. Pour la première moitié du XX^{e} siècle, ces taux correspondent simplement aux taux de croissance intrinsèques de la population stable équivalente aux paramètres de 1950-1955 (avec une certaine variabilité aléatoire). A partir de 1950, les taux issus des simulations sont comparés aux *taux de croissance naturelle* estimés par les Nations Unies. Dans la plupart des pays, on observe *une bonne adéquation entre les différentes valeurs*, comme nous l'avons annoncé plus haut. C'est le cas ici du Bénin, du Congo, du Botswana et du Ghana. En Afrique du Sud, les taux observés sur la période 1950-1980 sont plus élevés que les taux de croissance naturelle, ce qui, à nouveau, est lié au fait que la structure de la population au milieu du siècle était différente de la structure stable équivalente (*cf.* graphe 2.1). Mais ceci se résorbe à partir des années 1980. On l'a souligné, de tels décalages ne concernent qu'une petite minorité de pays et ils ne devraient pas avoir d'impact sur la suite.

Dans les pays fortement affectés par le VIH-sida (le Botswana et l'Afrique du Sud sur ce graphe), la comparaison entre les taux estimés par les Nations Unies et les taux recalculés à partir des microsimulations constitue un *premier type de validation des microsimulations*. En effet, à partir de 1975, nous avons greffé l'impact démographique du VIH-sida sur les estimations "sans sida" des Nations Unies (*cf.* infra). de plus, l'intégration du VIH-sida dans les simulations est partiellement indépendante des estimations des Nations Unies (par exemple, les courbes de survie sont différentes - *cf.* infra). Dès lors, lorsque nous comparons les taux de croissance simulées avec les taux de la variante "avec sida" des Nations Unies, nous pouvons déjà apprécier si l'intégration du VIH-sida dans les

structures par âge stables équivalentes et les structures estimées à partir de 1950. Comme les différences les plus importantes sont généralement dûes à des taux intrinsèques plus élevés, les populations finales font excéder 30 000 personnes.

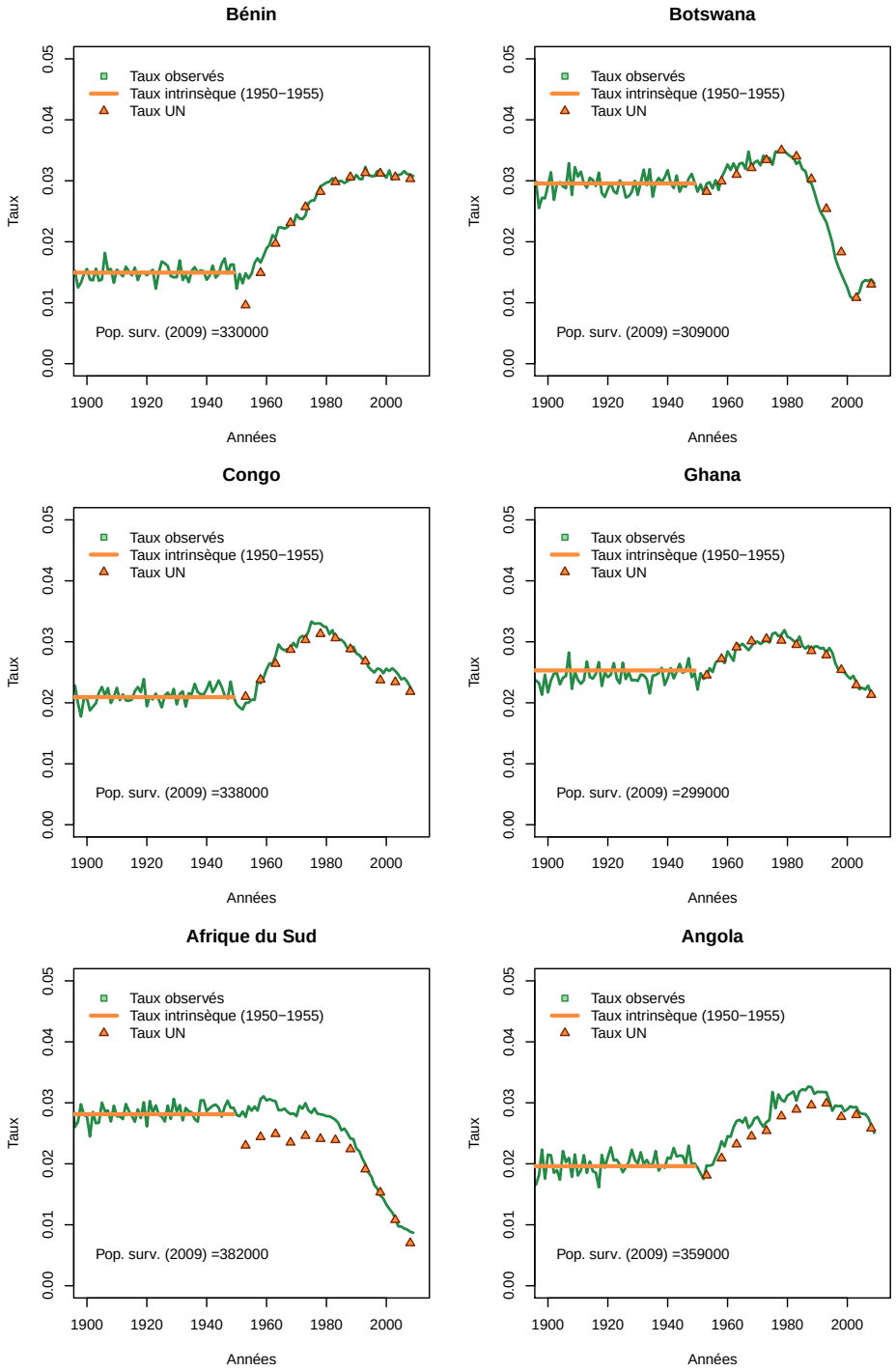

FIGURE 2.2 – *Comparaison des taux de croissance observés avec les taux intrinsèques de la population stable équivalente aux paramètres de 1950-1955 et estimés par les Nations Unies.*

simulations se fait de façon cohérente.

Ce premier examen est prolongé par une comparaison des structures par âge. Le graphe 2.3 (p. 65) compare, pour 1950, 1975 et 2009, la *répartition par âge observée dans les microsimulations* et celle *estimée par les Nations Unies* (toujours dans la variante avec sida). Cinq pays illustrent la diversité des situations : le Cameroun, l'Angola, l'Afrique du Sud, le Botswana et la Côte d'Ivoire. La bonne concordance observée à partir de 1975 entre les deux séries de structures par âge indique que les microsimulations reflètent bien les dynamiques démographiques des différents pays [41]. C'est particulièrement le cas ici du Cameroun et de l'Angola. En Afrique du Sud, nous avons déjà noté que les structures par âge diffèrent davantage en 1950 (en raison de notre hypothèse de stabilité) mais au terme des simulations, les deux répartitions par âge sont très proches l'une de l'autre. L'impact démographique du VIH-sida est manifeste en Afrique du Sud et au Botswana. Il faut préciser également que la migration n'est pas intégrée dans les microsimulations, ce qui peut avoir un impact sur la structure par âge [42].

2.3.2 Niveaux et structures de la mortalité "sans sida"

Nous détaillons dans cette section les niveaux et les schémas par âge de la mortalité "sans sida". Les taux introduits sont présentés, puis comparés aux taux recalculés à partir des microsimulations.

Taux "sans sida" introduits dans les simulations

Pour introduire des *risques de décès par âge*, nous utilisons les tables de mortalité estimées par la Division de la Population pour élaborer la révision 2008 des projections de population (UN 2009c) [43]. La mortalité estimée pour 1950-1955 est utilisée pour la période avant 1950. Ensuite, à partir des années 1975, pour la majorité des pays africains, l'impact du VIH-sida doit être intégré sur les niveaux et les structures de la mortalité. Nous imposons à tous les individus des *risques de mortalité "sans sida"*, puis nous intégrons des probabilités de transition vers l'infection, puis vers le stade clinique du sida, et enfin vers le décès. Sur toute la période 1950-2010, nous utilisons donc les tables de mortalité de la variante "sans sida" construite par les Nations Unies (Encadré 4), avant d'ajouter des risques de décès imputables au sida.

41. En tous cas telles qu'elles sont estimées par les Nations Unies.

42. C'est la raison pour laquelle la structure par âge simulée au Botswana en 1975 diffère de celle estimée par les Nations Unies : en raison des émigrations importantes entre 1950 et 1975, la répartition par âge en 1975 compte comparativement moins d'adultes que dans les simulations. C'est l'inverse en Côte d'Ivoire, où on observe en 1975 davantage d'adultes (surtout chez les hommes) que dans les simulations.

43. Ces tables sont distribuées dans les CD-ROM à partir de 1980 pour les pays affectés par le VIH-sida et à partir de 1995 pour les autres pays UN (2009d). Nous utiliserons les données non publiées à partir de 1950. Pour cette raison, nous ne présentons dans le texte aucun taux de mortalité par âge avant 1980.

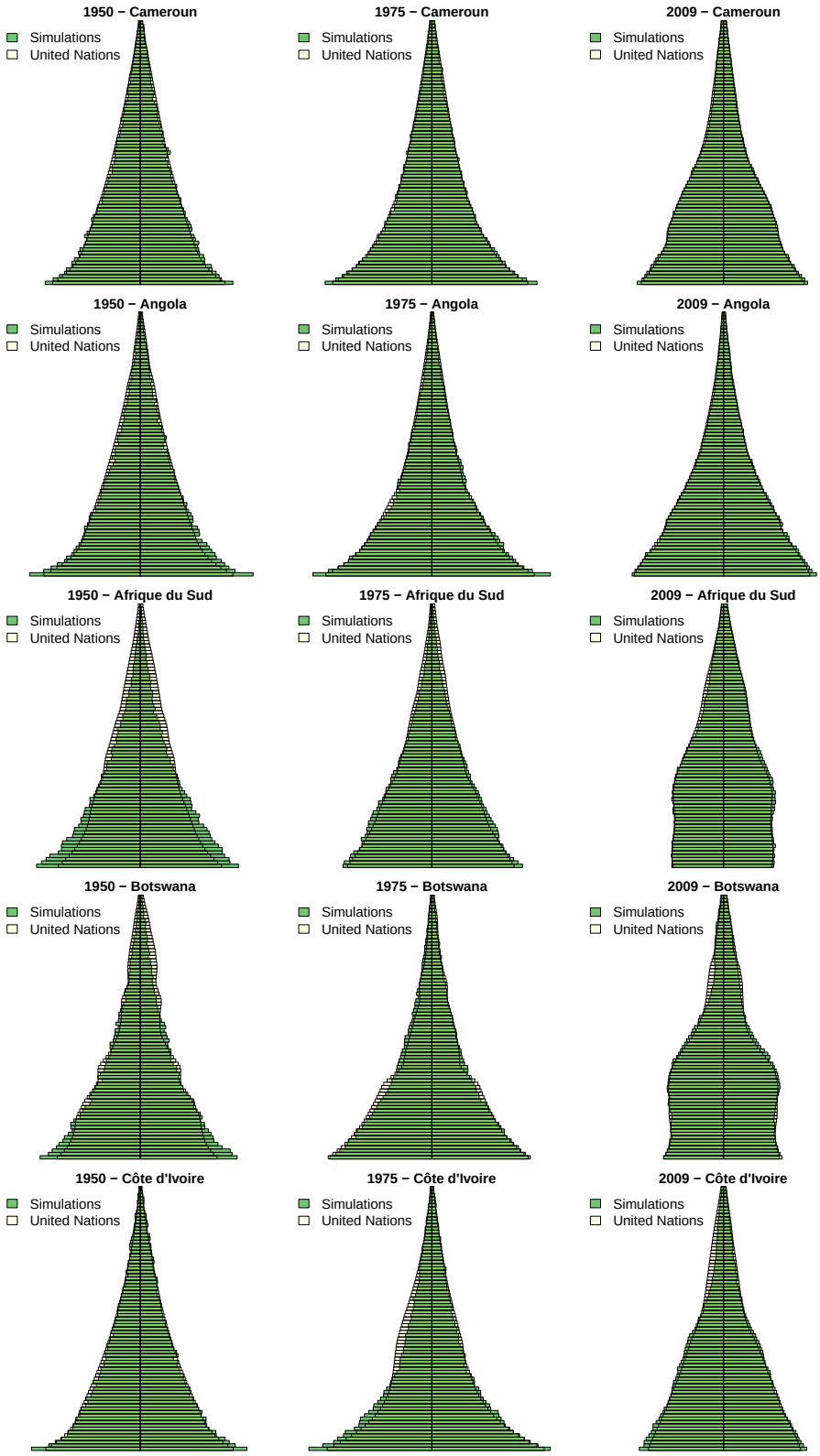

FIGURE 2.3 – *Structures par âge recalculées à partir des microsimulations et structures estimées par les Nations Unies (0-80 ans)*

Encadré 4 Construction des tables "sans sida" par les Nations Unies

La variante “sans sida” des Nations Unies est construite pour les pays ayant atteint une prévalence du VIH au moins égale à 1% durant la période 1980-2010[44]. Cette mortalité "sans sida" est celle dont on aurait déduit l’impact *direct* de l’épidémie. Elle tient toutefois compte du fait que la mortalité associée aux autres causes a été affectée par l’expansion du VIH-sida, en raison des effets sanitaires et socio-économiques *indirects*. En ce sens, il ne s’agit pas d’une situation hypothétique dans laquelle le sida n’aurait jamais existé. Cette variante est à la base des estimations et des projections de population dans les pays affectés par le VIH : elle constitue la première étape de la reconstitution des tendances de la mortalité, et dans un second temps, les décès dûs au sida y sont ajoutés.

En l’absence de données fiables relatives aux causes de décès en Afrique subsaharienne, il est difficile d’isoler dans les estimations les décès dûs au sida et ceux dûs aux autres causes. L’estimation de la mortalité sans sida procède donc par itération :

> "Beginning with an assumed trend of the background mortality, the HIV/AIDS epidemic is modelled and the results are then compared with overall mortality estimates, if they are available (. . .) or with results provided by a census or survey. If necessary, background mortality is adjusted, and the procedure is repeated until a reasonable agreement between the model output and available evidence is achieved." (UN (2005), pp. 109-110)

Cet ajustement se fait essentiellement en révisant la mortalité des enfants, qui est l’indice qui détermine le niveau de la mortalité “sans sida”. Rappelons que pour la majorité des pays africains, les Nations Unies utilisent les tables de Princeton, combinées avec des niveaux de mortalité des enfants, pour obtenir des valeurs provisoires de mortalité adulte générale. Cette procédure concerne encore quatre pays africains sur cinq (Reniers, Masquelier, et Gerland forthcoming).

La modélisation de la dynamique de l’épidémie repose sur le modèle appelé EPP, l’*UNAIDS Estimation and Projection Package* (Brown et al. 2008, Ghys et al. 2008). Il s’agit d’un modèle - et d’un logiciel - qui ajuste, à l’échelle nationale, les tendances de la prévalence adulte enregistrée dans les sites sentinelles ainsi que dans les enquêtes de séroprévalence (essentiellement les EDS). Il est développé pour les pays où l’épidémie du VIH est généralisée. Quatre paramètres clés sont utilisés pour l’ajustement du modèle :

1. le taux de croissance de l’épidémie (r),
2. l’année de départ de cette épidémie (t_0),
3. la fraction de la population qui, lors de l’année de départ, est entrée à 15 ans dans la population exposée au risque de contracter le VIH (f_0),
4. un paramètre de réponse (ϕ) qui renvoie au recrutement de nouveaux entrants dans la population à risque.

44. Ainsi que quelques pays dont la prévalence est plus faible mais dont la population est telle que le nombre de séropositifs en 2010 atteint 500 000 personnes : le Brésil, la Chine, l’Inde, la Russie et les Etats-Unis.

Ce modèle EPP peut être ajusté à partir du logiciel du même nom distribué par l'ONUSIDA[45]. Ces outils sont développés et révisés par le *Groupe de Référence d'ONUSIDA sur les Estimations, la Modélisation et les Projections*[46], dont les recommandations guident aussi les révisions du logiciel Spectrum[47]. Les prévalences et incidences générées par EPP sont transférées dans Spectrum pour estimer différents indicateurs, tels que le nombre de séropositifs, le nombre d'orphelins dus au sida, le nombre de nouvelles infections et les besoins en matière de traitement (Stover et al. 2008). Notons que depuis 2009, le programme EPP tient compte également de la distribution des anti-rétroviraux[48].

Du côté de la Division de la Population des Nations Unies, l'enjeu est d'estimer la mortalité avec et sans sida au départ des prévalences fournies par l'UNAIDS. La Division utilise abcDIM, qui est une combinaison de Spectrum et EPP (Mulder et Johnson 2005). Les résultats sont établis de façon à être cohérents avec les données de l'UNAIDS, et l'UNAIDS utilise en retour les projections sans sida de la Division pour calculer l'*incidence* à partir des *prévalences*.

Dans l'annexe méthodologique de la révision de 2004, les analystes de la Division de la Population détaillent leur démarche de la façon suivante (UN 2005) :

1. La reconstitution de la dynamique de l'épidémie par le modèle EPP permet de calculer le nombre de personnes nouvellement infectées chaque année, une population qui est ensuite ventilée par âge et par sexe. La distribution par âge des risques d'infection est supposée être constante et suivre une distribution de Weibull[49].
2. Une approche multi-états est ensuite utilisée pour projeter dans le futur le nombre de personnes qui vont vivre avec le VIH et le nombre de personnes qui décèderont du sida. Là encore, la durée de la période d'incubation est supposée constante et elle est modélisée au moyen d'une distribution de Weibull. Elle n'est pas modulée selon l'âge à l'infection.
3. A partir de là, le nombre d'enfants de moins de cinq ans qui vont décéder du sida est estimé, au départ du nombre de mères séropositives, en tenant compte d'une réduction de fécondité fixée à 20% chez ces dernières, ainsi que d'une probabilité de transmission mère-enfants, fixe elle aussi, et estimée à 35%.
4. Moyennant quelques autres hypothèses (notamment sur l'impact de la distribution des anti-rétroviraux, ainsi que sur l'impact des programmes de prévention de la transmission mères-enfants), le nombre de décès imputables au VIH-sida peut alors être calculé par âge et sexe.

45. Téléchargeable sur www.unaids.org/fr/KnowledgeCentre/HIVData/Epidemiology
46. UNAIDS Reference Group on Estimates, Modelling and Projections : www.epidem.org
47. Téléchargeable sur www.FuturesInstitute.org.
48. Voir Mulder et Johnson (2005) ou Dorrington et al. (2006b) pour certaines critiques de Spectrum. Voir Stover et al. (2008) pour les modifications récentes.
49. Avec un paramètre de forme noté k et un paramètre d'échelle noté λ, la densité de probabilité de la distribution de Weibull, correspond à

$$f(x) = (k/\lambda)(x/\lambda)^{(k-1)}e^{-(x/\lambda)^k}$$

La probabilité de survie x années après l'infection se note

$$s(x) = e^{(-x/\lambda)^k}$$

Nous utiliserons cette distribution par la suite.

5. Ce nombre de décès est ensuite ajouté à la table de mortalité "sans sida", qui peut être revue de façon itérative. Cette démarche procure la variante "moyenne".

Pour les microsimulations, les taux de mortalité sont repris des tables de la variante "sans sida" des Nations Unies. A partir de 1975, des risques d'infection et de décès spécifiques au VIH-sida sont ensuite ajoutés. Les taux de mortalité obtenus de cette façon (en combinant des tables sans sida et une mortalité due au sida) devraient cadrer avec la variante "moyenne" des Nations Unies (*cf.* infra).

Nous disposons de ces tables de mortalité "sans sida" par intervalles quinquennaux (1950-1955, 1955-1960...) et par groupes d'âges (0, 1, 5-10, ...). Les tendances sont toutefois modélisées sur base annuelle, et à partir de tables par année d'âge, afin de pouvoir mieux intégrer dans les simulations la dynamique de l'épidémie du VIH. Les tables de mortalité abrégées sont donc étendues en tables par année d'âge [50]. Les survivants aux âges exacts (${}_1l_x$) disponibles tous les cinq ans sont ensuite annualisés au moyen de splines naturelles d'interpolation (Annexe A.2.4, p. 408), et les tables de mortalité sont reconstruites sur cette base [51]. Une distribution des décès infantile est estimée à partir des enquêtes EDS, et présentée dans l'annexe A.2.5 (p.409).

Taux "sans sida" recalculés à partir des populations simulées

Pour calculer les taux de mortalité à partir des populations simulées, nous optons pour une approche biographique. Les fichiers sont restructurés sous forme de personnes-années, ce qui permet de tenir compte de la nature longitudinale des données. Les taux de mortalité "sans sida" s'obtiennent ici en rapportant les décès à la durée d'exposition observés dans chaque classe d'âge et pour chaque année, après avoir tronqué les données à la date d'infection au VIH pour les séropositifs.

50. Pour transformer les tables quinquennales en tables annuelles, nous combinons deux approches : d'une part, des splines d'interpolations monotones utilisées pour interpoler la courbe des survivants (Smith et al. 2004) et d'autre part, les multiplicateurs de Sprague, pour répartir les décès quinquennaux entre les âges (Siegel et al. 2004). Ces deux méthodes permettent de graduer le nombre de décès fournis par classe d'âge par saut de 1 an, en obtenant des valeurs de ${}_1l_x$ qui recoupent les valeurs de ${}_5l_x$ des Nations Unies. Chacune d'entre elles produit de légères irrégularités aux jeunes âges (Kostaki et Panousis 2001). Combiner ces deux méthodes permet d'obtenir un résultat plus lissé. Précisons que pour interpoler les survivants entre 1 et 5 ans, nous nous sommes servi du modèle général des Nations Unies.

51. Les valeurs de ${}_na_x$, le temps passé dans l'intervalle par les personnes qui décèdent au sein de l'intervalle, sont nécessaires pour reconstruire les tables annuelles mais elles ne peuvent pas être déduites de la courbe des survivants aux âges exacts. Pour les estimer, nous utilisons l'information sur la distribution par âge des décès, comme le propose Keyfitz (1966). Les valeurs obtenues par année d'âge sont ensuite re-pondérées pour cadrer avec les valeurs quinquennales fournies dans les tables des Nations Unies. Comme le programme requiert que les taux soient fournis jusqu'à 100 ans, la méthode de Coale et Kisker (1990) est utilisée pour étendre les tables de survie jusqu'à 100 ans.

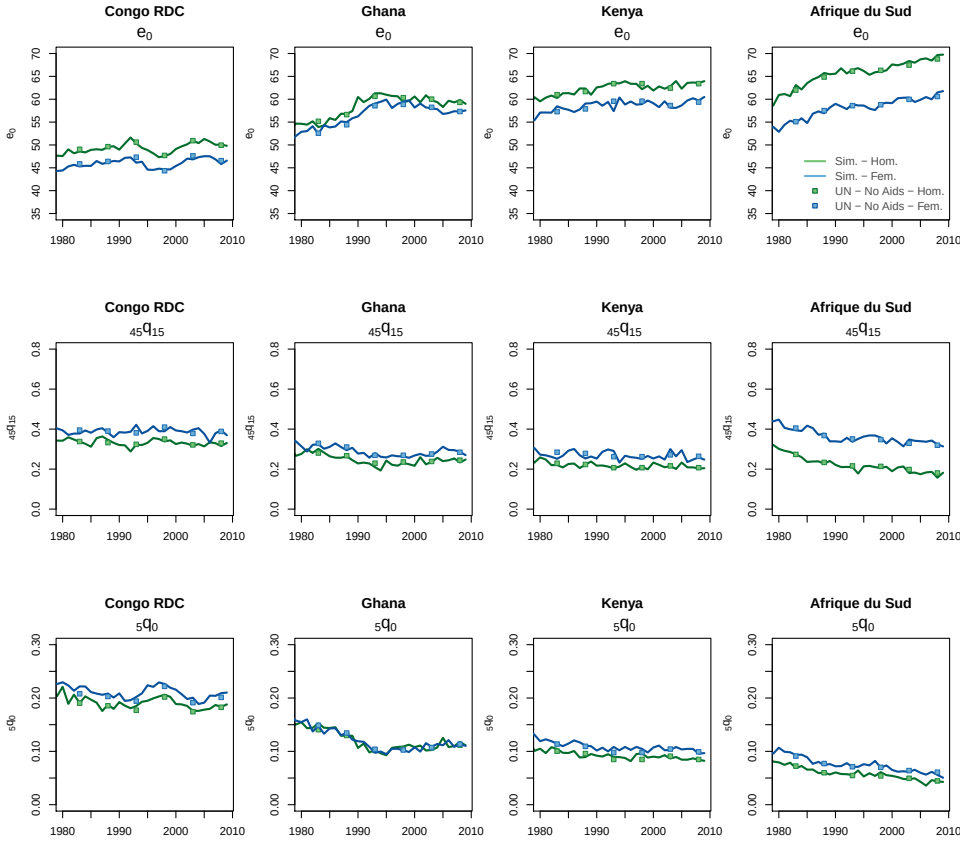

FIGURE 2.4 – *Niveaux de e_0, $_{45}q_{15}$ et $_5q_0$ introduits et recalculés à partir des simulations - mortalité sans sida*

A titre illustratif, le graphe 2.4 présente, pour quelques pays, les valeurs des indices e_0, $_{45}q_{15}$ et $_5q_0$ issus des tables de mortalité "sans sida" des Nations Unies et les compare aux valeurs recalculées à partir des populations simulées. Comme il s'agit de valeurs annuelles, les résultats des microsimulations sont caractérisés par une certaine variabilité aléatoire, mais ils cadrent bien avec les niveaux "sans sida" introduits. C'est sur cette mortalité sans sida que va venir se greffer les décès dus au VIH-sida (*cf.* infra).

2.3.3 Niveaux et structures de la fécondité

Schéma de fécondité par âge

Les taux de fécondité par âge qui sont retenus ici sont ceux issus des estimations des Nations Unies (Révision 2008) [52]. Ces taux sont disponibles par *groupes d'âge* et par *intervalles quinquennaux* (1950-1955, 1955-1960...). Comme les taux de mortalité, nous les décomposons en taux par année d'âge et par années calendrier.

52. Les taux de fécondité de la variante *sans sida* sont en fait les mêmes que ceux de la variante *moyenne*. Ces taux ne sont pas publiés avant 1980 pour les pays affectés par le VIH-sida, et avant 1995 pour les autres (UN 2009d). Nous utilisons donc des données partiellement non publiées. Source : United Nations, Department of Economic and Social Affairs, Population Division : World Population Prospects DEMOBASE extract. 2009.

Des taux par année d'âge sont d'abord calculés en lissant les naissances cumulées par âge [53]. Ces taux lissés, disponibles par intervalles quinquennaux, sont ensuite "annualisés" au moyen de splines naturelles d'interpolation. Cette procédure en deux étapes garantit que les ISF recalculés correspondent à ceux estimés par les Nations Unies. Le graphe 2.5 présente les résultats de ce lissage en deux étapes pour le cas du Zimbabwe (1980-2010).

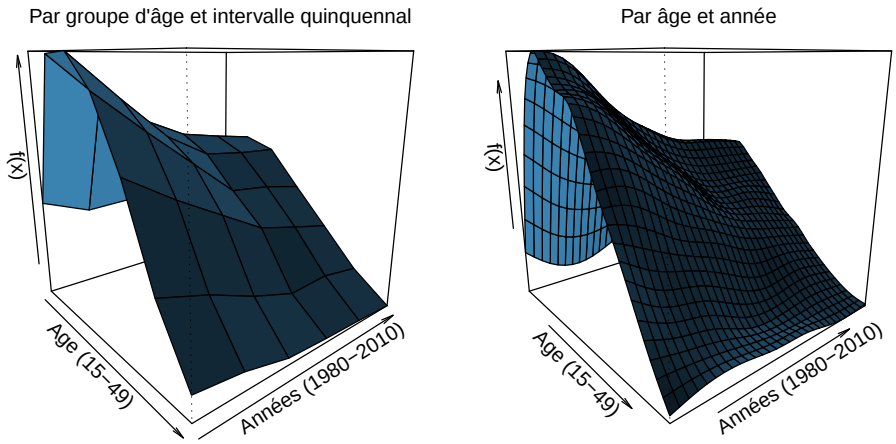

FIGURE 2.5 – *Schémas de fécondité estimés par les Nations Unies (par intervalles quinquennaux d'âge et d'années) et introduits dans les simulations (par intervalles annuels d'âge et d'années) - Zimbabwe 1980-2010*

Tendances de la fécondité dans le temps

Pour la période 1850-1950, au cours de laquelle l'hypothèse est celle d'une population stable, les taux de fécondité sont repris des estimations de 1950-1955. A partir de 1950, ils varient par année.

Les rapports de masculinité à la naissance varient également dans le temps et par pays. Ils sont repris des estimations des Nations Unies. Ils s'établissent pour la plupart des pays à 103, mais peuvent varier pour les pays considérés entre 101.6 (Madagascar 1950-2000) et 106.2 (Ghana 1950-2010).

Distribution des naissances par parité

Nous avons insisté, dans la section 2.2, sur l'importance de faire varier les taux de fécondité *selon la parité atteinte* dans les microsimulations (voir également la discussion en annexe, p. 394). En particulier, les taux introduits uniquement par âge sous-estiment la dispersion du nombre d'enfants nés vivants, et donc également la dispersion de la taille des fratries. Les réseaux de parenté construits uniquement sur base de taux par âge

53. Ce lissage est fait au moyen d'une fonction spline monotone (Smith et al. 2004). Les structures par âge féminines par années, par pays et par groupes d'âge sont reprises des Nations Unies pour calculer le nombre de naissances par groupe d'âge (UN 2009d). Les taux de fécondité sont ensuite recalculés à partir des naissances décumulées et des populations féminines par année d'âge.

sont donc peu plausibles. Malheureusement, les taux de fécondité des Nations Unies concentrent les naissances de tous rangs, et aucune base de données ne procure des taux par âge et par rang en couvrant tous les pays africains [54].

Pour pouvoir modéliser la fécondité par âge et parité atteinte, nous avons donc cherché à dégager des enquêtes EDS une répartition "typique" des naissances par rang. Toutes les enquêtes EDS accessibles et comprenant des histoires génésiques (n=93) ont été restructurées en personnes-années et agrégées en un seul fichier. Les trajectoires biographiques des femmes sont scindées par âge et par parité. Au total, les enquêtes couvrent une période allant de 1986 à 2009, et elles rassemblent plus de 4 millions de personnes-années.

A titre illustratif, les taux de fécondité par *groupe d'âge* pour chaque rang de naissance sont présentés dans le graphe 2.6. Ces taux sont obtenus en combinant toutes les enquêtes et calculés pour les cinq ans précédant chacune des enquêtes : la fécondité de premier rang est la plus faible quasiment partout (en raison du fait que certaines femmes restent sans enfants), puis, à mesure que la parité s'élève, le schéma par âge de la fécondité se déplace vers la droite (on retrouve ici la distribution déjà observée avec l'enquête congolaise de 2007 en annexe A.2.1, p.394). Son mode évolue de 20-24 ans pour la fécondité de rang 2 à 30-34 ans pour la fécondité de rang 9 et plus [55].

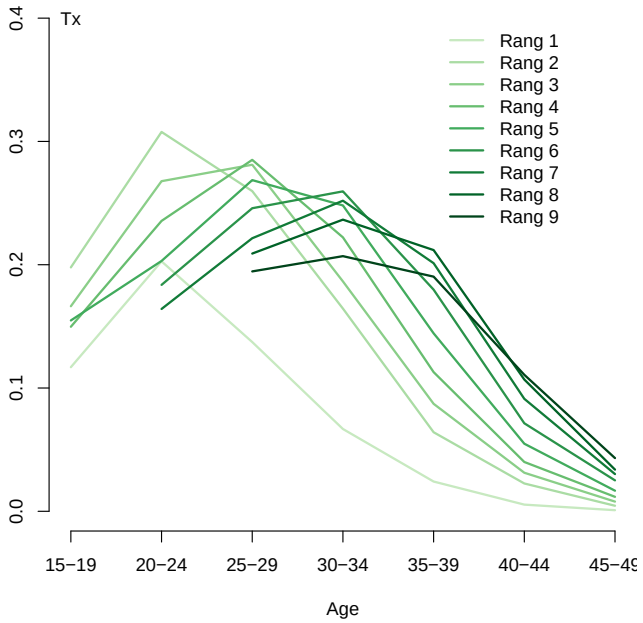

FIGURE 2.6 – *Taux de fécondité par âge pour chaque rang de naissance, estimés à partir des enquêtes EDS (n = 93)*

54. Pour la période récente (après 1990), l'*Annuaire Statistique des Nations Unies* ne fournit que des taux de fécondité par rang pour l'Île Maurice (et l'Afrique du Nord) (UN 2002a).

55. Sur ces graphes, les taux ne sont pas présentés lorsque le nombre total de personnes-années est inférieur à 1000. Les naissances de rang supérieure à 8 sont regroupées dans une même catégorie.

Ce graphe 2.6 présente des taux par âge pour chaque niveau de parité atteinte. Pour distribuer les taux de fécondité "tous rangs" en taux par rang, nous calculons plutôt *des taux par rang pour chaque âge*. Nous cherchons en effet à savoir comment, à un âge donné, la fécondité de rang i se distingue de la fécondité de rang 2, qui sert ici de référence. Pour chaque année d'âge, le nombre de naissance est donc régressé sur la parité atteinte au moyen d'une régression de Poisson, en contrôlant les durées d'exposition correspondantes (Schoumaker 2004). Les taux par rang pour chaque année d'âge sont présentés dans le graphe 2.7. Lorsque les taux diffèrent significativement des taux de fécondité de rang 2 (au seuil 0.05), nous retenons l'exponentielle des coefficients de la régression comme rapport de taux [56]. A partir là, nous pouvons estimer des taux par âge et par rang. La procédure de calcul est détaillée dans l'annexe A.2.6, page 410.

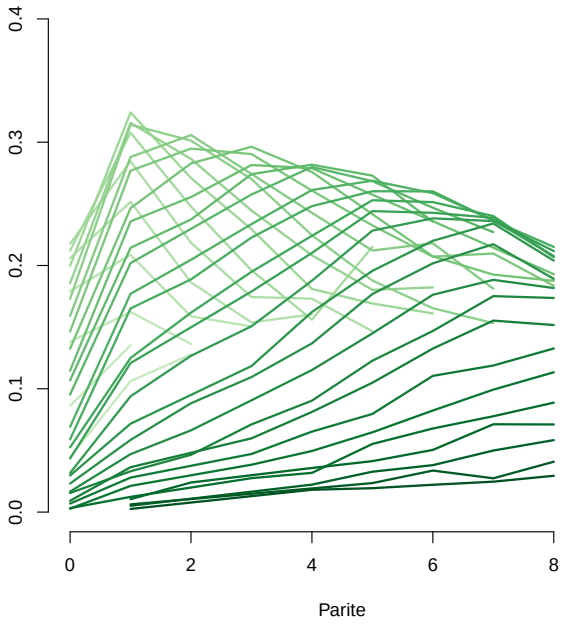

FIGURE 2.7 – *Taux de fécondité par rang de naissance pour chaque année d'âge, estimés à partir des enquêtes EDS (n = 93)*
Note : en allant du plus clair au plus foncé, on parcourt les âges de 15 à 49 ans.

Cette prise en compte de la parité pour modéliser la fécondité pourrait être améliorée à plusieurs égards. Notamment, un modèle multi-niveau permettrait de mieux tenir compte de spécificités régionales dans la distribution des naissances par rang. De même, dans les pays où la fécondité baisse sensiblement, il est peu probable que cette baisse se fasse au même rythme quelque soit la parité atteinte. Généralement, on observera une baisse plus rapide chez les femmes de parités élevées [57]. Il faudrait donc faire varier les rapports de taux dans le temps, et selon les pays [58].

56. Dans le cas contraire, nous retenons simplement 1. Le fait de travailler par année d'âge peut poser problème dans les cas atypiques (p.ex. parités élevées aux jeunes âges) mais ceci est sans effets sur le résultat des microsimulations car les taux s'appliquent dans ces cas sur une très petite minorité de femmes.

57. On pourra observer ceci dans le graphe 3.18 p.140.

58. Par ailleurs, la fécondité devrait être modélisée par statut matrimonial. Le fait que la

Niveaux de fécondité recalculés à partir des populations simulées

Le graphe 2.8 présente les ISF estimés par les Nations Unies et ceux recalculés depuis les microsimulations pour différents pays. On observe une bonne concordance entre les deux séries de valeurs. Cela signifie que *les taux de fécondité que nous avons introduits selon la parité atteinte permettent bien de récupérer ici les taux 'tous rangs'*. C'est ce que nous cherchions, afin de disposer de simulations dont les paramètres de fécondité cadrent avec les tendances passées estimées pour les différents pays africains.

Problèmes posés par la fécondité masculine

Selon nous, la principale limite du programme SOCSIM est que *les pères des enfants nés hors mariage ne sont pas connus*. Les naissances surviennent uniquement selon un rythme défini par les taux de fécondité féminine, puisque les actes sexuels en soi ne sont pas modélisés par SOCSIM.

En conséquence, les taux de fécondité masculine vont résulter des interactions entre d'une part, le calendrier de la *fécondité féminine*, et d'autre part, celui de la *nuptialité*. Mais rien ne garantit que l'on va récupérer des schémas de fécondité masculine proches de ceux qui pourraient être observés dans les enquêtes ou les recensements (*cf*. section 3.2.4, p. 129). Dès lors, par exemple, l'*âge des pères à la naissance des enfants* - une information clé pour estimer la mortalité adulte à partir de la survie des parents - tel qu'on va pouvoir le calculer à partir des simulations, ne cadrera pas avec les valeurs que l'on pourra observer en réalité.

Cette limite nous a posé problème, car nous souhaitions explorer les techniques permettant d'estimer la mortalité masculine à partir des pères. Notamment, l'un des objectifs était d'analyser comment évolue la probabilité d'être orphelin de père dans les contextes où le VIH-sida joue un rôle important. Nous verrons dans le chapitre suivant que dans la littérature, l'accent a surtout été mis sur la *survie des mères*. Les pères ont fait l'objet de moins d'attention. Dans cette littérature, on ne trouvera par exemple pas de méthode permettant d'estimer la survie des pères dans les contextes perturbés par le VIH-sida.

C'est une des raisons pour lesquelles *nous avons intégré dans les simulations des taux de primo-nuptialité*. Comme les enfants nés dans le mariage ont un père identifié, ceci permet d'établir un lien avec le père, au moins pour ces enfants. Puisqu'il subsiste un certain nombre de naissances hors mariage, nous avons ensuite alloué des pères à ces enfants nés hors mariage. Une seconde motivation derrière la prise en compte de la nuptialité dans les simulations était de pouvoir faire varier la fécondité selon le statut matrimonial.

Cette intégration de la nuptialité n'est toutefois pas présentée ici, pour deux raisons :

fécondité ne varie pas avec le statut matrimonial conduit ici à des proportions trop élevées de naissances hors mariage. Ceci pourrait faire l'objet d'améliorations par la suite, mais cet ajustement n'est pas simple : il faut tenir compte de la répartition des femmes par âge, par parité et par statut matrimonial, afin de garantir que ni les schémas par âge ni le niveau moyen de la fécondité ne soient altérés.

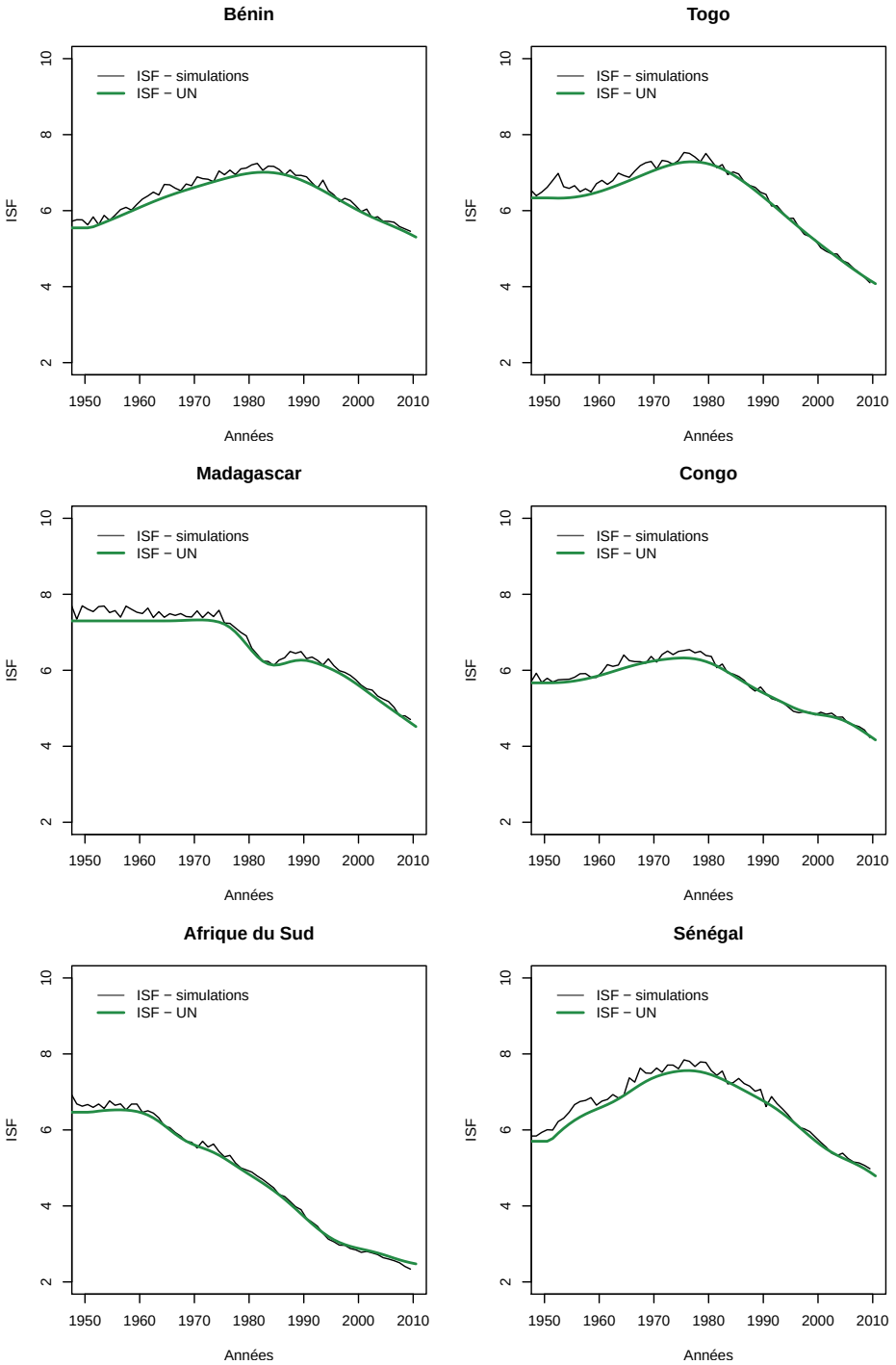

FIGURE 2.8 – *Valeurs des ISF estimés par les Nations Unies et recalculés à partir des populations simulées - exemples du Bénin, du Togo, de Madagascar, du Congo, de l'Afrique du Sud et du Sénégal.*

1. Comme le mariage est un événement qui concerne deux individus, il peut y avoir *un décalage entre les taux de nuptialité introduits par sexe dans les simulations et les taux recalculés à partir des populations finales*. Tout dépend de la vitesse avec laquelle les individus qui entrent sur le marché matrimonial trouvent un partenaire. Or, en raison d'importants écarts d'âge au mariage observés en Afrique, il est souvent nécessaire de prévoir des taux de *remariage* ou de la *polygamie* pour que les individus destinés au mariage puissent tous trouver un partenaire. Faute de données, nous n'avons simulé que les *premiers* mariages, ce qui a posé problème.
2. Par ailleurs, nous nous sommes aperçus que les *proportions d'orphelins de père* (que la simulation des mariages devait permettre d'identifier) était sous-estimées par notre modèle de microsimulation, pour des raisons détaillées plus loin.

A ce stade, l'intégration de la nuptialité dans les microsimulations est donc indépendante des développements qui vont suivre, raison pour laquelle elle n'est pas présentée ici. Nous lui avons toutefois accordé une attention considérable et nous renvoyons le lecteur intéressé par cette question à l'annexe A.2.7 (page 412). Nous y présentons une large série d'estimations relatives à l'âge moyen au premier mariage et à la distribution des célibataires par âge. Nous recourons ensuite au modèle de Coale et McNeil (1972) pour chacun des pays africains simulés. Nous utilisons également un modèle multi-niveaux pour obtenir des taux de primo-nuptialité qui varient dans le temps et par pays.

***

Après avoir spécifié les taux de mortalité sans sida, ainsi que les taux de fécondité et de nuptialité (en annexe), la section suivante détaille comment les simulations ont été conçues pour intégrer les différents impacts du VIH-sida sur la dynamique démographique.

2.4 Intégration du VIH-sida

Il existe aujourd'hui d'innombrables modèles *macro* visant à reproduire la dynamique de la transmission du VIH-sida et ses impacts démographiques (Stover et al. 2008, Dorrington et al. 2006a, Heuveline 2003, Gregson et al. 1994, Anderson et al. 1992). A ces modèles macro s'ajoutent quelques approches basées sur la microsimulation. Les modèles qui font référence en la matière sont sans conteste STDSIM (Van der Ploeg et al. 1998, Orroth et al. 2007) et SIMULaids (Auvert et al. 2000)[59]. Sam Clark (2006) a également développé un modèle très abouti au niveau de la prise en compte des comportements sexuels et de leur impact sur la propagation du VIH.

Ces différents modèles micro et macro mettent en lumière différents aspects de l'épidémie : ils permettent d'estimer les besoins en traitements antirétroviraux, de ré-estimer les tendances observées de prévalence et

59. Voir également Leclerc et al. (2009), Clark et Eaton (2008).

d'incidence, d'évaluer l'impact de différentes configurations de réseaux sexuels, etc. De nombreux modèles, tant micro que macro, intègrent des éléments comportementaux, comme le recours à des pratiques sexuelles à risques, la fréquence des rapports sexuels, les règles guidant la sélection des partenaires sexuels, etc.

Notre objectif est ici moins ambitieux puisqu'il s'agit uniquement d'intégrer des risques de transmission et de décès spécifiques au VIH dans des microsimulations menées avec SOCSIM. La tâche n'est toutefois pas facile, car SOCSIM n'a que très rarement été utilisé pour mener des microsimulations dans des contextes perturbés par le VIH-sida (Wachter et al. 2002; 2003, Zagheni 2010).

Comme nous l'avons annoncé, intégrer des risques spécifiques au VIH-sida dans SOCSIM doit se faire en recourant à différents sous-groupes de populations. Cette option permet de ne pas modifier le coeur du programme. Elle a toutefois ses limites car elle implique que la fécondité et l'infection soient découplés. En effet, rappelons que les actes sexuels en soi ne sont pas modélisés par SOCSIM [60].

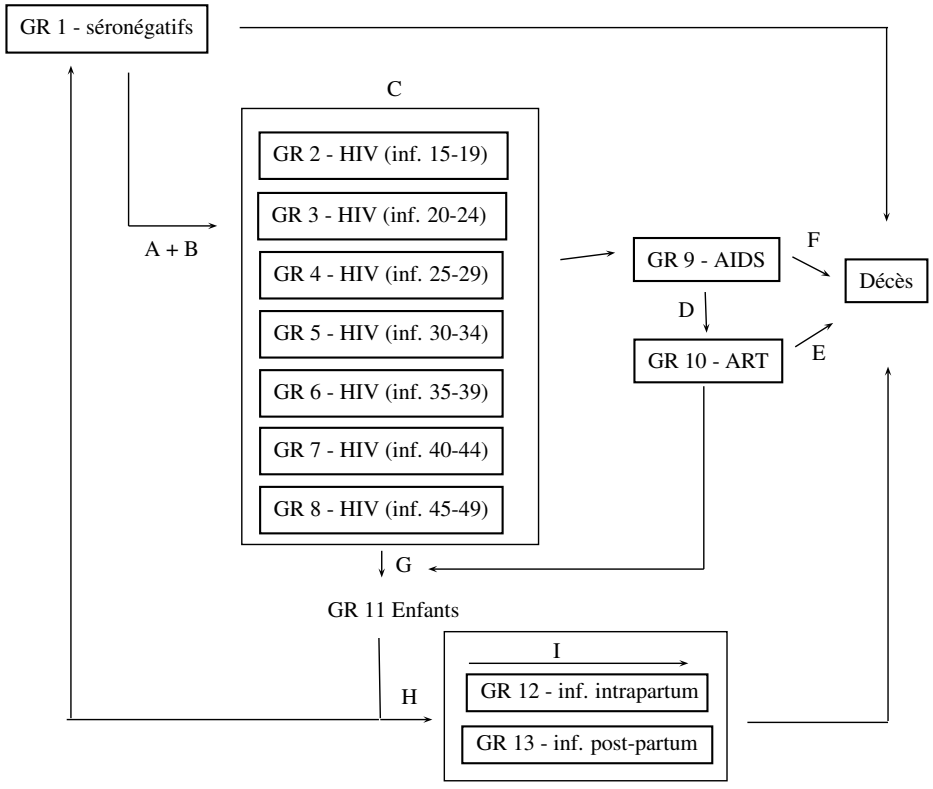

FIGURE 2.9 – *Sous-groupes de population définis pour intégrer le VIH-sida dans les simulations*

Le schéma 2.9 présente les différents sous-groupes que nous avons

60. Ceci pourrait être modifié, en "détournant" les procédures utilisées par SOCSIM pour gérer les mariages. Si l'on définit ce que SOCSIM entend par "mariage" comme étant plutôt des "partenariats sexuels", il est possible de former de tels partenariats en tenant compte des écarts d'âge des partenaires sexuels, de la formalisation de certains partenariats, etc.

spécifiés. Les sections suivantes détaillent ces sous-groupes, ainsi que les différentes transitions que les individus opèrent entre eux. En quelques mots, la logique d'ensemble est la suivante :

- *L'infection au VIH se traduit par le passage du groupe des séronégatifs (1) vers les séropositifs (2-8), selon l'âge auquel les individus contractent le VIH.*
- Le VIH se caractérise par une très longue période de latence clinique (environ 9-10 ans), durant laquelle les personnes infectées sont asymptotiques. *Dans les simulations, les individus sont maintenus au sein des groupes de 2 à 8 durant cette période asymptotique.*
- Au terme de cette période, les organes chargés de la régénération des lymphocytes CD4 sont peu à peu épuisés, et le système immunitaire s'effondre, car il ne parvient plus à adapter sa réponse face à un virus qui mute et se recombine très vite. Ce déficit immunitaire caractérise le stade du sida, la phase ultime de l'infection, qui dure nettement moins longtemps (environ 2 ans). *Dans les simulations, cette étape correspond au groupe 9* (schéma 2.9). L'immunodépression rend alors les séropositifs vulnérables aux agressions, notamment celles menées par les maladies opportunistes (tuberculoses, candidoses, pneumonies, etc.).
- Ce stade du sida est diagnostiqué sur base clinique[61] ou immunologique (à partir du dénombrement des cellules CD4+ dans l'organisme). Il correspond à la période durant laquelle des traitements antirétroviraux (ART) peuvent être administrés. *Dans les simulations, les individus sous traitement transitent par le groupe 10, avant de décéder.*
- La transmission du VIH peut également se faire verticalement, des mères aux enfants. *Dans les simulations, tous les enfants nés de mères infectés transitent par le groupe 11. Si les enfants contractent le VIH intra-partum, ils sont déplacés dans le groupe 12. S'ils contractent le VIH post-partum, ils sont déplacés vers le groupe 13. S'ils ne sont pas infectés, ils retournent dans la population de séro-négatifs.*

Pour chacun de ces groupes, des taux de transitions vers d'autres groupes (ou vers le décès) sont calculés. Afin de spécifier ces taux de transition, une série d'éléments doivent intervenir. Ils sont listés ci-dessous, et font référence aux flèches indiquées dans le schéma 2.9 (p.76) :

A - les tendances nationales d'incidence du VIH,

B - une série de probabilités individuelles d'infection au VIH par âge et sexe,

C - des modèles de progression des personnes infectées vers le stade du sida,

D - des estimations relatives à la distribution des traitements antirétroviraux,

E - un modèle de survie sous traitements ART,

F - un modèle de survie sans traitement pour les individus ayant atteint le stade du sida clinique,

61. C'est le stade 4 de la classification OMS.

G - des estimations relatives à la baisse de la fécondité des femmes séropositives,

H - des estimations relatives à l'intensité de la transmission verticale intra- et post-partum,

I - des modèles de survie des enfants infectés verticalement.

Deux transitions sont variables dans le temps et par pays : l'incidence du VIH et l'allocation des traitements ART. Toutes les autres transitions sont communes à tous les pays et fixes dans le temps. Pour intégrer ces différentes transitions dans les simulations, nous suivrons autant que possible les recommandations du *Groupe de Référence d'ONUSIDA sur les Estimations, la Modélisation et les Projections*.

La première transition, qui correspond à l'infection au VIH, requiert que soient définis des taux d'infection par âge et par sexe. C'est de loin l'élément le plus difficile à estimer et à intégrer dans les simulations. Pour faire passer une certaine proportion de séronégatifs vers la séropositivité, nous utilisons les *tendances de l'incidence du VIH*, combinées à une distribution standard des *risques d'infection par âge et par sexe*.

Le premier élément, l'incidence du VIH, est estimée par l'ONUSIDA, essentiellement à partir :

1. des tendances de la prévalence dégagés du modèle utilisé par l'ONUSIDA, appelé EPP, pour *Estimation and Projection Package* (nous en avons parlé dans l'encadré 4, p.66),
2. des niveaux de mortalité estimés par les Nations Unies,
3. de la survie des personnes infectées (sous traitement ou non).

Avant de discuter de la survie des personnes infectées, nous présentons les dernières estimations de la prévalence du VIH.

2.4.1 Prévalence du VIH

Selon les dernières estimations fournies par l'ONUSIDA, les prévalences nationales s'élèvent en moyenne sur l'Afrique subsaharienne à 5.2% [4.9-5.4] en 2009 chez les adultes de 15 à 49 ans, contre 5.8% [5.5-6.0] en 2001 (UNAIDS 2009). Ces dernières années, la prévalence a baissé de façon importante dans quelques pays (Ouganda, Zimbabwe, Côte d'Ivoire, Burundi, etc.). Elle s'est stabilisée dans d'autres pays, parfois à des niveaux très élevés (en Zambie et en Namibie autour de 15%, au Swaziland, au Lesotho et au Botswana à plus de 20%, par exemple). Globalement, la quasi-totalité des pays sub-sahariens ont atteint leur pic de prévalence.

Les cartes 2.10 détaillent les prévalences estimées pour les années 1990, 1995, 2000 et 2005 (tels qu'estimées en 2008) [62].

62. Les dernières estimations sont disponibles via http ://data.unaids.org/. Notons que jusqu'à la révision 2007, l'UNAIDS ne publiait pas des tendances mais uniquement la dernière estimation annuelle.

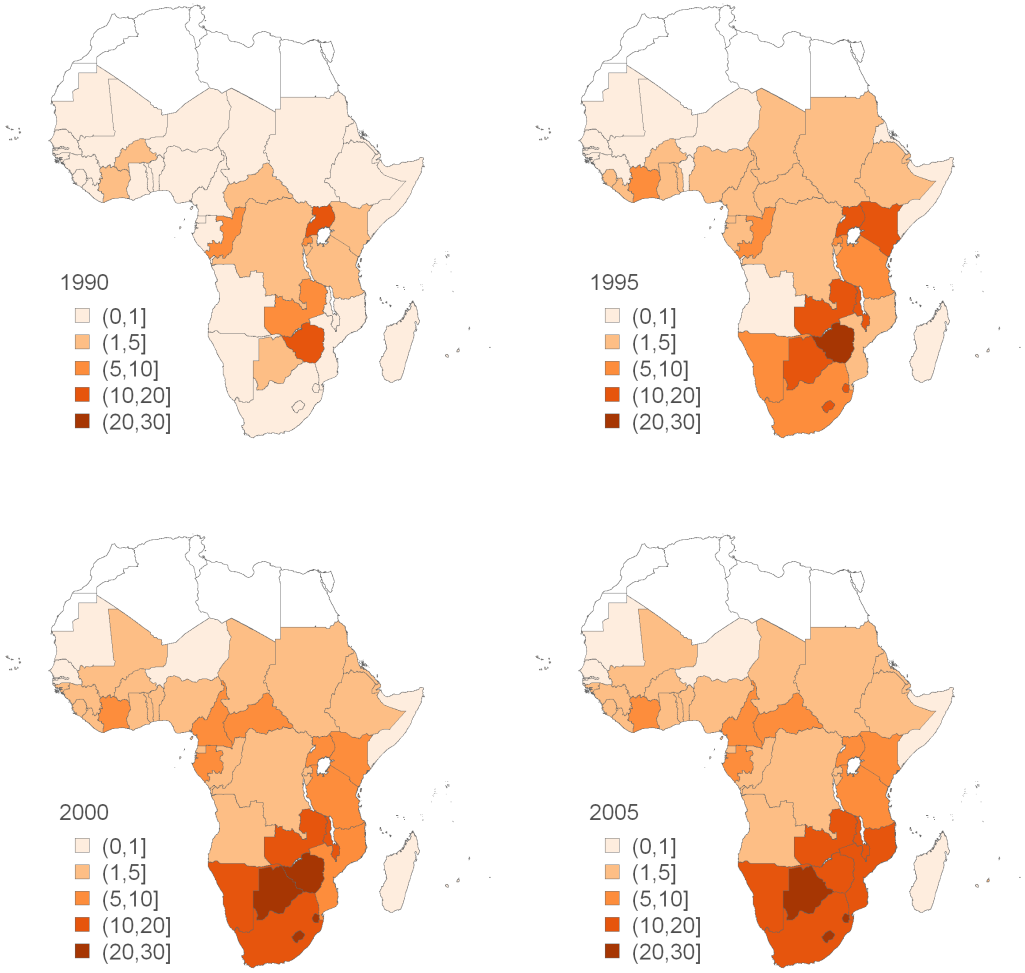

FIGURE 2.10 – *Prévalences estimées (en %) en 2008 pour les années 1990, 1995, 2000 et 2005 - Source : (UNAIDS 2009)*

Les prévalences reprises sur les cartes 2.10 sont celles de l'ONUSIDA. Nous avons rappelé que ces estimations sont issues du modèle EPP utilisé pour reconstruire une tendance générale de la prévalence à partir des différentes *estimations ponctuelles*. Durant les années 1980 et 1990, ces estimations ponctuelles étaient fournies par les systèmes de surveillance sentinelles mis en place dans les *services de consultations prénatales* (SCPN). Ces estimations sont connues pour être biaisées à de nombreux égards, surtout à la hausse, notamment parce que les femmes qui consultent ont récemment été actives sexuellement sans contraception. D'autres éléments interviennent, comme le faible recours aux consultations prénatales, ou l'inégale répartition des SCPN sur le territoire (avec une distribution disproportionnée des SCPN en faveur des milieux urbains) [63]. Un autre biais structurel est introduit par la diminution de la fécondité observée chez les femmes séropositives (Lewis et al. 2004) (*cf.* infra, p.94).

En dépit de ces différents biais, et surtout en l'absence de données d'autre nature, les prévalences observées dans les sites sentinelles ont longtemps été considérées comme étant suffisamment représentatives de la population d'ensemble. Les premières estimations officielles de prévalence ont donc été élaborées sur cette base (Schwartländer et al. 1999, Walker et al. 2003). Un des enjeux de cette approche est évidemment l'estimation des *prévalences masculines* à partir des valeurs observées dans les SCPN - et donc uniquement chez les femmes.

Depuis 2001, le programme d'enquêtes EDS a intégré des tests sanguins anonymisés chez les adultes de 15 à 49 ans dans une grande majorité d'enquêtes. Ceci constitue une avancée majeure pour l'estimation de la prévalence en Afrique subsaharienne. Ces enquêtes fournissent en effet des estimations représentatives au niveau national et sous-régional, tant pour les femmes que pour les hommes. Dans les pays où elles sont disponibles (plus de 25 pays subsahariens), ces EDS sont utilisées aujourd'hui pour "calibrer" les tendances dégagées des sites sentinelles. En d'autres termes, le modèle EPP est conçu de façon à ce que les tendances de la prévalence issues des SPCN cadrent avec l'estimation issue des EDS (par milieu de résidence). *Ceci conduit quasi systématiquement à une révision à la baisse des prévalences estimées*. Une comparaison systématique des prévalences observées dans les EDS avec celles obtenues à partir des sites sentinelles a révélé que ces dernières devaient être réduites de 20% environ [10-30%] pour cadrer avec celles déduites des enquêtes, et ce, tant en milieu urbain qu'en milieu rural (Gouws et al. 2008). Dans les

63. Ainsi, pour les estimations de 2001, l'UNAIDS a réduit de 20% les prévalences rurales. Cet ajustement visait à tenir compte du fait que la prévalence est généralement *plus élevée* dans les régions mieux desservies en sites sentinelles - qui sont davantage situées en milieu urbain. Les premiers SPCN retenus pour suivre l'évolution du VIH-sida ont souvent été délibérément choisis dans des zones à prévalence élevée. Ceci perturbe également les tendances observées. Au fil du temps, l'extension des systèmes de surveillance aux zones où la prévalence est plus faible va générer une fausse impression de baisse de la prévalence. A partir des estimations de 2005, le modèle EPP a donc intégré un facteur d'ajustement spécifique, variable par pays, appelé "level fit". Cet ajustement permet de mieux tenir compte de cette extension de la couverture du territoire par les SCPN vers des zones où la prévalence est plus faible. Ces différents éléments sont discutés dans Walker et al. (2003), Brown et al. (2006).

pays où il n'existe pas d'enquête de séroprévalence au niveau national, les prévalences déduites des sites sentinelles sont donc réduites de 20% environ depuis 2007 [64].

Cette révision à la baisse de la prévalence, guidée par les résultats des enquêtes EDS, ne va pas sans soulever de questions. En effet, les enquêtes de séroprévalence souffrent elles aussi de biais bien spécifiques. C'est le cas notamment du refus des tests sanguins - un phénomène qui varie significativement d'une enquête à l'autre : Reniers et Eaton (2009) citent les chiffres de 4.4% de refus du test au Cameroun (2004), et jusqu'à 27.4% de refus au Malawi (2004). Ces refus pourraient induire une sous-estimation importante de la prévalence, car il semble que *les séropositifs ont davantage tendance à refuser de se faire tester que les séronégatifs* (Reniers et Eaton 2009). Cette éventuelle sous-estimation de la prévalence, si elle se confirmait, se traduirait par une sous-estimation de la mortalité adulte dans les pays fortement affectés par le VIH-sida (Reniers, Masquelier et Gerland, *à paraître*).

A ce stade, de nombreuses incertitudes entourent donc toujours les estimations de prévalence, en dépit des récentes enquêtes nationales, qui ont permis d'apporter une lumière nouvelle sur un phénomène qui était jusque là essentiellement appréhendé à partir des femmes enceintes.

2.4.2 Distributions par sexe et par âge des risques d'infection

Distribution par sexe

Dans la majorité des pays sub-sahariens, la prévalence du VIH est plus élevée chez les femmes que chez les hommes, comme on l'observe sur le graphe 2.11, que nous avons construit à partir des prévalences observées dans les enquêtes EDS [65]. La médiane du rapport entre la prévalence féminine et la prévalence masculine (du VIH-1) se chiffre à 1.56 sur ces 27 enquêtes EDS, menées entre 2001 et 2009.

La composition par sexe de la population séropositive varie avec "l'âge de l'épidémie". Les hommes ont d'abord été plus nombreux à être infectés au début de l'épidémie, en raison de comportements sexuels plus risqués (recours aux prostituées, relations sexuelles lors de migrations de travail, etc.). Par la suite, ce rapport s'est inversé, à cause des écarts d'âge aux rapports sexuels, mais également d'une plus grande vulnérabilité biologique féminine. De plus, comme la létalité est plus élevée lorsque l'infection est contractée aux âges avancés, les hommes décèdent plus vite, ce qui tend à diminuer davantage la prévalence masculine.

Le programme Spectrum propose *un schéma par défaut pour distribuer la prévalence par sexe* : le rapport de la prévalence féminine sur la prévalence masculine avoisine 0.24 pour des prévalences en dessous du seuil de 1%, puis il s'élève graduellement chaque année jusqu'à atteindre 1.38

64. Ceci vaut pour les deux milieux de résidence alors que seules les prévalences rurales étaient réduites de 20% depuis 2001. *Cf.* note 63.

65. Source : HIV/AIDS Survey Indicators Database, consultable sur www.measuredhs.com/hivdata (8/10/2010) et rapports nationaux EDS, consultés sur www.measuredhs.com

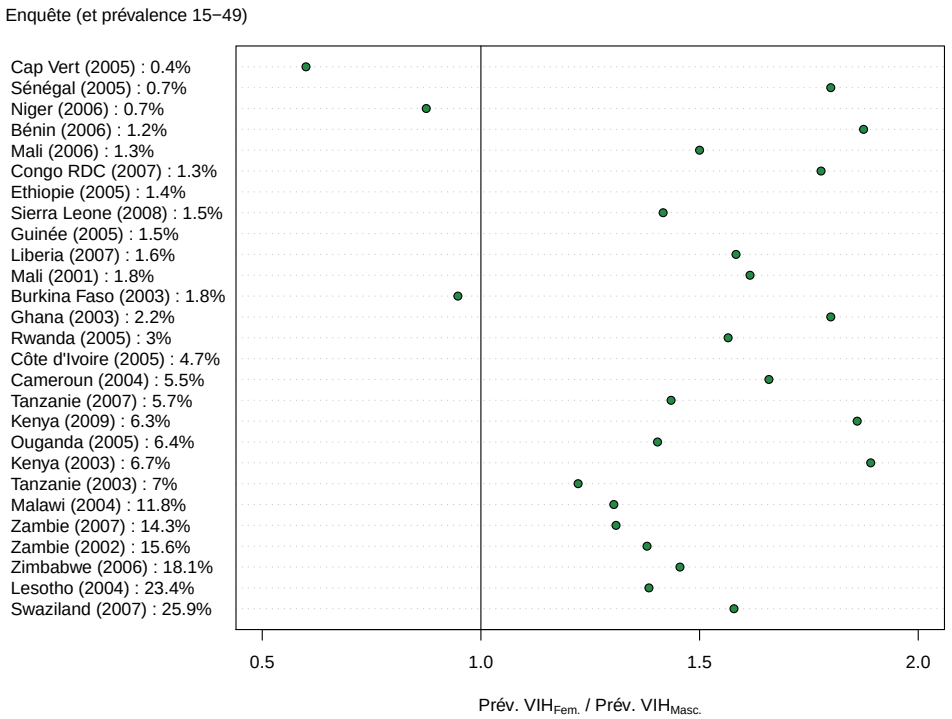

FIGURE 2.11 – *Ratio de la prévalence féminine sur la prévalence masculine chez les adultes de 15 à 49 ans (enquêtes EDS, nos calculs)*

après une vingtaine d'années, et se stabilise (Stover 2009). Ce ratio est présenté dans le graphe 2.15 (p. 86). Il semble que ce schéma soit également appliqué aux données d'incidence par Spectrum, ce qui constitue une approximation. En effet, le rapport de masculinité de la *prévalence* n'est pas égal au rapport de masculinité de l'*incidence*. La prévalence est influencée par la survie après l'infection, et donc par l'âge à l'infection, qui est moins élevé chez les femmes (*cf.* infra). Ces dernières vont vivre plus longtemps, et si les nouvelles infections sont distribuées à partir du ratio de la prévalence féminine sur la prévalence masculine, l'incidence féminine sera surestimée par rapport à l'incidence masculine. Dans les simulations, nous optons donc pour un ratio moins élevé, où les infections surviennent à 80% chez les hommes en début d'épidémie pour chuter à 45% au terme de quelques années. Il s'agit du ratio utilisé par les Nations Unies dans leur révision 2004 [66]. Notons que ce ratio est entouré de nombreuses incertitudes, et qu'il a un impact important sur les simulations. Nous y reviendrons.

Dans les simulations, les estimations d'incidence sont reprises de celles utilisées par les Nations Unies pour la Révision 2008 [67]. L'incidence totale est la moyenne des incidences masculine et féminine, pondérées par la population par sexe *non infectée*. Si R correspond au ratio de l'incidence féminine sur l'incidence masculine, et $pop_{t-1,\male}$ à la population adulte mascu-

66. Sans autres infomations, nous supposons qu'il est toujours d'application pour la Révision 2008.

67. A nouveau, nous nous basons ici sur des données non publiées, raison pour laquelle elles ne sont pas présentées - Source : United Nations, Department of Economic and Social Affairs, Population Division : World Population Prospects DEMOBASE extract. 2009.

line non infectée (Stover 2009), l'incidence féminine peut donc se calculer comme ceci :

$$\begin{aligned} Inc_{♀,t} &= Inc_{tot,t} \times (pop_{t-1,♂} + pop_{t-1,♀}) \\ &\quad / (pop_{t-1,♀} + pop_{t-1,♂} / R) \end{aligned}$$

L'incidence masculine s'obtient ensuite via l'équation :

$$Inc_{♂,t} = Inc_{♀,t} / R$$

Précisons que pour effectuer ce calcul, nous devons disposer de la population *non infectée* de 15 à 49 ans. Or celle-ci n'est pas déductible directement des structures de population distribuées par les Nations Unies [68]. Cette population est donc estimée annuellement à partir du résultat des microsimulations, et les taux d'incidence sont revus en conséquence au cours des microsimulations, tous les ans.

Distribution par âge

En Afrique subsaharienne, les écarts d'âge entre partenaires sexuels sont relativement élevés. Au Zimbabwe rural, par exemple, *les femmes ont des partenaires sexuels masculins qui sont en moyenne 7 ans plus vieux qu'elles, et dans un quart des partenariats, la différence d'âge s'élève au moins à 10 ans*. Ces écarts sont relativement constants avec l'âge des femmes (Hallett et al. 2007). De telles différences d'âge se traduisent par des schémas masculins et féminins de *prévalence par âge* très distincts. Il en va de même pour les schémas par âge de l'*incidence*. C'est ce qui nous intéresse ici, puisque nous cherchons à distribuer les risques d'infection entre les âges.

Les schémas par âge de la *prévalence* sont facilement observés à partir des tests sanguins collectés dans les enquêtes. Le graphe 2.12 présente de tels schémas calculés à partir de quatre EDS, menées au Cameroun en 2004, en Tanzanie en 2007-08, au Zimbabwe en 2005-2006 et en Zambie en 2007. Dans tous les cas, même si les niveaux de prévalence diffèrent d'un pays à l'autre, les pics de prévalence sont situés autour de 25-30 ans chez les femmes tandis que les hommes sont surtout affectés autour de 35-40 ans.

Si les schémas par âge de la prévalence s'observent facilement, les données requises pour caractériser le schéma par âge de l'*incidence* sont nettement plus rares, car les individus doivent être identifiés et testés au cours de plusieurs enquêtes sérologiques successives. Pourtant, ces estimations d'incidence sont beaucoup plus utiles que les prévalences observées, car la prévalence est déterminée *à la fois* par les risques d'infection *et* par la survie après l'infection. En d'autres termes, des modifications ou des spécificités régionales dans la durée de survie (par exemple en raison des ART) vont introduire des éléments de confusion.

68. La population estimée par le scénario "sans sida" ne correspond pas à la population non infectée mais à la population qui aurait été observée en l'absence de sida.

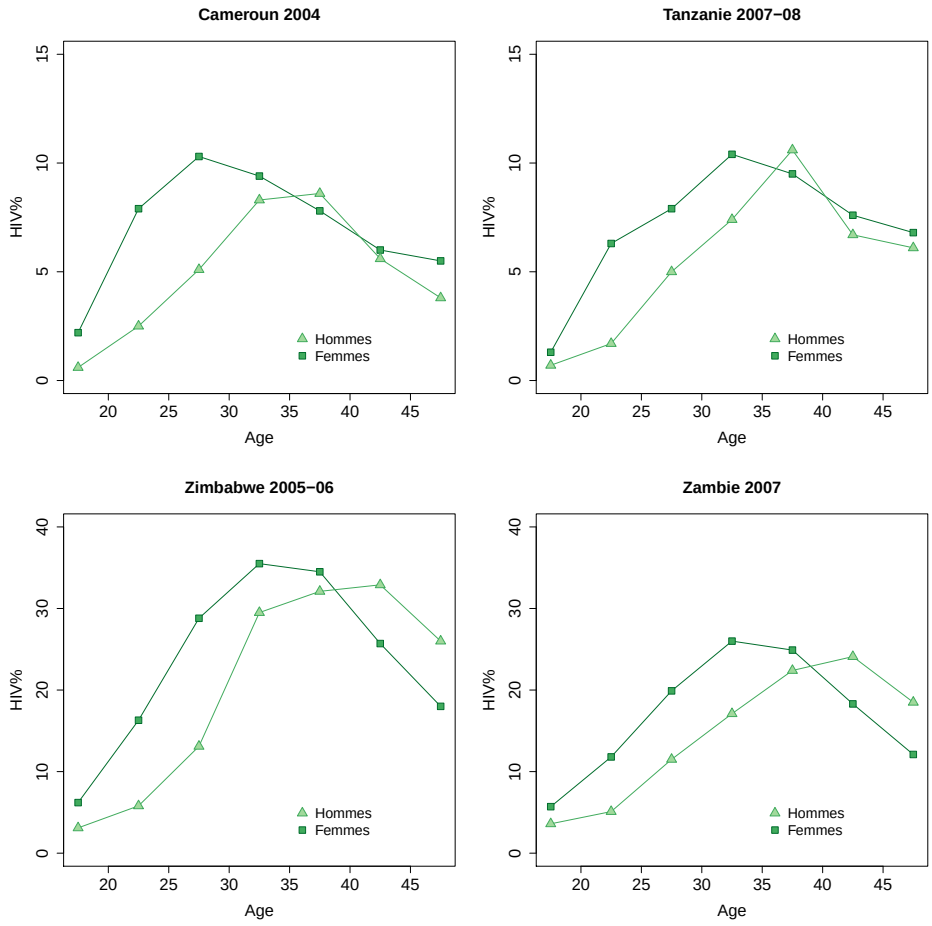

FIGURE 2.12 – *Schémas par âge de la prévalence (EDS Cameroun 2004, Tanzanie 2007-08, Zimbabwe 2005-2006 et Zambie 2007*

Pour pallier à ce manque de données sur l'incidence, deux méthodes ont récemment été proposées afin d'estimer l'incidence par âge à partir de plusieurs estimations de prévalence successives (Hallett et al. 2008). Elles visent à séparer, dans l'évolution de la prévalence, ce qui a trait aux *nouvelles infections* et ce qui tient à la *surmortalité des séropositifs*. La première méthode utilise des informations sur les taux de mortalité observés dans des sites de surveillance démographique (Zaba et al. 2007), la seconde se base sur des informations relatives à la survie après l'infection (Todd et al. 2007). Pour les détails techniques de l'implémentation de ces méthodes, nous renvoyons à l'article de Hallett et al. (2008).

FIGURE 2.13 – *Schémas par âge de l'incidence estimé par la méthode de Hallet et al. (2009) : Tanzanie 2003/4 - 07/8 et Zambie 2002-07*

La seconde méthode - basée sur la survie depuis l'infection - procure les schémas par âge d'infection présentés dans le graphe 2.13. Nous l'avons appliquée à deux pays qui font face à une épidémie mature et qui ont organisé deux enquêtes EDS successives avec tests sérologiques ; la Zambie (épidémie en baisse) et la Tanzanie (épidémie stable). Dans ces deux situations très diverses, on constate que l'incidence féminine démarre plus haut que l'incidence masculine, mais qu'elle baisse ensuite plus rapidement (ce qui est surtout apparent à partir des enquêtes zambiennes).

Les données manquent encore pour généraliser cette approche à de nombreux pays, mais à mesure que les enquêtes de séroprévalence se succèdent, il sera possible d'affiner l'estimation des schémas par âge d'*incidence*. Pour le moment, il faut encore recourir à des schémas théoriques d'infection par âge.

Quels sont ceux disponibles actuellement ? Le logiciel de projection Spectrum, utilisé par l'UNAIDS, prévoit un schéma théorique d'infection au VIH par âge et pour chaque sexe, basé sur les enquêtes sérologiques. [69] Depuis 2008, les Nations Unies utilisent de leur côté différents schémas

69. Un schéma est disponible pour les épidémies généralisées, d'autres schémas sont disponibles pour les épidémies concentrées. Ces différents schémas sont estimés en recourant à la méthode de Hallett et al. (2008), mais en postulant que la prévalence est stable dans le temps, ce qui permet d'utiliser les enquêtes des pays où une seule enquête est disponible.

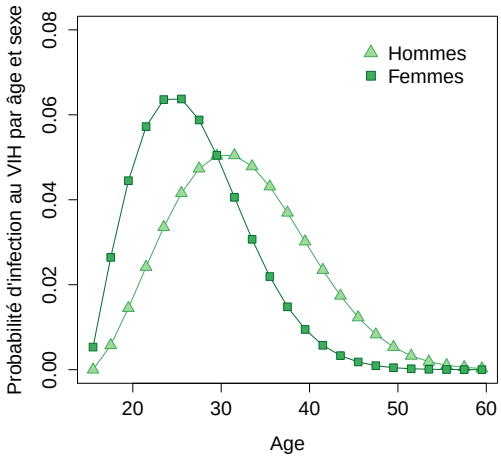

FIGURE 2.14 – *Schéma par âge d'infection au VIH par sexe - Modèle africain des Nations Unies*

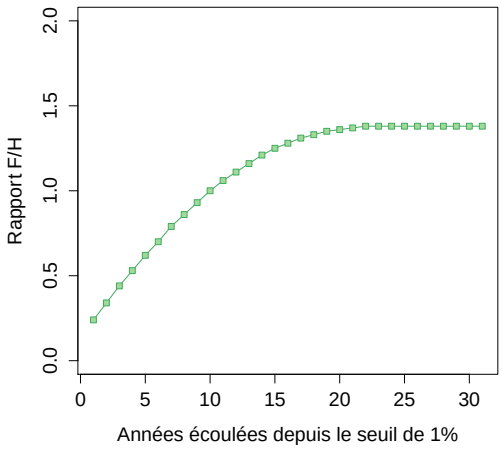

FIGURE 2.15 – *Distribution par sexe de la prévalence - modèle de Spectrum - Sc : Stover (2009)*

régionaux d'infection par âge, estimés également à partir des enquêtes EDS[70]. Ces schémas prennent tous la forme d'une distribution de Weibull. Le schéma "africain" procure un âge moyen à l'infection de 32 ans pour les hommes et 27 ans pour les femmes (UN 2009c). C'est cette répartition par âge que nous retiendrons pour mener les simulations. Elle est présentée dans le graphe 2.14[71]

La principale limite de notre série de microsimulations réside dans le fait que *les risques d'infection entre conjoints sont supposés être indépendants*. En effet, l'infection au VIH survient aléatoirement en fonction des estimations d'incidence nationale, des structures par âge atteintes et d'un schéma fixe de probabilité d'infection par âge. Ceci a au moins deux conséquences :

1. L'indépendance des risques d'infection entre partenaires signifie que sous sa forme actuelle, le modèle sous-estime sensiblement le nombre de *doubles orphelins.* Or un trait typique de l'épidémie du VIH-sida est qu'elle provoque une forte hausse de la fraction des orphelins qui perdent leurs deux parents (en raison des risques de

70. Voir http ://esa.un.org/unpp/index.asp ?panel=4

71. Pour chaque âge x, nous calculons les taux d'incidence par âge en multipliant l'incidence des 15-49 ans par ce schéma d'infection par âge puis en divisant cette valeur par la proportion de la population non infectée d'âge x parmi la population non infectée de 15-49 ans. Si $Pop_{x,\varphi}$ correspond à la population féminine *non infectée* d'âge x, on peut écrire :

$$\begin{aligned} Inc_{x,\varphi} &= \frac{Infections_{x,\varphi}}{Pop_{x,\varphi}} \\ &= Inc_{15-49,\varphi} * \frac{Infections_{x,\varphi}/Infections_{15-49,\varphi}}{Pop_{x,\varphi}/Pop_{15-49,\varphi}} \end{aligned}$$

Les taux d'incidence par âge et par sexe doivent donc être ré-éstimés annuellement durant les microsimulations, car ils dépendent de la distribution par âge et sexe de la population non infectée (une information dont nous ne disposons pas *a priori*). Les taux sont donc recalculés chaque année en fonction des résultats des années antérieures simulées.

transmission entre partenaires). Ainsi, les proportions de "doubles orphelins" parmi les orphelins s'élèvent de 10 à 17% dans les enquêtes EDS menées en Afrique de l'Est et en Afrique australe, contre 4 à 8% dans les autres pays d'Afrique centrale et de l'Ouest (Bicego et al. 2003). Au total, 19.3% des orphelins d'Afrique subsaharienne sont des "doubles orphelins" en 2010 (UNICEF 2006). A ce stade, nous ne pouvons donc pas analyser la situation des doubles orphelins au départ des microsimulations. Mais ceci n'a pas d'impact sur les analyses menées dans le cadre de cette thèse.

2. Par ailleurs, la mortalité des *orphelins de père* est sous-estimée dans les contextes perturbés par le VIH-sida. En effet, les pères décédés en raison du VIH-sida sont susceptibles d'avoir transmis ce risque d'infection aux mères, ce qui augmente la probabilité qu'un enfant de père décédé soit infecté au VIH par transmission verticale. Pour cette raison, les microsimulations ne rendent pas bien compte de l'évolution de la probabilité d'être orphelin de père dans les contextes perturbés par le sida. Cette limite nous a poussé à mener des simulations au niveau macro également, afin de tenir compte de cette transmission verticale, qui concerne indirectement les pères (*cf.* chapitre 4).

Dans l'objectif d'améliorer notre modèle, il faudra dans le futur tenter de lever cette hypothèse trop lourde de l'indépendance entre les risques d'infection des partenaires, afin d'obtenir des microsimulations mieux adaptées au suivi des relations de parenté dans les contextes perturbés par le VIH-sida[72].

2.4.3 Durée de survie des adultes infectés

Courbes types de survie après l'infection

La mortalité des séropositifs est généralement modélisée à partir de données relatives à la survie depuis l'infection (Todd et al. 2007), plutôt qu'à partir des taux de mortalité par âge des personnes infectées (Zaba et al. 2007). Dans les deux cas, les données pour l'Afrique proviennent essentiellement des observatoires de populations. Notamment, le réseau ALPHA, coordoné par la London School of Hygiene & Tropical Medicine, fédère six observatoires localisés en Afrique de l'Est et en Afrique aus-

72. Il est possible de recourir à une version alternative de SOCSIM, développée notamment dans le cadre de la thèse de E. Zagheni à Berkeley. Cette version prévoit *un facteur multiplicatif des taux de transition d'un groupe à l'autre* (ici des séronégatifs aux séropositifs) dans le cas où le partenaire (marié) est déjà infecté. En d'autres termes, les individus qui vivent dans un couple séro-discordant pourraient voir leur probabilité d'infection rehaussée par une constante. Grassly et Timaeus (2005) ont proposé une approche pour estimer à partir des EDS la probabilité qu'une femme soit infectée lorsque son partenaire est infecté, en faisant intervenir des données sur la prévalence nationale. Cette approche pourrait être utilisée pour estimer le facteur multiplicatif à retenir. Tout ceci rend toutefois la calibration du modèle plus délicate. Pour retomber, au terme des simulations, sur les incidences estimées par l'UNAIDS au niveau de la population générale, il faut pouvoir jongler avec les nombreux éléments en jeu ; les taux de nuptialité, les règles guidant la sélection des partenaires, les courbes d'incidence elles-mêmes, les schémas par âge d'infection, les durées de survie, etc. Ceci ne sera donc pas tenté dans le cadre de cette thèse mais il s'agit d'une piste de développements pour la suite.

trale[73]. Une récente méta-analyse des données de ces observatoires a permis de dégager des *courbes type de survies après l'infection* en l'absence de traitements anti-rétroviraux (Ghys et al. 2007, Todd et al. 2007). Ces données suggèrent que les durées de survie après infection sont globalement assez similaires à celles observées dans les pays occidentaux en l'absence de traitements[74]. Cette analyse n'a pas relevé de différences significatives par sexe dans les durées de survie, mais elle a par contre confirmé l'importance de l'*âge à l'infection*.

Dans le logiciel Spectrum, comme pour les estimations des Nations Unies, la survie n'est pas modulée selon l'âge à l'infection. Par contre, pour tenir compte du fait que les femmes sont infectées plus tôt et que les durées de survie diminuent avec l'âge à l'infection, *la distribution féminine procure une survie plus longue* (d'un an environ). Dans les deux cas, une distribution de Weibull est utilisée pour chaque sexe. D'autres distributions sont ensuite utilisées pour rendre compte de la survie au sein du stade sida, *avec* ou *sans* recours aux thérapies anti-rétrovirales (ART).

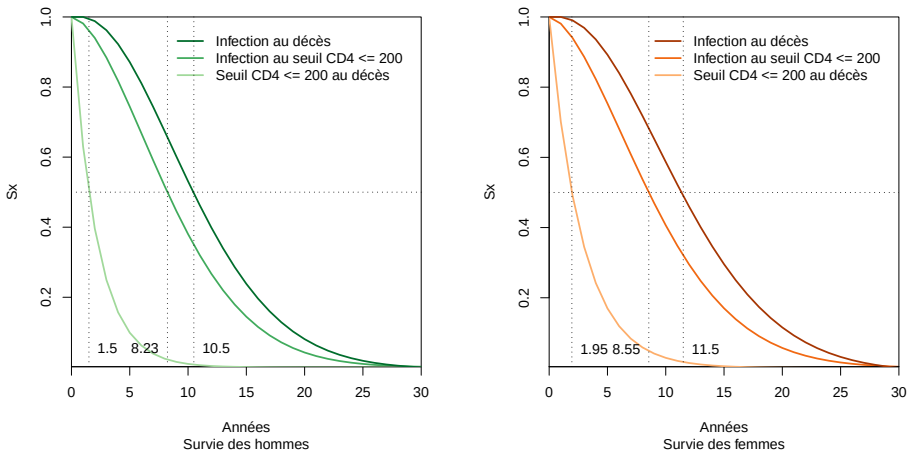

FIGURE 2.16 – *Courbes de survie des individus infectés progressant de l'infection au stade clinique, puis du stade clinique au décès, en l'absence de traitement - Sc : Stover (2009)*

La survie après l'infection est donc généralement modélisée *en deux temps*, en distinguant la phase du sida clinique (*full blown aids*) de la période qui la précède. Le graphe 2.16 présente les courbes de survie utilisés par Spectrum pour les individus progressant de l'infection au stade clinique (1e étape), puis du stade clinique au décès (1e étape) - en l'absence de traitement (Stover 2009). Les courbes masculines sont à gauche et les courbes féminines à droite. Ces courbes sont construites de façon à recouper la survie des séropositifs observée dans les observatoires du réseau ALPHA, mais sans tenir compte de l'âge à l'infection. Le seuil retenu pour délimiter la phase asymptotique et la phase du sida

73. Africa Centre (Afrique du Sud), Karonga (Malawi), Kisesa (Tanzanie), Masaka et Rakai (Ouganda), et Manicaland (Zimbabwe). Voir *Analysing Longitudinal Population-based HIV/AIDS data on Africa* http ://www.lshtm.ac.uk/cps/alpha/

74. Soit des durées médianes de survie d'un peu plus de 11 ans.

est établi à partir du nombre de protéines CD4 dans l'organisme [75]. Différents seuils sont proposés, en fonction de l'éligibilité des malades aux traitements anti-rétroviraux (qui sont administrés dans la seconde phase de la maladie). Jusqu'il y a peu, l'OMS recommandait de démarrer les traitements lorsque le nombre de protéines CD4 dans l'organisme était ≤ 200 cellules/mm^3, ce qui correspond (approximativement) au stade du sida lorsqu'il est diagnostiqué cliniquement. *Les courbes de survie que nous retiendrons pour les simulations sont celles qui correspondent à ce seuil de 200 cell./mm*^{3}. Ce seuil a été relevé à 350 cellules/mm^3 en 2009 par l'OMS, suite à l'observation que les traitements procuraient de meilleurs résultats quand ils étaient démarrés plus tôt. Mais dans la pratique, les traitements démarrent souvent avec l'apparition des premiers signes cliniques.

Pour les hommes, la durée médiane de l'infection jusqu'au seuil ≤ 200 cellules/mm^3 est de 8.23 ans (k = 10.2, λ =1.71) dans le logiciel Spectrum, et la durée médiane de ce seuil au décès est de 1.5 ans (k = 2.16, λ =1). Pour les femmes, ces valeurs valent respectivement 8.55 ans (k = 10.65, λ =1.67) et 1.95 ans (k = 2.82, λ =1). Si une distribution de Weibull est utilisée de l'infection au stade du sida, la survie dans le stade du sida est donc modélisée par le biais d'une distribution exponentielle (puisque le paramètre λ vaut 1). Précisons que les paramètres de ces distributions sont revus quasiment annuellement, ce qui complique quelque peu la comparaison des différentes projections [76].

Notons également que ces courbes sont conçues en combinant les décès dûs au VIH-sida et ceux non associés au VIH-sida. Elles ne saisissent pas l'effet 'net' de la mortalité due au VIH-sida. Chez les personnes infectées, la mortalité due au sida est telle qu'elle rend négligeable les variations dans la mortalité sans sida (Marston et al. 2007).

Du côté de la Division de la Population des Nations Unies, des distributions comparables sont utilisées. Pour la Révision 2008, les durées totales (de l'infection au décès) s'établissent à environ 10.5 ans pour les hommes et 11.5 ans pour les femmes. Par rapport à Spectrum, la période asymptotique est plus courte dans les estimations des Nations Unies et la période clinique est plus longue [77]. Les durées médianes passées dans la phase asymptotique se chiffrent à 7.5 ans pour les hommes et 8.5 ans pour les femmes (contre 8.23 et 8.55 ans dans Spectrum) [78], et ces durées s'élèvent à 3 ans dans la période clinique pour les deux sexes (contre 1.5 ans dans Spectrum). Là encore, les changements qui surviennent entre chaque Révision compliquent quelque peu la comparaison entre les diffé-

75. Lorsqu'un seuil plus élevé est retenu, la phase d'incubation est plus longue et la phase clinique plus courte.

76. En 2004, le temps de survie médian entre l'infection et le décès était estimé à 8.6 ans pour les hommes et 9.4 pour les femmes pour les pays en développement dans Spectrum (Stover 2004). Les paramètres utilisés en 2009 procurent des temps de survie médians (en l'absence de traitement) de deux ans supérieurs environ : 10.5 ans pour les hommes et 11.5 ans pour les femmes (Stover 2009).

77. La survie au sein du stade clinique n'est pas modélisé sous forme exponentielle mais au moyen d'une distribution de Weibull.

78. Sc : http ://esa.un.org/unpp/

rents modèles [79].

La modélisation de la survie en deux temps est nécessaire pour intégrer les *traitements anti-rétroviraux* (ART), qui réduisent la vitesse de la réplication des particules virales. L'Afrique subsaharienne a enregistré une diffusion rapide des ART dans les dernières années. En 2007, la proportion de séropositifs en besoin de traitement qui pouvaient disposer d'ART s'élevait déjà à 30% [27-34%] dans la région, contre à peine 2% quatre ans plus tôt (WHO 2007b). Cette couverture en ART était même estimée à plus de 50% dans quatre pays sub-sahariens : le Rwanda, le Sénégal, le Bostwana et la Namibie. Ces ART sont administrés dans la seconde étape de la progression de la maladie [80].

Alors que le programme Spectrum tient compte des ART mais pas de l'âge à l'infection, nous menons des microsimulations avec SOCSIM en tenant compte des deux à la fois. Dans la perspective d'analyser la probabilité d'être orphelin de mère dans les contextes perturbés par le sida (chapitre 3 et 4), il nous paraît important de prendre en compte ces deux éléments, qui vont avoir un impact sur les âges au décès.

Les données d'observatoires du réseau ALPHA permettent de dégager des courbes de survies après l'infection *variant avec l'âge à l'infection*. Pour chaque groupe d'âge, la survie peut être modélisée avec une distribution de Weibull dont le paramètre d'échelle λ varie (Hallett et al. 2008). Ces paramètres sont repris dans les 4 premières colonnes du tableau 2.1 [81]. Les courbes correspondantes sont présentées dans le graphe 2.17. *Ce sont ces paramètres qui seront utilisés dans nos microsimulations* [82].

Groupe d'age	k	λ	Médiane (Weibull)	Médiane (PH-type)
15-19	2	16.0	13.3	13.0
20-24	2	15.4	12.8	12.5
25-29	2	14.1	11.7	11.5
30-34	2	12.1	10.1	9.8
35-39	2	11.0	9.2	9.0
40-44	2	10.1	8.4	8.2
45-49	2	7.9	6.6	6.5

TABLE 2.1 – *Paramètres des distributions de Weibull modélisant la survie après l'infection - Sc : (Hallett et al. 2008)*

Ces distributions renvoient à la survie *de l'infection jusqu'au décès*. Or, si nous voulons reclasser certains individus comme étant sous traitement

79. A ce jour, les annexes méthodologiques n'ont plus été publiées par la Division de la Population depuis 2006.

80. Comme on l'a relevé, l'OMS recommande aujourd'hui de démarrer les traitements au seuil de 350 cellules/mm^3.

81. La quatrième colonne est discutée plus loin, dans la note 84 (p.84).

82. Notons que ces paramètres sont issus d'observatoires d'Afrique australe et d'Afrique de l'Est. L'épidémie y a atteint la maturité dans de nombreux pays, et le VIH-1 est le type prédominant. Il est possible que la distribution des durées d'incubation diffèrent d'une région à l'autre, notamment du fait que la charge virale du VIH-2, surtout présent en Afrique de l'Ouest, est moindre.

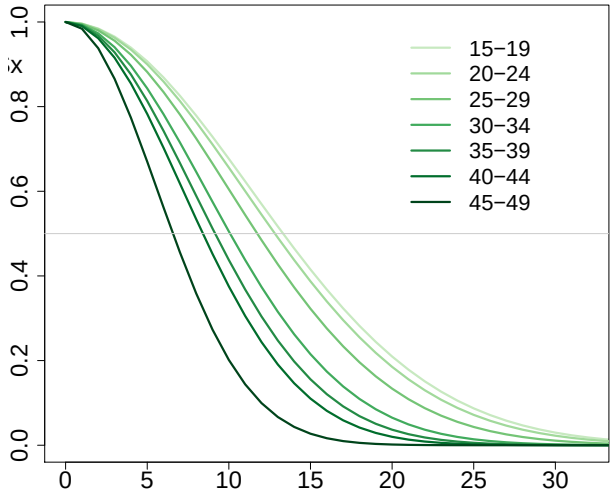

FIGURE 2.17 – *Courbes de survie des individus infectés progressant de l'infection au décès, par âge - Sc : Hallett et al. (2008)*

ART, nous devons pouvoir identifier les individus éligibles à ces traitements (ceux qui ont atteint le stade clinique).

Notre objectif est donc de décomposer chacune de ces distributions en deux parties, la première renvoyant au temps passé dans la phase *asymptotique*, la seconde au temps passé en phase de *sida clinique*. Pour cela, nous devons d'abord modéliser la survie au sein du stade clinique.

Survie au sein du stade du sida clinique

La survie dans la période du sida clinique peut être modélisée au moyen d'une distribution exponentielle, dont le taux vaut 1/2.49. Ceci correspond à la moyenne des valeurs masculines et féminines estimées par Spectrum [83].

A partir de là, nous pouvons isoler facilement, dans les fonctions de survie, le temps passé dans l'étape du sida clinique et le temps passé en phase asymptotique. La démarche que nous utilisons est détaillée dans l'annexe A.2.8. En deux mots, il s'agit d'approximer les distributions de Weibull par une combinaison de quatre exponentielles, en fixant à l'avance le taux de la dernière exponentielle (qui correspond au taux de mortalité au sein du stade clinique). Pour obtenir le temps écoulé de l'infection au stade clinique, il suffit ensuite de ne retenir que les trois premières exponentielles. Le graphe 2.18 présente le résultat de cet ajustement. Les courbes de Weibull pour les âges à l'infection de 15-19, et 45-49 ans (en traits pleins) sont comparées aux distributions approximées (en pointillés) [84]. Toujours sur ce graphe 2.18, les courbes correspondant à la survie de l'infection au stade clinique sont présentées sous forme de points.

83. Il s'agit plus précisément de la moyenne des deux valeurs de λ^{-1} de la distribution de Weibull avec $k = 1$, utilisées par Spectrum lorsque le seuil est fixé à $\leq$ 200 cellules CD_4/mm^3.

84. Les médianes des courbes de survie (totales) ré-estimées sont reprises dans la quatrième colonne du tableau 2.1. Elles sont légèrement en dessous de celles des distributions de Weibull de départ, mais de 3 mois au maximum. On peut donc considérer que l'ajustement est bon.

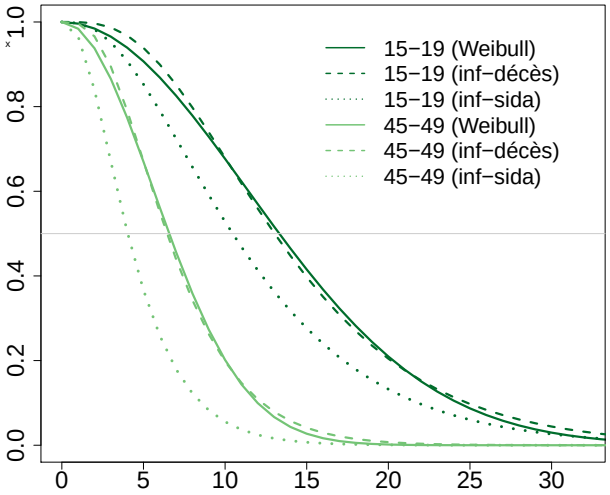

FIGURE 2.18 – *Courbes de survie des individus infectés progressant de l'infection au décès, par âge (Hallett et al. 2008), ajustées par distribution phase-type en quatre étapes*

Traitements anti-rétroviraux

Les *traitements ART* peuvent ajoutés dans les simulations, puisque l'étape du sida clinique est maintenant clairement identifiée. Une certaine proportion d'adultes ayant atteint l'étape du sida sont classés comme étant sous traitement et voient leur risques de mortalité baisser. La probabilité de survie sous ART est estimée à 85% la première année, puis en moyenne à 95% pour chaque année supplémentaire sous traitement (chez les adultes et en l'absence d'autres causes de décès) (Stover et al. 2008) [85].

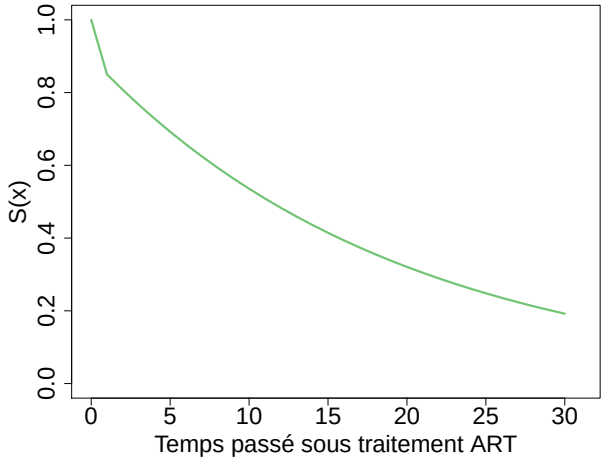

FIGURE 2.19 – *Courbes de survie des individus progressant de l'initiation des traitements ART vers le décès*

Les plus récentes estimations de la couverture des ART (c'est-à-dire le nombre de personnes en traitements sur le nombre de personnes en besoin de traitements) sont publiées dans WHO (2007b). Cette couver-

85. Le fait que de nombreuses personnes débutent les traitements à des stades avancés de l'infection explique partiellement la plus forte mortalité en début de traitement (Braitstein et al. 2006)

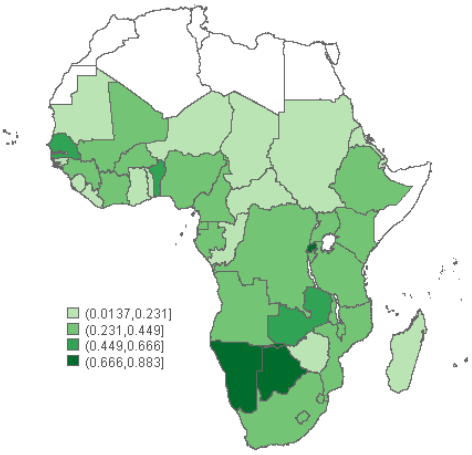

FIGURE 2.20 – *Couverture estimée des traitements antirétroviraux (tous âges) (décembre 2007) - Sc : WHO (2007b)*

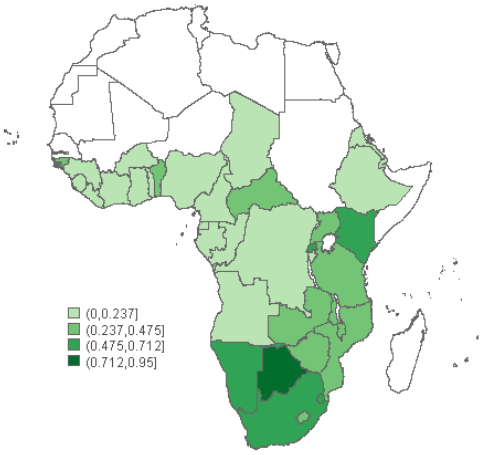

FIGURE 2.21 – *Estimation du pourcentage de femmes enceintes séropositives qui ont reçu des traitements antirétroviraux pour la prévention de la transmission verticale (2007) - Sc : WHO (2007b)*

ture est présentée pour 2007 sur la carte 2.20, pout tous âges confondus. Pour les simulations, nous utilisons des séries temporelles au niveau national et pour les adultes uniquement, reprises des estimations WHO/UNAIDS/UNICEF[86]. Sur cette base, une série de séropositifs adultes arrivés au stade du sida clinique sont tirés au hasard et redistribués vers le traitement en ART (en début de chaque année). C'est la seule transition qui n'est pas opérée de façon stochastique au cours des simulations.

Notons qu'à ce stade, nous n'intégrons pas de possibilité de traitement chez les enfants, car nous ne disposons pas de séries temporelles relatives à la distribution des traitements chez ceux-ci. Il conviendrait également de tenir compte d'un schéma par âge et sexe du recours aux ART chez les adultes. Les traitements antirétroviraux permettant de limiter la transmission verticale (PMTC) ne sont pas non plus intégrés dans les simulations[87].

2.4.4 Transmission verticale

Sans traitement, le risque de transmission verticale du VIH-1 est généralement estimé à 15-25% *in utero* ou lors de l'accouchement. *Nous fixerons ce risque à 20% sans ART dans les simulations.* 5-20% d'enfants non infectés

86. Il s'agit de séries temporelles - non publiées - utilisées par la Division de la Population des Nations Unies pour la Révision 2008 des WPP (UN 2009c).

87. Ceux-ci se sont rapidement diffusés dans les dernières années. En Afrique subsaharienne, on estime qu'en 2007, environ 45% [37-58%] des femmes séropositives enceintes ont reçu des antirétroviraux permettant de limiter la transmission verticale, alors que cette proportion ne s'élevait qu'à 9% en 2004 (WHO 2007b). Le pourcentage de femmes enceintes séropositives pouvant bénéficier de traitements pour réduire la transmission verticale est présenté par pays pour 2007 sur la carte 2.21. Mais ces traitements PMTC ne sont pas intégrés aux simulations.

intra-partum sont également infectés durant l'allaitement (WHO 2007a). La probabilité de transmettre le VIH par la voie de l'allaitement varie de 0.75% par mois en cas d'allaitement exclusif à 1.5% par mois en cas d'allaitement mixte (chez les enfants de 1 à 6 mois). Stover (2009) présente des proportions de séropositives qui pratiquent l'allaitement exclusif ou mixte [88]. *Dans les simulations, une probabilité mensuelle d'infection est calculée par âge en tenant compte à la fois des risques variables de transmission et de ces proportions.*

Deux types de progression du VIH peuvent être distingués chez les enfants ; une progression *rapide* et une progression plus *lente*. Pour tenir compte de cette progression à deux vitesses, la survie des enfants infectés hors autres causes de décès (ou, en d'autres termes, l'effet net du VIH sur la probabilité de survie) est modélisée avec une double distribution de Weibull (UNAIDS 2002a, Marston et al. 2005). Selon les paramètres utilisés par les Nations Unies pour leur dernière Révision, ceux qui progressent lentement ont, en moyenne et en l'absence de traitement, une durée de survie de 15.2 ans, et les autres une durée de survie de 1.3 ans (UN 2009c) [89]. Ces durées de survie sont plus longues que celles initialement estimées au début des années 2000, comme on l'observe sur le graphe 2.22, qui compare les distributions retenues en 2002 (UNAIDS 2002a), en 2005 (Marston et al. 2005) et 2008. Pour les enfants infectés in-utéro, la survie varie avec l'âge, et la fonction de hasard de la distribution de Weibull correspondante peut être directement introduite comme telle dans les simulations [90]. Pour les enfants infectés par allaitement, nous procédons de la même façon, à ceci près que la survie est modélisée selon la durée passée depuis l'infection et pas depuis la naissance. Dans les deux cas, ces enfants restent exposés aux autres causes de décès [91].

2.4.5 Réduction de la fécondité chez les femmes séropositives

Différents facteurs expliquent que les femmes séropositives aient une moindre fécondité ; la forte morbidité et la mortalité de leur partenaire, la moindre fertilité des femmes infectées, ou la plus grande proportion de fausses couches, par exemple (Zaba et Gregson 1998, Lewis et al. 2004). Cette moindre fécondité caractérise les séropositives de 20 à 45 ans, car, par un effet de sélection du à une activité sexuelle précoce, les plus jeunes séropositives sont *plus fécondes* que les femmes séronégatives de leur âge.

88. Ces proportions sont établies pour l'ensemble de l'Afrique subsaharienne.

89. Bien qu'il existe peu de données sur les durées de survie par type d'infection (intra-partum ou post-partum), les *slows progressors* sont généralement considérés comme ayant été infectés post-partum et les *fast-progressors* comme ayant été infectés *in utero* (Stover et al. 2006)

90. La fonction de hasard de la distribution de Weibull vaut :

$$h(x;k,\lambda) = (k/\lambda)(x/\lambda)^{(k-1)}$$

91. Cette prise en compte de la transmission verticale dans les simulations pourrait être améliorée à plusieurs égards. Notamment, il faudrait tenir compte à la fois des procédés visant à réduire la transmission intra-partum (*cf.* supra) et de ceux destinés à diminuer les risques associés à l'allaitement. Ces différents éléments sont bien intégrés au sein du logiciel Spectrum ou du programme ASSA (Stover 2009, Dorrington et al. 2006a).

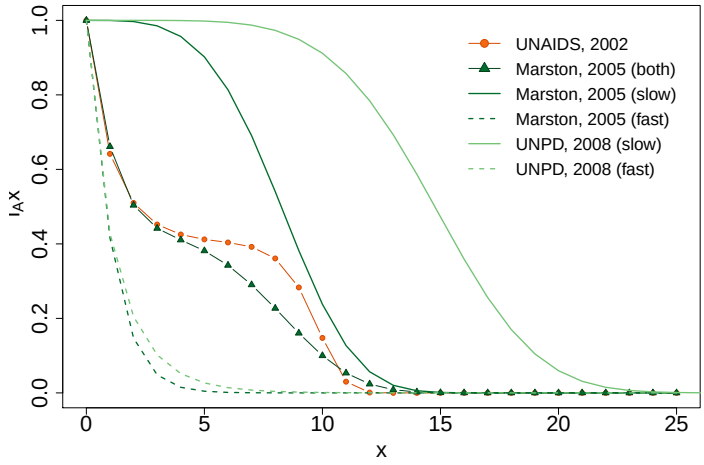

FIGURE 2.22 – *Courbes de survie des enfants de l'infection au décès, en l'absence d'autres causes*

Le ratio de la fécondité des séropositives sur la fécondité des sérogénatives tend également à diminuer avec l'âge. Lewis et al. (2004) estiment ce ratio à 1.26 [1.04-1.52] pour les 15-19, 0.76 [0.66-0.89] pour les 20-24 et 0.67 [0.6-0.74] pour les 25-44 ans [92]. *Ce sont ces valeurs que nous retiendrons dans les simulations* [93]. Nous supposons que cette baisse de la fécondité est constante dans le temps [94].

***

Afin de synthétiser les différentes transitions relatives au VIH-sida introduites dans les simulations, le tableau 2.2 en reprend les principales caractéristiques. Comme nous l'avons précisé, dans certains cas, les taux de transitions varient avec l'âge (c'est le cas des risques d'infection), et dans d'autres cas, ils varient avec la durée (c'est le cas de progression des personnes infectées vers le stade du sida clinique). De plus, ce tableau distingue les groupes qui restent exposés aux autres causes de décès (par exemple, les personnes sous traitement), des autres groupes qui ne le sont pas - parce que les courbes de survie contiennent déjà les décès non liés au sida.

92. Lewis et al. (2004) ont également cherché à estimer l'impact de cette réduction sur l'ISF total ; ils estiment que pour chaque point de pourcentage supplémentaire de prévalence chez les femmes de 15 à 44 ans, l'ISF de la population générale diminue de 0.37%.

93. En fixant par défaut une baisse de 0.31 pour les personnes ayant atteint le stade du sida clinique ou en traitement (Grassly et Timaeus 2005).

94. Les taux de fécondité de l'ensemble de la population sont ensuite ré-ajustés pour cadrer avec les niveaux estimés par les Nations Unies. Notons que les projections de population des Nations Unies se basent sur une réduction de fécondité de 20% chez les femmes séropositives, quelque soit leur âge (UN 2005).

Transition	Groupes	Type	Mortalité autres causes (avant la transition)
Infection	1>(2 :8)	Age	Grp 1 exposé
Vers le stade clinique	(2 :8)>9	Durée	Grps 2 :8 non exposés
Vers le décès (sida)	9>D	Durée	Grp 9 non exposé
Vers le décès (ART)	10>D	Durée	Grp 10 exposé
Vers l'infect. (postpartum)	(2 :11)>13	Age	Grps 2 :8 non exposés Grp 11 exposé
Vers le décès (intrapartum)	12>D	Age	Grp 12 exposé
Vers le décès (postpartum)	13>D	Durée	Grp 13 exposé

TABLE 2.2 – *Résumé des transitions, des groupes concernés et des types de transitions pour intégrer le VIH-sida dans SOCSIM*

Conclusion

Dans ce chapitre, nous avons détaillé les différents éléments utilisés pour calibrer une série de microsimulations sur les trajectoires démographiques des pays d'Afrique subsaharienne.

L'attention a été surtout été accordée ici à l'intégration de risques de décès spécifiques au VIH-sida dans le programme SOCSIM. Pour modéliser l'impact démographique du VIH-sida au niveau individuel, nous partons des tables "sans sida" des Nations Unies, puis nous décomposons l'infection et la survie en différentes étapes, ce qui permet à la fois de définir une durée de survie qui varie par âge *et* par durée depuis l'infection. La prise en compte de ces deux éléments est à notre connaissance assez rare dans les programmes de microsimulations consacrés au VIH-sida.

Nous procédons donc en deux temps, (1) en partant d'abord de tables sans sida, puis (2) en y ajoutant des décès liés au sida. Nous avons vu précédemment que la mortalité "sans sida" recalculée à partir des simulations cadrait bien avec les niveaux de mortalité "sans sida" introduits à partir des estimations de l'ONU (p. 69). Pour vérifier que la seconde étape se fait de façon cohérente, nous pouvons comparer les niveaux de mortalité "toutes causes" issus des simulations à ceux de la variante "avec sida" des Nations Unies. C'est ce que nous faisons dans le graphe 2.23, où sont présentées les probabilités féminines de décès de 15 à 60 ans, pour l'Ouganda et la Zambie. Les carrés correspondent aux estimations des Nations Unies, tandis que les traits correspondent aux valeurs recalculées à partir des microsimulations. Les traits et les points en vert renvoient à la mortalité non liée aux sida, à laquelle s'ajoute la mortalité due au sida, représentée en orange.

Dans ces deux pays, les microsimulations parviennent bien à capturer la forte hausse de la mortalité en raison du VIH-sida. Ceci indique que l'intégration du sida dans les simulations s'est faite de façon cohérente, ce que nous avions déjà pu observer à partir des pyramides des âges des populations simulées (p. 65). Les simulations prévoient toutefois une baisse de la mortalité un peu plus précoce que ne l'estiment les Nations Unies. Nous pensons que ceci tient aux hypothèses relatives à la distribution par

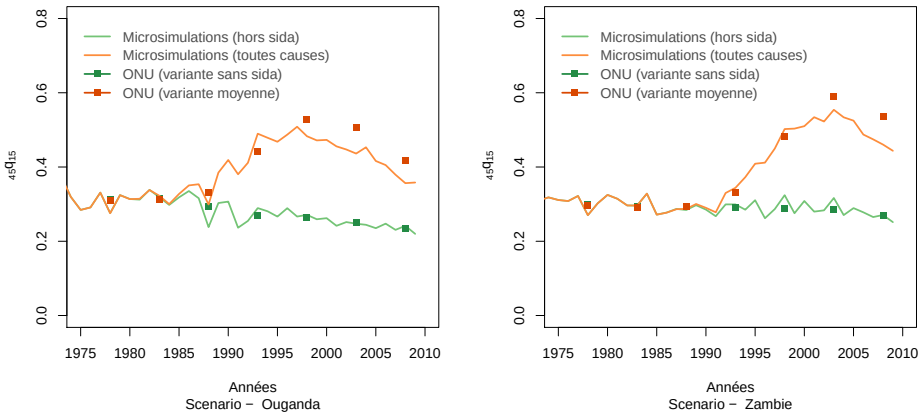

FIGURE 2.23 – *Probabilités féminines de décès entre 15 et 60 ans, mortalité "sans sida" et mortalité "toutes causes", comparaison des estimations des Nations Unies avec la mortalité recalculée des microsimulations - Ouganda et Zambie*

sexe de l'incidence, ainsi qu'au fait que nous avons modélisé la survie selon l'âge à l'infection.

La tentation est grande de raffiner le modèle de microsimulation pour qu'il colle au plus près de la réalité, mais il s'agit d'une course sans fin. Certaines limites de cette série de simulations sont toutefois évidentes et elles constituent autant de pistes de développements pour la suite :

- La première piste de développements consiste à *simuler des divorces, des remariages ainsi que de la polygamie*. Sans cela, nous observons des perturbations au niveau des schémas de nuptialité, comme nous l'avons expliqué brièvement plus haut (et détaillé en annexe A.2.7). Ceci est problématique, car tout ce qui dépend de la fécondité masculine est affecté par ce qui se passe au niveau des mariages simulés. En effet, SOCSIM n'attribue un père que lorsque les naissances surviennent au sein d'un mariage. Pour cette raison, nous ne pourrons pas exploiter les simulations pour analyser les méthodes qui permettent d'estimer la mortalité masculine à partir de la survie des parents.
- La seconde voie permettant d'affiner le modèle consiste à lever l'hypothèse selon laquelle *les risques de conception et d'infection au VIH sont indépendants*. Il conviendrait tout particulièrement de réfléchir à une meilleure articulation de ces deux phénomènes, ainsi qu'à une modélisation des risques d'infection dans les couples sérodiscordants [95]. Cela suppose d'intégrer des actes sexuels au sein des simulations, et de faire intervenir la fréquence des rapports, la probabilité d'infection par rapport, le recours aux préservatifs, etc. Afin de modéliser les risques d'infection, il s'agit également de mieux tenir compte de l'*age mixing* entre partenaires ; c'est-à-dire des différents types de sélection du partenaire par âge. Ceci pourrait être

95. *cf.* note 72, p. 87.

fait en recourant toujours à SOCSIM, mais au prix d'une modification importante du programme [96].

– Une troisième piste de développements futurs concerne la *prise en compte de la surmortalité des orphelins*. Des études récentes menées dans des observatoires de populations montrent que les enfants font face à des risques sensiblement plus élevés durant l'année qui précède et l'année qui suit le décès d'un parent, indépendamment du statut sérologique de leur mère (Nakiyingi et al. 2003, Ng'weshemi et al. 2003, Zaba et al. 2005). Nous y reviendrons dans le chapitre suivant.

A ce stade, malgré les limites que nous venons de relever, la série de simulations développée ici va permettre de reproduire assez fidèlement les changements démographiques survenus dans les différents pays africains en matière de mortalité, de fécondité, et d'infection au VIH. Elles sont menées au niveau individuel et les liens de parentés sont enregistrés pour tous les individus. Tout en reproduisant assez fidèlement les estimations des Nations Unies, elles nous donnent donc une clé d'entrée sur ce qui se passe au niveau individuel, et au niveau des structures de parenté. Dans les trois chapitres suivants, ces simulations vont nous permettent de revisiter les différentes techniques utilisées pour estimer la mortalité adulte à partir des mères, des frères et des soeurs.

96. *cf.* note 60, p. 76.

Estimer la mortalité adulte à partir de la survie des parents

3

Sommaire

3.1 Comment formaliser la probabilité d'être orphelin ? . 103
3.2 Convertir les proportions de survivants en probabilités de survie 110
3.3 Applications aux cas de fécondité et de mortalité non stables 136
3.4 Survie des parents aux âges adultes 154
3.5 Biais de sélection 163
3.6 Application aux populations simulées pour l'Afrique subsaharienne 176
Conclusion 194

La probabilité de décéder à l'âge adulte est étroitement liée à la proportion d'orphelins. Différentes méthodes ont donc été développées pour dégager des estimations de mortalité à partir de la survie des parents.

Dès les débuts, les chercheurs qui ont conçu ces méthodes ont toutefois insisté sur la nécessité de considérer les résultats obtenus comme sujets à caution. Trois éléments justifient cette prudence. Premièrement, cette méthode est construite sur des *hypothèses* qui sont rarement rencontrées dans la pratique. Deuxièmement, les informations sont collectées rétrospectivement et elles sont donc entachées de *biais de sélection*. Troisièmement, cette méthode s'applique typiquement sur des *données* fréquemment entachées d'erreurs. La sous-déclaration des décès, les erreurs sur les âges ou les approximations autour des liens de parenté posent les plus gros problèmes.

De ces trois aspects, le problème de la qualité des données est celui qui a reçu le plus d'attention - surtout l'effet d'adoption. Une proportion importante des jeunes enfants adoptés sont déclarés comme non-orphelins car le chef de ménage fait référence aux parents adoptifs et pas aux parents biologiques. Nous y reviendrons dans le chapitre 6. Par contre, relativement peu de recherches se sont consacrées à l'examen des biais de sélection, ainsi qu'aux effets de la violation des hypothèses à la base

des méthodes. Quelques travaux ouvrent toutefois des perspectives intéressantes. Notamment, Palloni et al. (1984) offrent une discussion très complète des différents biais de sélection qui affectent la méthode des orphelins. Un peu plus récemment, Timaeus et Nunn (1997) ont ouvert la voie à une révision de la méthode pour les contextes perturbés par le VIH-sida. Depuis, il y a eu peu de développements dans la littérature scientifique. Dans ce chapitre, nous souhaitons présenter les principaux aspects méthodologiques qui sous-tendent la méthode des orphelins, afin d'identifier les questions sur lesquelles des avancées méthodologiques pourraient être faites. Pour cela, nous proposons d'abord une brève synthèse des différentes méthodes disponibles, puis nous analyserons les biais de sélection et l'importance des hypothèses de base. Nous abordons les biais induits par le VIH-sida, mais ces derniers font surtout l'objet du chapitre suivant.

Nous l'avons noté, les données sur la survie des proches peuvent être exploitées de deux façons distinctes, selon le type de données à disposition. L'*exploitation directe* rapporte simplement les événements aux durées d'exposition correspondantes, quand celles-ci sont connues. Cette option directe est rarement utilisée, puisque l'âge des parents au décès est généralement inconnu là où est appliquée la méthode des orphelins. Lorsque les informations manquent sur la durée d'exposition (ou lorsqu'elles sont jugées de qualité douteuse), il faut mobiliser des *techniques indirectes*. Il s'agit alors de convertir des proportions de survivants en probabilités de survie après avoir contrôlé l'exposition au moyen d'approximations. Notons qu'il existe une option intermédiaire, dans les cas où une question sur la date de décès est ajoutée au questionnaire (Chackiel et Orellana 1985), mais c'est très rarement le cas (*cf.* chap. 1).

Nous nous concentrons donc ici sur l'option indirecte, qui est de loin la plus utilisée. Son principe est assez simple. Le chercheur observe une proportion d'orphelins, qui dépend à la fois du *niveau de mortalité* et de la *durée d'exposition* des parents au risque de mourir. Ce sont en fait l'*âge des enquêtés* et le *calendrier de la fécondité* qui vont déterminer l'exposition au risque :

- L'exposition débute lorsque les individus sont susceptibles de décéder comme parents. Les probabilités de survie estimées sont donc "conditionnelles" [1]. En moyenne, l'exposition débute à l'*âge moyen à la naissance des enfants* (encadré 5, p.102).
- Les individus sont ensuite soumis au risque de décès jusqu'à ce que leurs enfants soient interrogés dans l'enquête... En moyenne, la durée d'exposition se termine donc à un âge qui correspond à *l'âge moyen à la naissance des enfants majoré de l'âge des individus qui font référence à la survie de leurs parents*.

Après avoir contrôlé l'exposition, des coefficients correcteurs peuvent être calculés afin de passer des *proportions* de proches survivants aux *probabilités* de décéder aux âges adultes. Pour ce faire, la proportion de

1. Certaines méthodes procurent des probabilités de survies inconditionnelles (Hill et Trussell 1977), mais le calcul des coefficients se base sur des probabilités conditionnelles.

parents survivants est d'abord exprimée comme une combinaison de schémas de fécondité, de mortalité et de taux de croissance, en supposant que la population est stable. Une série de simulations est ensuite construite sur base de modèles démographiques (qui synthétisent l'intensité et l'évolution par âge de la fécondité et de la mortalité). Par exemple, la mortalité est modélisée à partir du système des logits de Brass (Encadré 6, p.112). Les paramètres de ces modèles sont amenés à varier pour représenter différents régimes de fécondité et de mortalité. Chaque fois qu'un paramètre varie, une nouvelle proportion de non-orphelins par âge est calculée par approximation numérique. Les probabilités de survie qui correspondent aux différentes valeurs des paramètres sont ensuite régressées sur les proportions de non-orphelins par âge, en contrôlant uniquement un nombre limité de paramètres (l'âge moyen à la naissance des enfants, en l'occurrence).

Brass et Hill (1973) sont les concepteurs de la méthode indirecte basée sur la survie des parents. Hill et Trussell (1977) ont par la suite élaboré la méthode basée sur la survie des frères et sœurs, dont la logique générale est identique. Elle est présentée dans le chapitre suivant. De nombreux autres coefficients ont été proposés dans le même esprit, pour les deux types de données (UN 1983b, Graham et al. 1989, Palloni et Heligman 1985, Timaeus 1992, Timaeus et al. 2001).

Afin de présenter la méthode des orphelins, de détailler ses limites et de suggérer certains développements, ce chapitre se structure en cinq temps.

- Les deux premières sections sont consacrées aux *principes généraux* et aux *méthodes* qui permettent de transformer des proportions de parents survivants en probabilités de survie.
- La troisième section présente les approches disponibles pour dater les estimations ou pour réduire les perturbations induites par les *changements de fécondité et de mortalité*.
- La quatrième section discute des approches basées sur la *survie des parents parmi les descendants ayant atteint l'âge adulte*, qui sont conçues notamment pour limiter l'impact du biais d'adoption.
- La cinquième section discute des biais de sélection qui ne sont pas spécifiquement liés au VIH-sida.
- La sixième et dernière section applique les différentes méthodes d'estimation et de datation aux microsimulations présentées dans le chapitre 2. Elle permet de montrer que la méthode des orphelins procure des résultats satisfaisants dans les contextes peu perturbés par le VIH-sida, mais que les estimations dégagées des contextes affectés par le sida sont assez éloignées des probabilités de décès qui sont à la base des simulations.

Le chapitre suivant s'inscrit dans la continuité de celui-ci. Il est consacré plus particulièrement aux perturbations introduites par le VIH-sida et aux amendements qui pourraient être apportés à la méthode des orphelins.

Encadré 5 L'âge moyen à la naissance des enfants

L'*âge moyen à la naissance des enfants* (mean age at childbearing) se calcule en prenant la moyenne pondérée de l'âge des femmes par le nombre de naissances. On le notera M ($M_{♀}$ pour les mères et $M_{♂}$ pour les pères). Il est différent de l'*âge à la maternité* (ou mean age of the fertility schedule), que l'on notera MAC : celui-ci s'obtient en prenant la moyenne de l'âge des femmes pondéré par les taux de fécondité. L'âge moyen à la naissance des enfants est donc influencé par l'effet de la mortalité sur le nombre de femmes.

Dans les recensements, il est possible de calculer approximativement la valeur de cet âge moyen en interrogeant les femmes sur le nombre d'enfants auxquels elles ont donné naissance dans les douze mois précédents. Dans ce cas, les naissances sont survenues six mois en moyenne avant le recensement. En supposant qu'elles sont distribuées uniformément dans l'année en cours, les centres de classes de l'âge des mères doivent donc être simplement décalés de six mois (17, 22, 27, etc.). Le nombre de naissances dans chaque classe est ensuite multiplié par le centre de classe, et tous les produits sont additionnés, puis cette somme est divisée par le nombre total de naissances :

$$M = \sum_{\alpha}^{\beta} a_i N_i / \sum_{\alpha}^{\beta} N_i$$

Il peut être calculé de la même façon dans les enquêtes.

Ce calcul n'est pas à l'abri des erreurs de déclaration ; on sait par exemple que les naissances récentes sont sous-déclarées dans les enquêtes EDS, surtout autour de la période de référence (3 ans avant l'enquête) notamment pour alléger la charge de travail des enquêteurs (Schoumaker 2008) [2]. Des sous-déclarations des enfants décédés rapidement sont également régulièrement observées. Si ces erreurs ne sont pas uniformément réparties dans les différentes classes d'âge, la valeur estimée de l'âge moyen sera vraisemblablement biaisée. Par ailleurs, la tendance à se "vieillir" souvent observée chez les femmes va conduire à une sur-estimation de M, et donc à une sous-estimation de la mortalité.

2. On pourrait tenter de contourner ce problème en travaillant à partir des histoires génésiques, et en calculant l'âge moyen par année. Mais ce calcul est biaisé à la baisse à mesure que l'on s'éloigne de l'enquête par le fait que les femmes âgées sont de moins en moins représentées dans l'échantillon (en raison de la troncature à 50 ans).

3.1 Comment formaliser la probabilité d'être orphelin ?

La démographie de la parenté a une longue histoire derrière elle, depuis les travaux de Lotka, qui, en 1931, s'interroge sur les relations entre la mortalité et la perte des parents dans le cadre des populations stables (Lotka 1931). Elle a connu ses plus importants développements avec les recherches de Goodman et al. (1974), Le Bras (1973), ainsi qu'avec les travaux recourant à la microsimulation.

De cette littérature, on peut dégager au moins deux approches pour parvenir à exprimer de façon analytique la proportion de parents survivants ; une méthode dite de *comptage* et une méthode dite *probabiliste* (Keyfitz et Caswell 2005).

Méthode de comptage

1. Tout part du nombre total de femmes de l'année en cours, qui est noté P.
2. Le taux d'accroissement annuel de cette population féminine est noté r et supposé constant. L'hypothèse est que la population est stable.
3. Dans ces conditions, la population va connaître une croissance exponentielle de sorte qu'il y a a années, elle s'exprimait comme $P\ e^{-r\,a}$. On s'intéresse à cette population parce que c'est celle qui caractérisait la situation à la naissance des enfants d'âge a interrogés dans l'enquête.
4. Combien de mères ont donné naissance à ces enfants d'âge a ? En travaillant par âge dans un premier temps, on va chercher à exprimer ce nombre de femmes d'âge x, observé il y a a années, qui est noté $n(x,a)$. Pour cela, il suffit de retourner encore en arrière jusqu'à la naissance des mères puis de les faire survivre : il s'agit donc d'appliquer aux naissances survenues il y a $a+x$ années (c'est-à-dire lors de la naissance des mères) une probabilité de survie de la naissance jusqu'à l'âge x ($_xp_0$). Pour obtenir ces naissances, le nombre total de femmes (il y a $a+x$ années) doit être multiplié par le taux de natalité (b). En définitive, le nombre de femmes d'âge x, observé il y a a années, peut donc s'exprimer comme ceci : [3]

   $$n(x,a) = b\ P\ e^{-r(x+a)}\ {}_xp_0$$

   Si B correspond au nombre de naissances de l'année en cours, on obtient :

   $$n(x,a) = B\ e^{-r(x+a)}\ {}_xp_0 \tag{3.1}$$

5. Toutes ces femmes d'âge x il y a a années n'ont pas survécu, puisqu'il leur restait encore a années à parcourir. Pour ne conserver que les femmes survivantes, ce nombre $n(x,a)$ est donc multiplié par une probabilité de survie de x jusqu'à a :

3. La littérature présente différentes terminologies, et nous avons cherché à privilégier l'uniformité.

$$n(x,a) = B\ e^{-r(x+a)}\ {}_{x+a}p_0$$

6. Nous disposons donc maintenant du nombre de femmes survivantes à l'enquête qui étaient âgées de x ans il y a a années. Comme on s'intéresse uniquement aux *mères* survivantes, il s'agit d'intégrer des informations sur la probabilité que ces femmes survivantes, il y a a années, aient accouché à l'âge x. Pour ce faire, l'expression $B\ e^{-r(x+a)}\ {}_{x+a}p_0$ doit être multipliée par $m(x)dx$, en supposant une fécondité constante [4]. On obtient le nombre d'enfants d'âge a dont la mère est survivante.

7. Jusque là, seules les femmes d'âge x ont été prises en compte. Pour considérer maintenant les femmes survivantes de tous âges *qui ont connu une naissance* il y a a années, l'expression $B\ e^{-r(x+a)}\ {}_{x+a}p_0\ m(x)dx$ est intégrée entre les âges de reproduction :

$$\int_\alpha^\beta B\ e^{-r(x+a)}\ {}_{x+a}p_0\ m(x)dx$$

8. La proportion de mères ayant accouché il y a a années qui ont survécu peut alors s'exprimer facilement. Il suffit de prendre au dénominateur toutes les femmes qui ont donné naissance il y a a années, et au numérateur, les femmes qui, parmi elles, ont survécu jusqu'à l'enquête ou au recensement :

$$S(a) = \frac{\int_\alpha^\beta e^{-r(x+a)}\ {}_{x+a}p_0\ m(x)dx}{\int_\alpha^\beta e^{-r(x+a)}\ {}_{x}p_0\ m(x)dx} \tag{3.2}$$

   On a là l'équation de base de la méthode des orphelins.

9. Cette équation peut être simplifiée ; en factorisant le terme e^{-ra} aux numérateur et dénominateur puis en l'annulant. Dans ce cas, le dénominateur vaut 1, en vertu de l'équation de Lotka-Euler (reprise en page 60). On retrouve alors l'expression dérivée pour la première fois par Lotka (1931) pour exprimer la probabilité qu'un enfant d'âge a ait une mère survivante sous les conditions actuelles de mortalité et de fécondité [5] :

$$S(a) = \int_\alpha^\beta {}_ap_x\ {}_xp_0\ e^{-rx} m(x)dx \tag{3.3}$$

10. En travaillant par groupe d'âge des enfants (de y à $y+5$), et en intégrant des risques de décès chez ces enfants sur l'intervalle d'âge, on obtient la proportion de mères survivantes parmi celles qui ont donné naissance à un enfant aujourd'hui âgé de y à $y+5$ ans :

4. La fécondité concerne ici tant les naissances féminines que masculines.

5. $S(a)$ est souvent noté $M_1(a)$, $M_2(a)$ se référant au nombre de filles dont la grand-mère est survivante.

$$
\begin{aligned}
{}_5S_y &= \frac{\int_y^{y+5} e^{-r(a)} \times_a p_0 \int_\alpha^\beta e^{-r(x)} \times_{x+a} p_0 \times m(x)dx\ da}{\int_y^{y+5} e^{-r(a)} \times_a p_0 \int_\alpha^\beta e^{-r(x)} \times_x p_0 \times m(x)dx\ da} \\
{}_5S_y &= \frac{\int_y^{y+5} e^{-r(a)} \times_a p_0\ S(a)\ da}{\int_y^{y+5} e^{-r(a)} \times_a p_0\ da}
\end{aligned} \tag{3.4}
$$

Après avoir spécifié différents schémas de fécondité et de mortalité, l'équation 3.4 peut être utilisée pour calculer par approximation numérique la série correspondante de proportions de non-orphelins de mères. Il est possible ensuite de mettre les proportions en vis-à-vis des probabilités qui les ont générées, afin de fournir des coefficients correcteurs.

Remarquons que ce qui est observé à partir des enquêtes et des recensements, c'est une proportion d'*enfants dont la mère est survivante*, les mères étant comptées autant de fois qu'elles ont d'enfants survivants. La méthode repose donc sur l'approximation de la proportion de mères survivantes par la proportion d'enfants dont la mère est survivante. Pour être exacte, cette approximation suppose que *chaque femme ait exactement un enfant survivant*. Plus largement, il faut au moins que :

1. la survie des mères ne soit pas associée à la survie des enfants,
2. la fécondité des mères ne soit pas associée à leur mortalité,
3. la taille de la fratrie ne soit pas associée à la mortalité des enfants.

Ces hypothèses sont rarement rencontrées dans la pratique, ce qui peut introduire des biais de sélection (nous y reviendrons plus loin).

Méthode probabiliste

Goodman et al. (1974) parviennent à l'équation 3.3 par une autre approche, qu'ils qualifient de probabiliste. Ils prennent l'exemple d'un enfant d'âge a et cherchent à exprimer la probabilité que cet enfant - pris au hasard dans une population caractérisée par les schémas de fécondité $m(x)dx$ et de mortalité $l(x)$ - ait une mère survivante. Si cette mère était âgée de x années à la naissance, la probabilité qu'elle soit survivante est simplement ${}_ap_x$. Ceci est conditionnel au fait que cette mère ait donné naissance à l'âge x. Pour éliminer cette condition, il faut prendre une moyenne sur tous les âges. Il faut toutefois tenir compte du fait que les mères ne sont pas distribuées *uniformément* entre les âges de reproduction. Cette moyenne doit donc être prise en donnant à chaque âge *un poids correspond au nombre de naissances qui surviennent à cet âge*. Si l'on note $W(x|t-a)$ la distribution par âge des mères il y a a années (en comptant les âges des mères au moment des naissances), on peut alors exprimer la proportion d'enfants d'âge a dont la mère est survivante comme :

$$
S(a) = \int_\alpha^\beta {}_ap_x\ W(x|t-a)dx \tag{3.5}
$$

Dans le cas d'une population stable, la structure de la population des femmes est constante dans le temps ; elle peut être exprimée comme $l_x m_x e^{-rx}$ (où r correspond au taux de croissance intrinsèque). En remplaçant $W(x|t-a)$ par $l_x m_x e^{-rx}$, on retrouve donc l'équation 3.3 :

$$S(a) = \int_{\alpha}^{\beta} {}_a p_x \; {}_x p_0 \; e^{-rx} m(x) dx$$

L'intérêt de cette approche "probabiliste" est qu'elle fait ressortir le fait que la probabilité (pour un enfant d'âge a) d'avoir une mère survivante correspond à la *moyenne des probabilités de survie féminine pondérées par la structure par âge des mères à la naissance il y a a années*.

Développements

La proportion de non-orphelins d'âge a peut être exprimée comme une pure mesure de mortalité qui dépend de l'âge moyen à la naissance des enfants (M), si l'hypothèse est faite que la fonction ${}_a p_x$ est linéaire en x (Preston et al. 2001). Dans ce cas, on peut alors introduire une constante K et une pente j, de sorte que la probabilité ${}_a p_x$ puisse s'écrire comme :

$${}_a p_x = K - (j \; x)$$

A partir de là, on peut facilement simplifier l'équation 3.5 obtenue de façon probabiliste :

$$\begin{aligned} S(a) &= \int_{\alpha}^{\beta} {}_a p_x \; W(x|t-a) dx \\ &= \int_{\alpha}^{\beta} K - (j \; x) \; W(x|t-a) dx \\ &= K - j \int_{\alpha}^{\beta} x \; W(x|t-a) dx \\ &= K - j \; M \\ &=_a p_M \end{aligned}$$

Ce développement nous montre que *la proportion de non-orphelins par âge devrait être proche de la probabilité de survie entre l'âge moyen à la naissance des enfants et cet âge majoré de l'âge des enfants* (a). C'est précisément ce qui motive le développement de la méthode des orphelins. Toutefois, si pour des petites valeurs de a (disons de 5 à 10), et pour certaines tables de mortalité, l'hypothèse de linéarité de ${}_a p_x$ est recevable, ceci n'est clairement pas le cas pour d'autres tables ou pour des valeurs de a plus élevées. Ceci peut s'observer sur le graphe 3.1, où nous faisons intervenir à titre d'exemple les tables Ouest et Sud de Princeton. Pour un nombre a valant 5 (panel du haut), la fonction ${}_a p_x$ est quasiment linéaire en x (surtout avec la table Ouest de Princeton). Par contre, avec des valeurs de a valant 20 (panel du bas), la linéarité n'est clairement plus de mise.

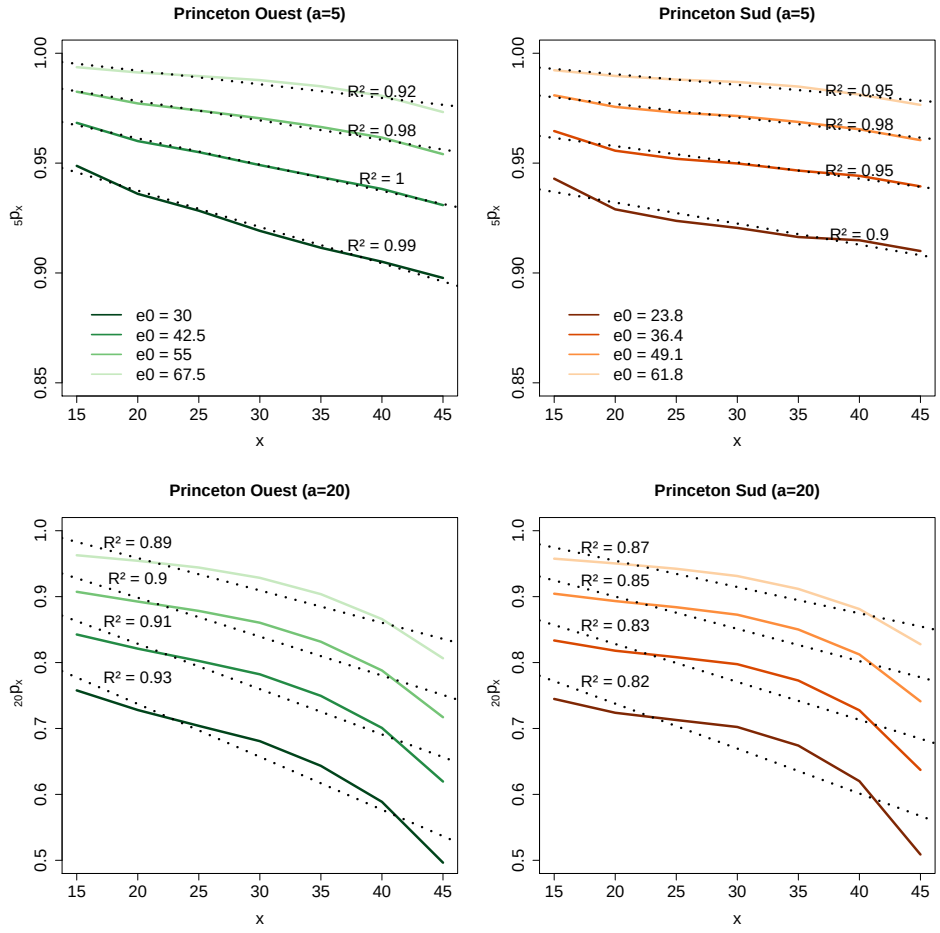

FIGURE 3.1 – *Valeurs des probabilités de survie ${}_5p_x$ et ${}_{20}p_x$ - Tables de Princeton Ouest et Sud, pour différents niveaux de mortalité et différentes valeurs de x.*
NOTE *: les valeurs des espérances de vie en soi n'ont pas d'importance ici.*

Une approximation plus fine de $S(a)$ est fournie par Keyfitz et Caswell (2005). Elle repose sur l'expansion de ${}_ap_x$ en une série de Taylor [6] centrée sur l'âge moyen à la naissance des enfants :

$$ {}_ap_x =_a p_M + (x - M)\ [{}_ap_M]' + \frac{(x-M)^2}{2!}\ [{}_ap_M]'' + ... \tag{3.6}$$

En intégrant les deux côtés de l'égalité sur tous les âges de reproduction après avoir multiplié par $e^{-rx}m(x)$, le second terme de la somme s'annule, ce qui procure - si l'on note σ^2 la variance des âges des mères à la naissance des enfants - l'équation suivante :

$$ S(a) \approx_a p_M + \frac{\sigma^2}{2}\ [{}_ap_M]'' \tag{3.7}$$

Cette équation fait ressortir deux éléments qui déterminent la qualité de l'approximation de la proportion de non-orphelins d'âge a par ${}_ap_M$:

1. Le premier élément est $[{}_ap_M]''$, la seconde dérivée de la fonction de survie ${}_ap_M$ évaluée en M. Cette seconde dérivée correspond à la *courbure* de la fonction de survie. Là où la courbe de la fonction ${}_ap_M$ sera orientée vers le bas (*concave down*), sa dérivée seconde sera négative, ce qui signifie que la probabilité ${}_ap_M$ sera une approximation qui va surestimer la proportion $S(a)$. C'est généralement le cas pour des valeurs de M inférieures à 40, quelle que soit la valeur de a. Ceci peut s'observer sur le graphe 3.2, que nous avons construit à partir du standard général de Brass. Les trois graphes de gauche présentent simplement l'évolution des probabilités ${}_5p_x$, ${}_{15}p_x$ et ${}_{35}p_x$ en fonction de x. Les secondes dérivées de ces trois courbes sont présentées dans les trois graphes de droite. Quelle que soit la valeur de x, ces secondes dérivées sont négatives avant l'âge de 40 ans. La courbure de la fonction de survie est ensuite orientée vers le haut (et la seconde dérivée devient positive), mais cela survient au delà des valeurs possibles de l'âge moyen à la naissances des enfants. En définitive, la courbure sera généralement orientée vers le bas, ce qui signifique que la probabilité ${}_ap_M$ sera plus élevée que la proportion $S(a)$.

   On note également, en comparant les graphes représentant $[{}_ap_M]''$ (graphes de droite de la figure 3.2), que pour des valeurs de M variant entre 20 et 40 ans, plus la valeur de a augmente (en comparant par exemple ${}_{35}p_x$ à ${}_5p_x$), plus la seconde dérivée $[{}_ap_M]''$ sera élevée en valeur absolue. Cela signifie qu'aux jeunes âges, les proportions de mères survivantes sont étroitement liées à la probabilité de survie ${}_ap_M$, mais que l'approximation est moins bonne aux âges élevés. On retrouvera ce constat à plusieurs reprises par la suite.

6. Une série de Taylor d'une fonction f (au voisinage d'un point a) est définie comme "une série entière construite à partir de f et de ses dérivées successives en a" (Wikipédia - 07/2010). Elle prend la forme suivante :

$$ f(a) + (x - a)\ [f(a)]' + \frac{(x-a)^2}{2!}\ [f(a)]'' + \frac{(x-a)^3}{3!}\ f^{(3)}(a)...$$

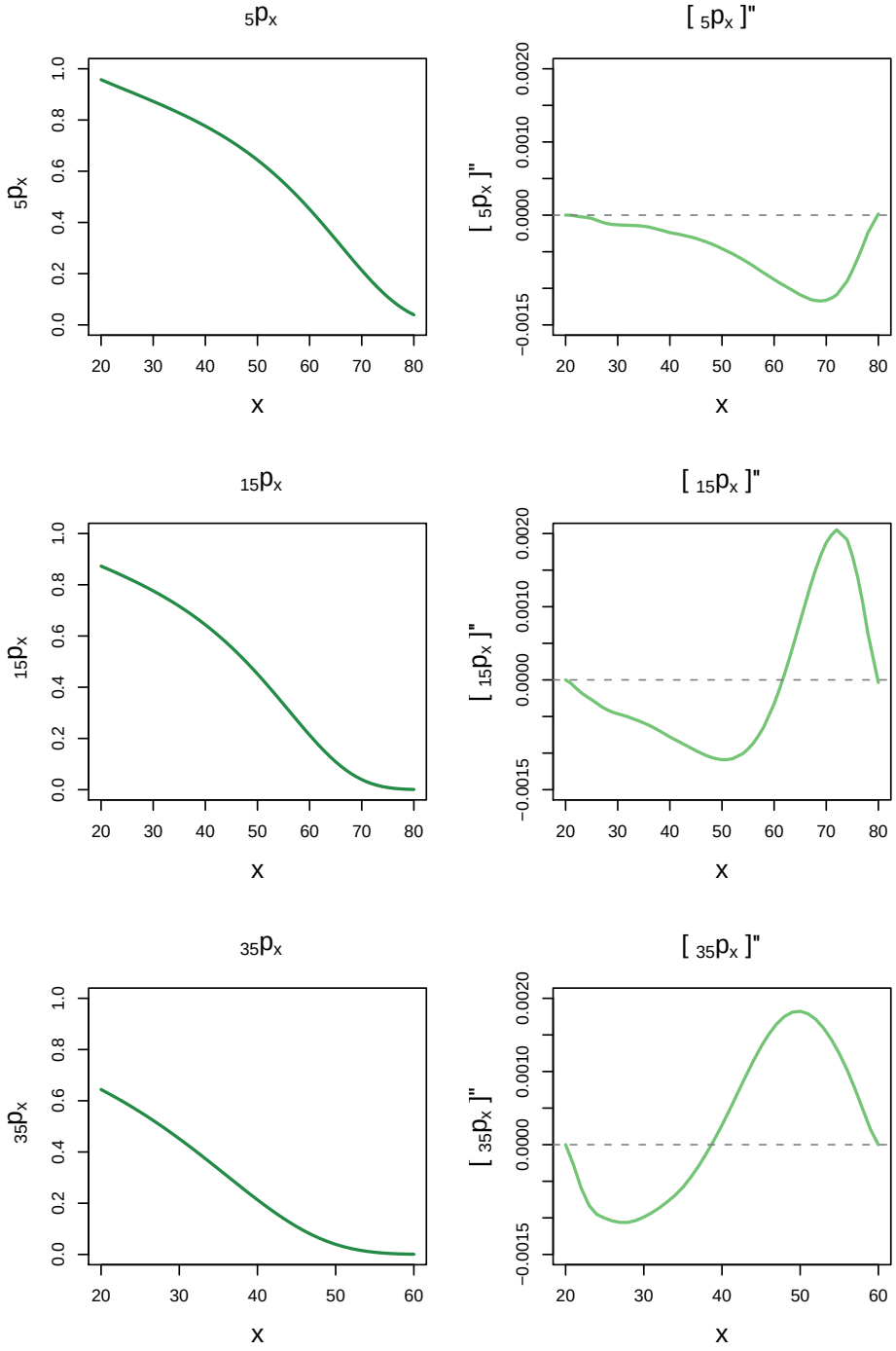

FIGURE 3.2 – *Illustration de l'évolution de* $[{}_a p_M]''$ *pour trois valeurs différentes de a (5, 15 et 35) - A partir du standard général de Brass et d'une valeur de M fixée à 25 ans.*

2. Le second élément qui détermine la qualité de l'approximation de $S(a)$ par ${}_ap_M$ est σ^2, la variance des âges à la reproduction. L'équation 3.7 indique que la *dispersion des schémas de fécondité* rend la relation entre la proportion d'orphelins et la probabilité de survie moins directe ; en d'autres termes, là où la dispersion des schémas de fécondité entre les âges est élevée, l'approximation de $S(a)$ par ${}_ap_M$ sera moins bonne.

Nous reviendrons sur ces différents éléments par la suite, notamment lorsque nous aborderons le cas de la mortalité des pères (dont les âges de reproduction sont davantage dispersés) ou le cas des schémas de mortalité perturbés par le VIH-sida (où la courbe de survie est sensiblement différente).

3.2 Convertir les proportions de survivants en probabilités de survie

La section précédente a montré que la proportion de non-orphelins par âge ($S(a)$) pouvait être exprimée de façon analytique, sous une forme qui fait intervenir des schémas de mortalité, des schémas de fécondité, et le taux de croissance intrinsèque de la population stable. Par ailleurs, nous avons rappelé que les valeurs de $S(a)$ constituent de bonnes approximations de la probabilité de survie ${}_ap_M$. Ces deux constats sont à la base de la méthode des orphelins.

Cette dernière procède par simulation (au niveau analytique) pour générer de nombreuses proportions et probabilités, et ce, pour chacune des classes d'âge des enfants (typiquement de 5 à 45 ans). Les proportions et probabilités sont ensuite mises en vis-à-vis dans une équation du type :

$$ {}_np_M = a + b\ M + c\ S(n) $$

Dans cette équation, M renvoie toujours à l'âge moyen à la naissance des enfants et n renvoie à l'âge d'entrée de la classe d'âge.

Les classes d'âge déterminent la part *fixe* de l'exposition, et les différentes valeurs de l'âge moyen à la naissance des enfants contrôlent la *variabilité induite par les différents schémas de fécondité*. Les coefficients a, b et c permettront par la suite d'approximer les probabilités grâce aux proportions, quelles que soit les valeurs paramétriques de base. Toutefois, il n'est pas aisé d'obtenir des valeurs pour estimer ${}_np_M$, c'est à dire des probabilités de survie qui débutent à l'âge moyen à la naissance des enfants. En effet, cet âge moyen va varier d'une population à l'autre. Il est donc préférable de remplacer M par 25 (une valeur moyenne de M), puis de recalculer les coefficients en conséquence. Pour les hommes, on utilisera plutôt des valeurs au-delà de 30 ans, en raison d'une fécondité plus tardive (*cf.* p. 129).

Plutôt que de recourir à un *modèle de régression*, un *système de pondération* peut alternativement être utilisé. C'est d'ailleurs cette option qui a d'abord prévalu, avec le travail de Brass et Hill (1973). Ces auteurs font intervenir deux proportions adjacentes de non-orphelins :

$$ {}_np_{25} = W \times S(n) + (1 - W) \times S(n - 5) $$

Dans cette équation, les valeurs de W dépendent de M. Ce système de pondération à deux proportions est en fait identique à un système de régression avec deux proportions, à ceci près qu'il ne comprend pas de constante[7].

Les facteurs de pondération de Brass et Hill (1973) sont les premiers disponibles pour convertir les proportions de parents survivants en indices de mortalité. La "méthode des orphelins" a par la suite été retravaillée par différents auteurs, et notamment par Ian Timaeus (Timaeus 1986; 1991b; 1992, Timaeus et Nunn 1997). Plutôt que d'opter pour une démarche chronologique, nous centrons la section suivante sur le travail de Timaeus (1992), car celui-ci a proposé une série de coefficients pour la mortalité féminine et pour la mortalité masculine. Les options alternatives seront brièvement présentées par après.

3.2.1 Coefficients introduits par Timaeus (1992)

Nous avons vu que la proportion de mères survivantes observée à partir des enfants âgés de x à $x + n$ pouvait s'exprimer sous la forme suivante :

$$ {}_5S_x = \frac{\int_x^{x+5} e^{-r(t)} \times {}_tp_0 \int_\alpha^\beta e^{-r(a)} \times {}_{a+t}p_0 \times f(a) da \; dt}{\int_x^{x+5} e^{-r(t)} \times {}_tp_0 \int_\alpha^\beta e^{-r(a)} \times {}_ap_0 \times f(a) da \; dt} $$

Cette expression peut être évaluée numériquement en rapportant, pour chaque groupe d'âge (c'est-à-dire pour les différentes valeurs de x), la somme des numérateurs et dénominateurs obtenus à partir des différentes valeurs de a et t[8]. Ce calcul procure des valeurs de ${}_5S_x$, qui peuvent ensuite être régressées sur les valeurs de ${}_np_{25}$ correspondantes au schéma de mortalité introduit, tout en contrôlant l'âge moyen à la naissance des enfants. Une option - parmi d'autres - consiste à calculer des probabilités de survie de 25 ans à $25 + n$ à partir de données déclarées par les enquêtés de $n - 5$ à n années. Ceci procure l'équation retenue par Timaeus (1992) :

$$ {}_np_{25} = \beta_0 + \beta_1 \times M + \beta_2 \times {}_5\, S_{n-5} \tag{3.8} $$

Les coefficients β_0, β_1 et β_2 sont estimés à partir de 96 simulations différentes. Pour modéliser la mortalité, Timaeus (1992) a recours au système des logits et au standard général de Brass (Encadré 6). Il retient 4

7. C'est Timaeus (1992) qui en fait la démonstration. Il indique que le système de pondération peut être décomposé en deux étapes : (1) un facteur multiplicatif f_n est calculé pour l'âge n puis (2) une moyenne est prise sur deux proportions adjacentes. Dans ce cas, on peut écrire :

$$ {}_np_{25} = W\ S(n) + (1 - W)\ S(n - 5) $$
$$ {}_np_{25} = w\ f_n S(n) + (1 - w)\ f_n\ S(n - 5) $$

Si, en plus du facteur multiplicatif, un terme d'interception a_n est ajouté, on obtient :

$$ {}_np_x = w\ (a_n + f_n)\ S(n) + (1 - w)\ (a_n + f_n)\ S(n - 5) $$
$$ {}_np_x = a_n + w\ f_n\ S(n) + (1 - w)\ f_n\ S(n - 5) $$
$$ {}_np_x = \beta_0(n) + \beta_1(n)S(n) + \beta_2(n)S(n - 5) $$

8. La fonction de croissance et de survie en a est calculée à mi-année.

valeurs pour le paramètre α et 2 valeurs de β, soit 8 schémas de mortalité distincts. Les niveaux d'espérances de vie à la naissance correspondant varient entre 36 et 73 ans, avec une moyenne à 55 ans.

Encadré 6 Le modèle des logits de Brass

Le modèle des logits de Brass a été développé pour construire des tables de mortalité à partir (1) de quelques probabilités de survie observées et (2) d'une table de mortalité qui peut servir de standard (Brass 1971). Il peut également être utilisé pour corriger certaines tables de mortalité empiriques entachées d'erreurs.

Il s'agit d'un système relationnel, au sens où il est basé sur la relation structurelle qui existe entre les courbes de survie de deux tables de mortalité. En d'autres termes, il est possible d'utiliser une transformation pour passer des probabilités observées (de la naissance à l'âge x) aux probabilités de la table standard. Cette transformation consiste à exprimer les probabilités de survie observées et celles du standard sous la forme logit. Pratiquement, si la probabilité ${}_xp_0$ est une probabilité observée et ${}_xp_{0(S)}$ est la probabilité issue de la table standard, il est possible de passer des unes aux autres grâce à l'équation suivante :

$$0.5 \times ln({}_xp_0/{}_xq_0) = \alpha + \beta \times \left[0.5 \times ln({}_xp_{0(S)}/{}_xq_{0(S)})\right]$$

Cette équation contient deux paramètres, α et β, qu'il s'agit d'estimer à partir des probabilités dont on dispose pour reconstituer toute la table. Ces deux paramètres ont un sens : le premier, α, rend compte du niveau de mortalité, tandis que le second, β, rend compte de la relation entre la mortalité adulte et la mortalité infanto-juvénile. A titre d'illustration, nous avons fait varier ces paramètres à partir d'une table standard développée par Brass (le "standard général") pour produire les graphes 3.3. Ces graphes présentent des courbes de survie par âge.

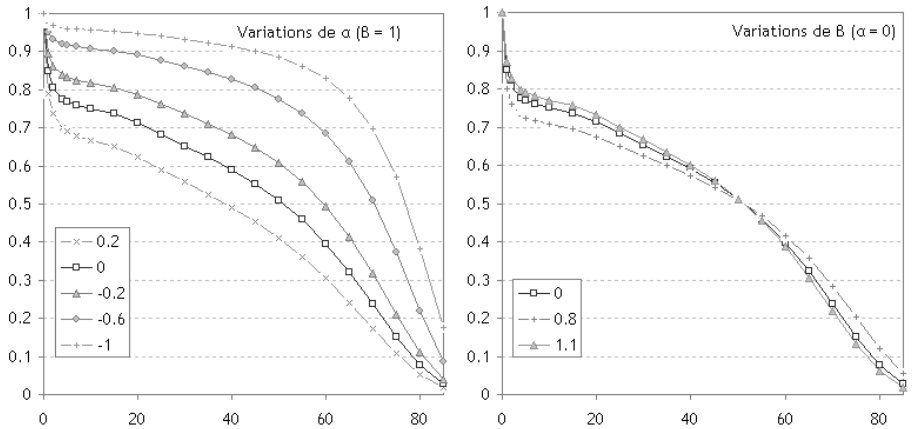

FIGURE 3.3 – *Illustration du rôle des paramètres α et β du modèle logit*

Pour un paramètre β donné, plus la valeur de α est élevée, plus la mortalité est élevée. Pour une valeur de α donné, plus le paramètre β est élevé, plus la mortalité *adulte* (ou plus précisément la mortalité au dessus de l'âge auquel la moitié de la cohorte est décédée) est élevée au regard de la mortalité *infanto-juvénile*. L'intérêt du système des logits est qu'il suffit de faire varier les paramètres α et β pour générer une série de tables de mortalité à partir d'une table standard.

Pour les femmes, les schémas de fécondité retenus par Timaeus (1992) sont basés sur le modèle relationnel de Gompertz (Encadré 7).

Encadré 7 Le modèle relationnel de Gompertz étendu à la fécondité

Ce modèle est également relationnel, parce qu'il relie la fécondité à un standard de fécondité plutôt qu'il ne la modélise selon l'âge (Booth 1984, Paget et Timaeus 1994, Brass 1974). Plus précisément, il transforme la fécondité cumulée $F(x)$ et la relie à une transformation de la fécondité cumulée standard $F_s(x)$. Tout comme pour le modèle logit (encadré précédent), une transformation du phénomène observé est donc reliée linéairement à une transformation du standard. La fécondité cumulée $F(x)$ est transformée en $Y(x)$ par le biais de l'équation suivante (dans laquelle intervient la fécondité totale notée TF) :

$$Y(x) = -ln(-ln\ F(x)/TF)$$

Le standard de fécondité cumulée $F_s(x)$ est lui aussi transformé de sorte que :

$$Y_s(x) = -ln(-ln\ F_s(x)/TF)$$

A partir de là, la fécondité observée peut être reliée linéairement au standard par le biais de l'équation suivante :

$$Y(x) = \alpha + \beta \times Y_s(x)$$

Comme précédemment, cette équation permet donc de générer différentes séries de schémas de fécondité en faisant varier les paramètres α et β, après avoir retenu un schéma de fécondité standard ($Y_s(x)$).

Le paramètre α rend compte de la localisation de la fécondité sur l'intervalle d'âge, c'est-à-dire de la plus ou moins grande précocité de la fécondité. Dans les populations réelles, il varie généralement de -0.5 à 0.5. Une valeur de -0.5 correspond au schéma de fécondité le plus âgé, tandis qu'une valeur de 0.5 correspond à une fécondité très précoce (Paget et Timaeus 1994).

Le paramètre β rend compte de la dispersion de la fécondité, c'est-à-dire de son caractère concentré sur un intervalle d'âge restreint (avec des intervalles intergénésiques assez courts) ou à l'inverse étendu sur toute la période d'exposition au risque de grossesse. Ce paramètre β varie dans les populations réelles de 0.65 à 1.5 environ (Paget et Timaeus 1994). Une valeur de 0.65 correspond à un schéma de fécondité très dispersé, tandis qu'une valeur de 1.5 correspond à l'inverse à une concentration élevée des naissances sur un bref intervalle de temps.

Comme pour le système des logits de Brass relatif à la mortalité, il faut également retenir un standard pour la fécondité. Celui retenu par Timaeus (1992) pour générer ses simulations est celui développé par Booth (1984). Il s'agit d'un standard construit sur base de 33 schémas de fécondité générés pour les populations à fécondité élevée (ISF > 5) [9]. Il concerne la fécondité féminine.

Il faut ensuite sélectionner des valeurs de α et de β. Timaeus (1992) retient quatre valeurs de α [10], et trois valeurs de β [11], ce qui couvre une large gamme de situations possibles. Cela procure donc 12 schémas de fécondité différents.

9. Ce standard est lui-même construit à partir du modèle de Coale et Trussell (1974).
10. -0.5, -0.2, 0.1, et 0.4
11. 0.7, 1.0 et 1.3

Ces schémas sont illustrés dans le graphe 3.4. En parcourant le graphe de haut en bas, on passe des fécondités les plus tardives aux fécondités les plus précoces. En allant de gauche à droite, on passe des fécondités les plus étendues sur la période d'exposition aux fécondités les plus concentrées.

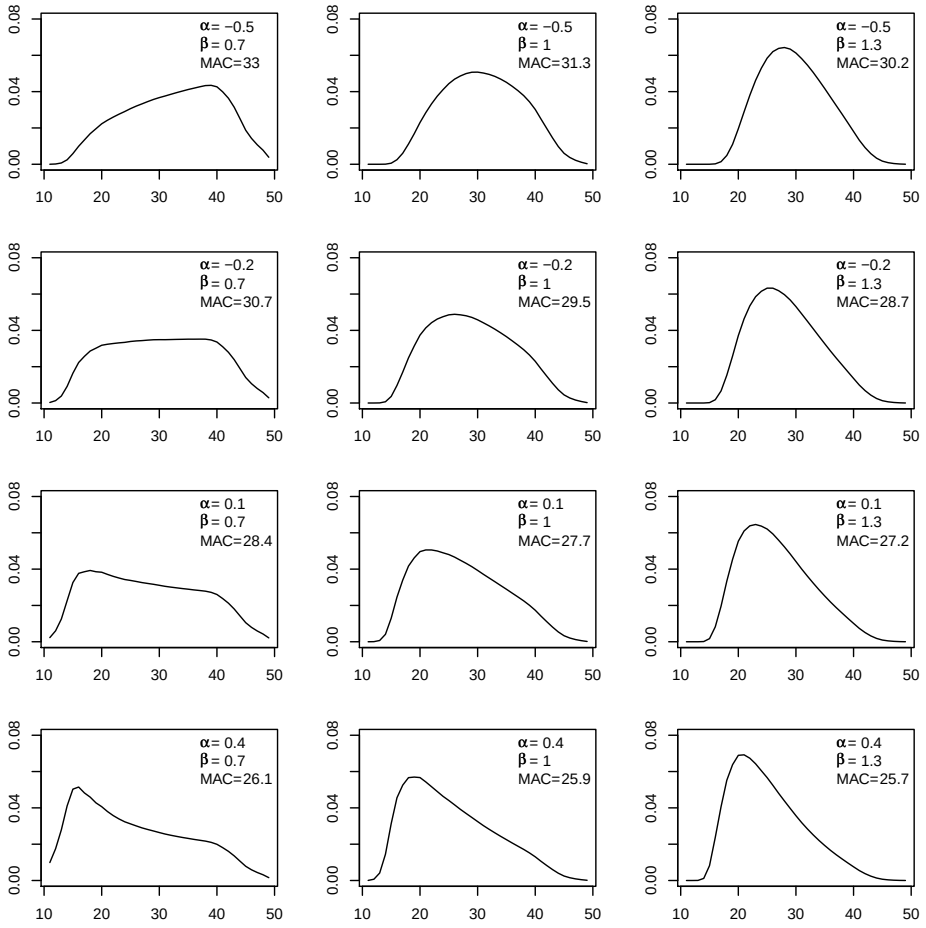

FIGURE 3.4 – *Illustration du rôle des paramètres α et β, calculs personnels à partir du standard publié dans Booth (1984) et des paramètres α et β publiés dans Timaeus (1992).* NOTE *: L'axe des abscisses renvoie à l'âge des femmes tandis que l'axe des ordonnées correspond à des taux de fécondité relatifs (leur somme vaut 1).*

Les valeurs de l'âge moyen à la naissance des enfants qui correspondent à ces paramètres varient entre 23.5 et 31.5 ans. Les valeurs de l'ISF varient quant à elles de 3.5 à 7.4 enfants par femme [12].

Le taux de croissance est fixé à 2%. Timaeus (1992) a montré que la méthode s'avère en fait très robuste aux déviations par rapport à la structure par âge de la population utilisée pour générer les coefficients.

Pour résumer, les différents paramètres pour la fécondité et la mortalité, ainsi que leurs valeurs, sont détaillés dans la table 3.1. Pour la mortalité féminine, leur combinaison procure 96 scenarios. Nous revien-

12. Comme les taux de mortalité par âge sont définis et que le taux de croissance est fixe, les taux bruts de natalité correspondants peuvent être estimés, ainsi que la structure par âge de la population stable correspondante, et donc les valeurs de l'âge moyen à la naissance des enfants.

drons plus loin sur le cas de la mortalité masculine.

Composante	Paramètres	Valeurs
Mortalité	α	0.2, -0.2, -0.6, -1.0
	β	0.8, 1.1
Fécondité (féminine)	α	-0.5, -0.2, 0.1, 0.4
	β	0.7, 1.0, 1.3
Structure par âge	r	2%

TABLE 3.1 – *Paramètres utilisés par Timaeus (1992) pour générer une série de 96 simulations*

A partir de cette série de simulations, Timaeus (1992) fournit les coefficients repris dans le tableau 3.2, qui reprend également les valeurs du R^2 des différentes régressions ainsi que le coefficient de variation [13]. On constate que l'ajustement est globalement très bon, avec des valeurs de R^2 très élevées, mais les erreurs relatives augmentent rapidement avec l'âge des enquêtés, atteignant 4% pour les enquêtés de 45-49 ans ($n = 50$). Nous avons relevé le fait que la relation entre la probabilité de survie ${}_ap_M$ et la proportion $S(a)$ dépendait de la variance des âges de reproduction ainsi que de la seconde dérivée de la fonction ${}_ap_M$. Or cette seconde dérivée augmente avec l'âge, ce qui explique que les régressions menées aux âges élevés ajustent un peu moins bien les données que celles menées aux jeunes âges.

n	β_0	β_1	β_2	R^2	CV
10	-0.2894	0.00125	1.2559	0.997	0.0015
15	-0.1718	0.00222	1.1123	0.996	0.0031
20	-0.1513	0.00372	1.0525	0.995	0.0058
25	-0.1808	0.00586	1.0267	0.993	0.0088
30	-0.2511	0.00885	1.0219	0.992	0.0126
35	-0.3644	0.01287	1.0380	0.992	0.0172
40	-0.5181	0.01795	1.0753	0.992	0.0222
45	-0.6880	0.02342	1.1276	0.993	0.0271
50	-0.8054	0.02721	1.1678	0.992	0.0400

TABLE 3.2 – *Coefficients utilisés pour convertir les proportions de mères survivantes en probabilités de survie - Timaeus (1992)*

Nous avons procédé aux mêmes simulations que Timaeus (1992) pour calculer les coefficients correcteurs. Notre objectif était de voir si nous pouvions retomber sur les mêmes coefficients à partir des mêmes valeurs paramétriques [14], avant d'étendre l'analyse aux microsimulations (dans la section 3.2.5, p. 133). Nous parvenons à des coefficients quasiment identiques (tableau 3.3) - les légers écarts sont vraisemblablement dus à des différences d'arrondis et à une approximation différente de la double intégrale. Nous repartons de nos simulations pour examiner certains aspects de la relation entre les probabilités de survie ${}_np_{25}$ et les proportions de non-orphelins ${}_5S_{n-5}$.

13. CV = RMSE/moyenne et RMSE = $\sqrt{E((\hat{\theta} - \theta)^2)}$

14. En utilisant la fonction `integrate` de R pour procéder à l'intégration numérique.

n	β_0	β_1	β_2	R^2	CV
10	-0.2996	0.00125	1.2662	0.996	0.0022
15	-0.1788	0.00227	1.1182	0.996	0.0036
20	-0.1585	0.00379	1.0582	0.995	0.0057
25	-0.1899	0.00598	1.033	0.994	0.0087
30	-0.259	0.00898	1.0267	0.993	0.0124
35	-0.3716	0.013	1.0418	0.992	0.0166
40	-0.5235	0.01805	1.0774	0.993	0.0212
45	-0.6907	0.02347	1.1285	0.994	0.0262
50	-0.8075	0.02725	1.1681	0.992	0.0394

TABLE 3.3 – *Coefficients utilisés pour convertir les proportions de mères survivantes en probabilités de survie - nos calculs à partir des paramètres retenus par Timaeus (1992)*

Le graphe 3.5 oppose tout d'abord les valeurs observées des proportions de non-orphelins de 25 à 29 ans (${}_5S_{25}$) (en abscisses) et les probabilités de survie correspondantes (${}_{30}p_{25}$) en ordonnées, selon les valeurs de l'âge moyen à la naissance des enfants. Pour les huit niveaux de mortalité, plus l'âge moyen M est élevé, plus la proportion d'individus de 25-29 ans dont la mère est survivante (${}_5S_{25}$) est faible ; les parents ont été exposés à la même mortalité mais à des âges plus avancés quand M est plus élevé.

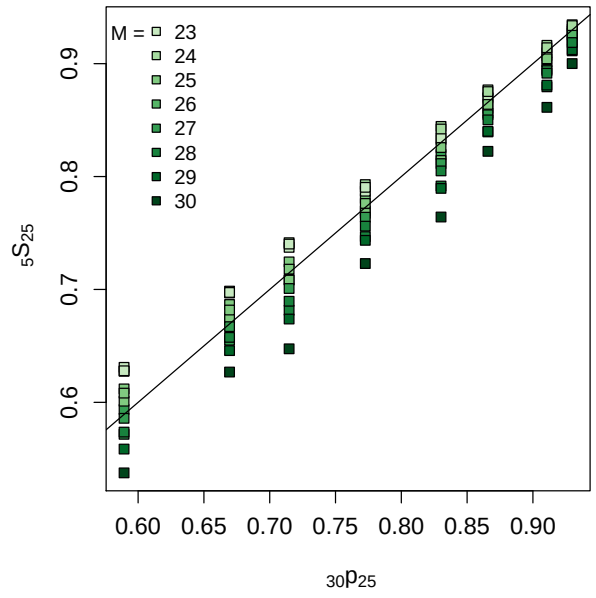

FIGURE 3.5 – *Valeurs de ${}_5S_{25}$ et ${}_{30}p_{25}$ pour différentes valeurs de M*

Nous avons vu que la probabilité ${}_ap_M$ procurait une bonne approximation de $S(a)$, et que l'approximation était encore meilleure avec la formule ${}_ap_M + \frac{\sigma^2}{2}\ [{}_ap_M]''$. A partir des simulations, nous pouvons juger de la qualité de ces deux approximations, en calculant, pour chaque série de simulations, les valeurs de ${}_ap_M$ et les valeurs de ${}_ap_M + \frac{\sigma^2}{2}\ [{}_ap_M]''$. La figure 3.6 présente les écarts (sous forme de rapports) entre ces approximations calculées à partir des simulations et les valeurs correspondantes de la proportion $S(a)$. Deux valeurs de a sont présentées ; 10 et 35, et les

écarts sont présentés selon les valeurs prises par les différents paramètres de base. Les quatre graphes (A, B, C et D) comparent les écarts selon les paramètres de la fécondité α et β. Les quatre graphes du bas (E, F, G et H) présentent les mêmes écarts, mais selon les valeurs prises par les paramètres de mortalité α et β.

Quels constats peut-on en tirer ?

On note tout d'abord que l'approximation ${}_ap_M$ surestime la proportion $S(a)$ dans tous les cas, ce qui reflète le fait que la courbure de la fonction ${}_ap_x$ est orientée vers le bas. De 0.1% de moyenne quand a vaut 10, cette surestimation s'élève à 4% en moyenne quand a vaut 35. L'approximation par ${}_ap_M + \frac{\sigma^2}{2}\ [{}_ap_M]''$ est par contre nettement meilleure. On observe cette fois une *sous-estimation*, qui se chiffre en moyenne à moins de 0.03% lorsque a vaut 5, et 0.4% de moyenne quand a vaut 35 [15].

En outre, ce qui ressort clairement de cette série de graphes, c'est que les erreurs sont positivement associées à la dispersion de la fécondité (le paramètre β de la fécondité) - ce qui est attendu. Elles sont également positivement associées au niveau de la mortalité (pour de petites valeurs de a) et positivement associées au paramètre β du système logit. Pour un même niveau de mortalité, plus ce paramètre β est élevé, plus la fonction ${}_ap_x$ est incurvée et plus sa dérivée seconde est négative.

En conclusion, on relèvera le fait que la proportion de non-orphelins par âge permet d'approcher avec une précision satisfaisante les probabilités de survie correspondantes, et que la qualité de cette approximation diminue surtout avec (1) l'âge des enquêtés, (2) la dispersion des schémas de fécondité, et (3) la courbure de la fonction de survie.

3.2.2 Une application pratique

A titre d'application pratique de la méthode des orphelins, nous utilisons les coefficients de Timaeus (1992) pour convertir en probabilités de survie les proportions de non-orphelins observées dans huit recensements est-africains. Il s'agit des données relatives à la survie des mères collectées dans les recensements menés au Kenya en 1989 et 1999, en Tanzanie en 1988 et 2002, en Ouganda en 1991 et 2002, et au Rwanda en 1991 et 2002.

Les proportions de non-orphelins proviennent d'un échantillon de données individuelles de chacun de ces huit recensements, accessibles via le projet IPUMS [16]. Les échantillons couvrent 10% de la population énumérée (sauf au Kenya, 5% de la population) [17]. Le tableau 3.4 présente

15. Cette légère sous-estimation est vraisemblablement liée au fait que nous avons négligé les termes supérieurs à 3 de la série de Taylor (*cf.* équation 3.6, page 108).

16. Minnesota Population Center. Integrated Public Use Microdata Series, International : Version 6.0. Minneapolis : University of Minnesota, 2010.
https ://international.ipums.org/international/.

17. L'âge moyen à la naissances des enfants est calculé soit à partir d'une question sur les naissances survenues dans les douze derniers mois (Rwanda et Tanzanie), soit à partir de la date de la dernière naissance (au Kenya et en Ouganda). Nous ne calculons les proportions de non-orphelins que sur les réponses non manquantes. Cette question des données manquantes est discutée plus loin (chap. 6).

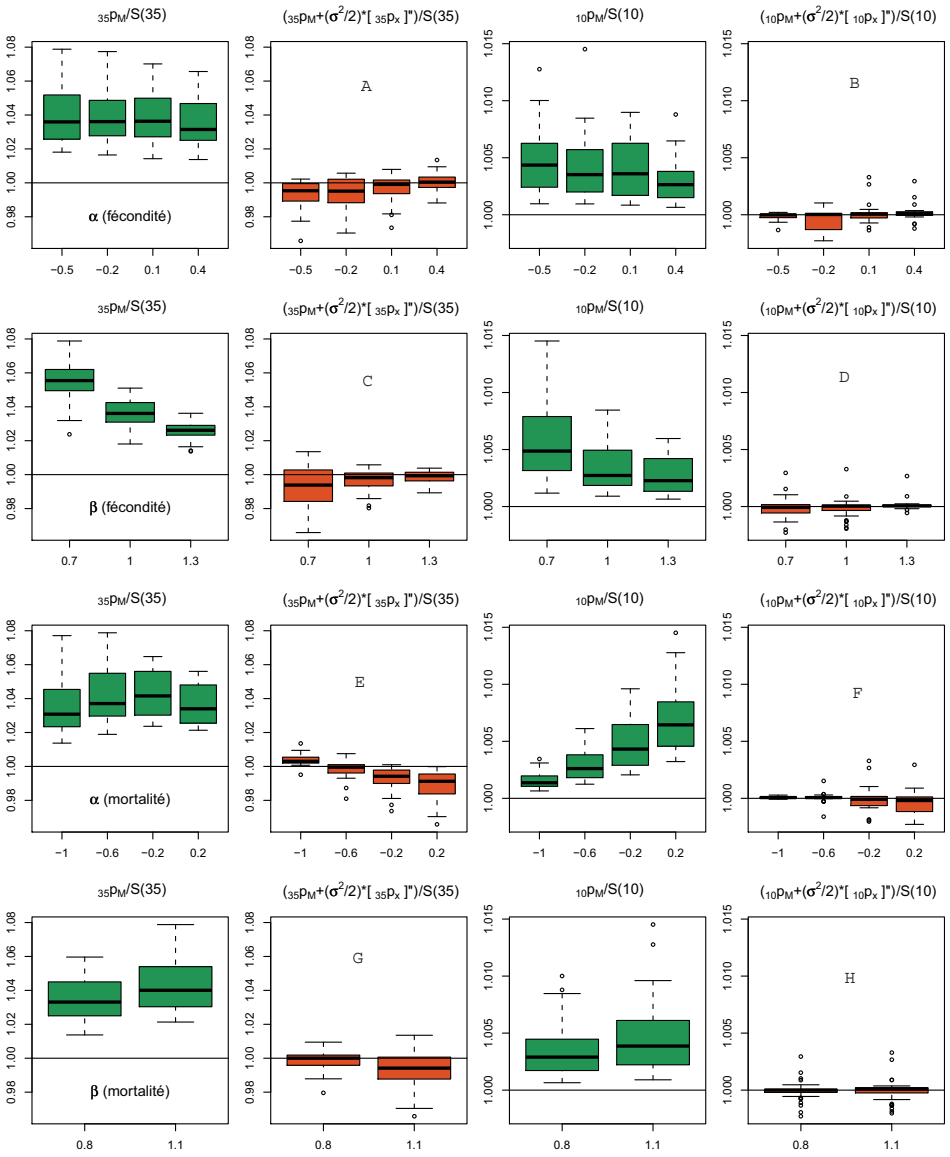

FIGURE 3.6 – *Illustration des approximations de S(a) par* ${}_ap_M$ *et* ${}_ap_M + \frac{\sigma^2}{2}\ [{}_ap_M]''$, *pour des valeurs de a valant 10 et 35*

les calculs pour les recensements ougandais. Les tableaux relatifs aux six autres recensements sont repris en annexe A.3.1 (p.426).

	Année	Age	${}_5S_{n-5}$	M	${}_np_{25}$	${}_{45}q_{15}$
10	1991	5-9	0.966	25.4	0.956	0.280
15	1991	10-14	0.942	25.4	0.932	0.275
20	1991	15-19	0.912	25.4	0.903	0.273
25	1991	20-24	0.876	25.4	0.868	0.271
30	1991	25-29	0.831	25.4	0.823	0.266
35	1991	30-34	0.763	25.4	0.754	0.270
40	1991	35-39	0.665	25.4	0.654	0.280
45	1991	40-44	0.526	25.4	0.501	0.306
50	1991	45-49	0.408	25.4	0.363	0.293
10	2002	5-9	0.959	25.1	0.947	0.324
15	2002	10-14	0.914	25.1	0.901	0.369
20	2002	15-19	0.873	25.1	0.861	0.367
25	2002	20-24	0.825	25.1	0.813	0.363
30	2002	25-29	0.788	25.1	0.776	0.329
35	2002	30-34	0.743	25.1	0.730	0.297
40	2002	35-39	0.678	25.1	0.661	0.274
45	2002	40-44	0.588	25.1	0.562	0.255
50	2002	45-49	0.477	25.1	0.435	0.234

TABLE 3.4 – *Application de la méthode des orphelins aux recensements ougandais de 1991 et 2002 (mortalité féminine)*

Comme telles, les différentes estimations indirectes ne sont bien sûr pas comparables entre elles, car elles se réfèrent à des intervalles d'âges différents pour chaque valeur de n. La cohérence des estimations est donc difficilement évaluable, et la reconstitution des tendances est impossible à partir d'une seule enquête.

Une première étape consiste donc à convertir les différentes valeurs de ${}_np_{25}$ en indices généraux de mortalité, du type ${}_{35}p_{15}$ ou ${}_{45}p_{15}$. Il s'agit d'une opération pour laquelle il existe différentes possibilités mais aucune solution pleinement satisfaisante. Généralement, les différentes probabilités de survies sont converties en une probabilité de décès au moyen du système des logits de Brass, ou par interpolation au sein des tables de Princeton [18].

Le tableau 3.5 présente différentes estimations de la probabilité ${}_{45}q_{15}$, à partir des valeurs des ${}_np_{25}$ calculées au départ du recensement ougandais de 2002. Les valeurs de ${}_{45}q_{15}$ sont obtenues successivement à partir du standard général de Brass, du schéma Ouest des tables de Princeton, puis de deux schémas par âge construits à partir des données d'observatoires de population africains (INDEPTH 2004).

On observe des niveaux de mortalité globalement similaires lorsque des schémas "standards" de mortalité sont utilisés pour générer les proba-

18. Brass et Hill (1973) suggèrent d'utiliser les logits pour aller plus loin, afin d'associer les estimations dégagées de la survie des parents avec d'autres estimations qu'il serait possible d'obtenir pour la mortalité infanto-juvénile. Dans le même esprit, Page et Wunsch (1976) proposent de recourir au modèle relationnel de Ledermann.

	Age	${}_np_{25}$	${}_{45}p_{15}$ Brass	CD Ouest	INDEPTH 1	INDEPTH 2
10	5-9	0.947	0.324	0.302	0.303	0.205
15	10-14	0.901	0.369	0.338	0.361	0.243
20	15-19	0.861	0.367	0.341	0.363	0.260
25	20-24	0.813	0.363	0.346	0.365	0.279
30	25-29	0.776	0.329	0.322	0.328	0.284
35	30-34	0.730	0.297	0.297	0.292	0.290
40	35-39	0.661	0.274	0.277	0.276	0.309
45	40-44	0.562	0.255	0.264	0.264	0.336
50	45-49	0.435	0.234	0.249	0.258	0.366

TABLE 3.5 – *Conversion des valeurs de ${}_np_{25}$ en ${}_{45}p_{15}$ - Recensement Ouganda 2002 (mortalité féminine)*

bilités ${}_{45}q_{15}$. C'est le cas des trois premières séries de ${}_{45}q_{15}$. Par contre, la conversion des valeurs de ${}_nq_{25}$ en une probabilité de décès de 15 à 60 ans procure des niveaux de mortalité très différents lorsque nous recourons au schéma perturbé par le VIH-sida (dernière colonne).

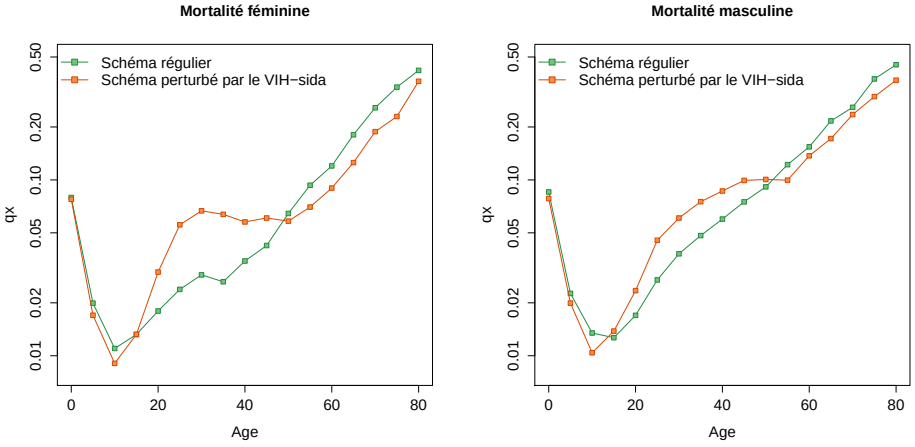

FIGURE 3.7 – *Schémas par âge de mortalité "régulier" et "perturbé par le VIH-sida" - Sc : (INDEPTH 2004)*

Ces deux schémas INDEPTH sont présentés dans le graphe 3.7. Le premier est obtenu à partir des observatoires de populations africains où la prévalence du VIH-sida reste faible. Il s'agit des observatoires situés au Sénégal, au Ghana, au Burkina, en Gambie et en Afrique du Sud [19]. Le second est construit à partir des observatoires où la prévalence du VIH-sida est plus élevée (Ethiopie, Mozambique, Zambie et Tanzanie) [20]. On y retrouve la hausse caractéristique des risques de décès entre 20 et 45-50 ans dans les contextes fortement touchés par le VIH-sida. Cette concentration des décès entre 20 et 45-50 ans implique que pour un même niveau de mortalité aux jeunes âges adultes (par exemple ${}_{10}q_{25}$), les probabilités de décès ${}_{45}q_{15}$ seront plus faibles avec le schéma perturbé par le VIH qu'avec

19. Agincourt, Mlomp, Bandafassi, Navrongo, Farafenni, Niakhar, Nouna et Oubritenga.

20. Butajira (Ethiopie), Manhica (Mozambique), Gwembe (Zambie), ainsi que les sites tanzaniens : Ifakara, Rufiji,Hai, Dar es Salam et Morogoro.

un schéma plus classique, comme le standard INDEPTH-1 ou le standard de Brass. Ceci s'observe sur le graphe 3.8, où nous comparons différentes probabilités de décès à partir du standard de Brass et du standard INDEPTH-2. Pour une valeur de la probabilité ${}_{10}q_{25}$ de 0.1, le standard de Brass nous donne un quotient ${}_{45}q_{15}$ de 0.5, contre 0.34 avec le standard INDEPTH-2 (féminin). Pour des valeurs de n avoisinant 35 ans, la relation entre les probabilités ${}_{n}q_{25}$ et ${}_{45}q_{15}$ est quasiment identique dans les deux schémas de mortalité (quotients non présentés ici). Ensuite, plus n augmente, plus les valeurs de ${}_{45}q_{15}$ seront comparativement plus élevées dans le schéma "perturbé" que dans un schéma classique (pour un même niveau de ${}_{n}q_{25}$).

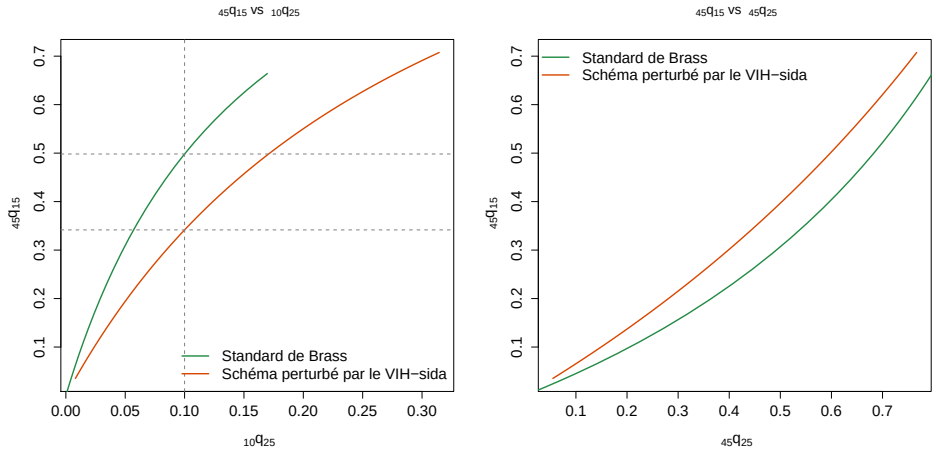

FIGURE 3.8 – *Relation entre les probabilités* ${}_{10}q_{25}$, ${}_{45}q_{15}$ *et* ${}_{45}q_{25}$ *selon le type de schéma de mortalité (mortalité féminine) - Sc : (INDEPTH 2004)*

Ceci explique les différences entre les valeurs de ${}_{45}q_{15}$ calculées à partir des trois schémas "classiques" et celles obtenues à partir du schéma "perturbé" (tableau 3.5, p.120). Les valeurs de ${}_{45}q_{15}$ valent approximativement 70% des niveaux observés chez les moins de 20 ans avec un des trois schémas classiques. A partir des déclarations faites par les 30-34 ans, le rapport s'inverse et les quotients de mortalité sont plus élevés lorsqu'ils sont convertis à partir du schéma perturbé. En définitive, *dans les contextes perturbés par le VIH-sida, le choix du standard a un impact décisif sur les niveaux de mortalité dégagées de la survie des parents*. C'est donc le cas également sur les tendances. Par exemple, si le VIH-sida induit une modification progressive des schémas de mortalité, convertir les ${}_{n}p_{25}$ au moyen de schémas par âge classiques sur toute la période considérée va conduire à une sur-estimation de la hausse de mortalité (puisque les plus jeunes enquêtés procurent les estimations les plus récentes).

Nous reviendrons sur les perturbations induites par le VIH-sida (dans le chapitre suivant), notamment parce que les coefficients de corrections devraient être revus pour tenir compte de ce schéma de mortalité atypique. A ce stade, afin de rendre les différentes valeurs de ${}_{n}p_{25}$ issues des recensements comparables, nous les avons simplement converties en valeurs de ${}_{45}q_{15}$ à partir du standard général de Brass, en interpolant linéairement entre différents niveaux de mortalité. Ces valeurs sont reprises dans la dernière colonne du tableau 3.4 (p.119) et des tableaux repris en

annexe A.3.1 (p.426).

Même traduites en un niveau unique de mortalité (${}_{45}q_{15}$), ces différentes estimations est-africaines sont difficilement comparables, car elles font référence à des périodes de référence variables. Il faut donc procéder à une datation des différentes estimations, une question discutée dans la section 3.3.2. A ce stade, nous pouvons déjà relever :

1. une légère baisse de l'âge moyen à la première naissance dans 3 pays sur les quatre (tandis que l'âge moyen reste stable en Tanzanie),
2. une forte hétérogénéité des niveaux de mortalité obtenus : les valeurs de ${}_{45}q_{15}$ avoisinent 0.150 à partir du recensement kenyan de 1989, elles vont jusqu'à 0.360 environ à partir des plus jeunes au recensement ougandais de 2002 et atteignent même 0.480 à partir du recensement rwandais de 2002. On perçoit déjà là les fortes hausses de mortalité, dues au VIH-sida (et aux conflits dans le cas du Rwanda).
3. la hausse des risques de mortalité adulte, qui est apparente dans les quatre pays examinés ici, lorsque l'on compare les niveaux obtenus d'un recensement à l'autre.

3.2.3 Autres modèles et paramètres utilisés

Préalablement aux coefficients de Timaeus (1992), il existait de nombreuses séries de coefficients qui permettent de convertir les proportions de parents survivants en probabilités de survie. Dans cette section, nous parcourons brièvement les différentes approches disponibles. L'objectif est surtout de présenter leurs avantages et désavantages respectifs, et d'insister sur le fait qu'en dépit d'hypothèses différentes sur les schémas de fécondité et de mortalité, ces différentes approches procurent globalement des résultats comparables.

Brass et Hill (1973)

Brass et Hill ont proposé la première série de valeurs dès 1973. Comme nous l'avons déjà indiqué, elle prend la forme de facteurs de pondération, qui s'appliquent aux proportions déclarées par deux groupes quinquennaux contigus. Pour les femmes, les probabilités de survie estimées partent de 25 ans et l'équation s'écrit sous la forme suivante :

$$ {}_{n}p_{25} = W(n)\,{}_{5}S_{n-5} + (1 - W(n))\,{}_{5}S_{n} $$

Les valeurs de $W(n)$ sont reprises en annexe A.3.2 (page 428) pour différentes valeurs de l'âge moyen à la naissance des enfants. Il faut interpoler entre ces valeurs pour obtenir les facteurs de pondération correspondant à la valeur de M réellement observée.

Pour calculer les probabilités de survie pour une valeur de $n = 10$, par exemple, on utilisera ${}_{10}p_{25} = W(10)\,{}_{5}S_{5} + (1 - W(10))\,{}_{5}S_{10}$. Dans ce cas, ${}_{10}p_{25}$ renvoie à la probabilité de survie de 25 ans à 35 ans, ${}_{5}S_{5}$ à la proportion de mères survivantes déclarée par les individus de 5 à 10 ans, et ${}_{5}S_{10}$ à la même proportion déclarée par les individus âgés de 10 à 15 ans.

Ainsi, durant l'enquête EDS menée en Tanzanie en 2004, 97.5% et 94% des enfants âgés respectivement de 5 à 9 et de 10 à 14 ans ont déclaré que leur mère vivait toujours. L'âge moyen à la naissance des enfants peut être estimé à 27.5 ans à partir des naissances déclarées des douze derniers mois. Après avoir interpolé les valeurs-types de $W(n)$ entre 27 et 28, il est donc possible de calculer la probabilité de survie de 25 à 35 ans comme :

$$_{10}p_{25} = 0.6549 \times 0.975 + 0.3451 \times 0.939 = 0.963$$

Cette valeur pourra ensuite être transformée en indices standards ($_{35}q_{15}$, $_{45}q_{15}$ ou e_0) par interpolation au sein d'une table type. En utilisant plutôt les coefficients de Timaeus (1992), on obtient :

$$_{10}p_{25} = -0.2894 + 0.00125 \times 27.5 + 1.2559 \times 0.975 = 0.969$$

L'avantage d'une procédure basée sur une pondération des différentes proportions contigües (plutôt que sur un modèle de régression) est que les résultats obtenus seront moins sensibles aux changements de calendriers de mortalité ou de fécondité, qui vont avoir tendance à s'annuler mutuellement (ceci est discuté par Timaeus (1992)).

Le principal désavantage des coefficients de Brass et Hill (1973) est qu'ils sont basés sur un nombre très restreint de simulations, ce qui peut poser problème, comme nous le verrons plus loin.

Le schéma de mortalité retenu pour calculer les valeurs de $W(n)$ est unique : il s'agit du standard général de Brass. Cela correspond à une mortalité relativement élevée, avec une espérance de vie à la naissance qui avoisine 43.5 ans. La structure par âge de la fécondité suit quant à elle une fonction cubique assez simple (annexe A.3.3), les différents schémas de fécondité ne variant qu'en fonction de l'âge auquel débute la période de reproduction [21]. Avec une mortalité et un taux de croissance fixes, seul cet âge introduit une variation dans les simulations. Les coefficients sont donc obtenus à partir de 7 simulations, et d'un seul schéma de mortalité [22].

Hill et Trussell (1977)

Hill et Trussell (1977) ont proposé la première série de valeurs prenant la forme de coefficients de régression. Ils font intervenir une seule proportion $_5S_{n-5}$, et contrôlent, en plus de l'âge moyen à la naissance des enfants (M), la probabilité de survie de la naissance au second anniversaire (l_2). Leur méthode génère une probabilité de survie *non conditionnelle*, de la naissance à l'âge $25 + n$. Les probabilités calculables varient de l_{45} à l_{75}. Les coefficients proposés s'appliquent *exclusivement à la survie féminine* :

21. Brass et Hill (1973) ne définissent pas directement de niveaux de fécondité, mais ils posent un taux de croissance de 2%. Ils font varier l'âge auquel débute la période de reproduction par sauts de 2.5 ans de 10 à 25 ans.

22. Comme précédemment, l'équation qui relie les proportions de non-orphelins aux probabilités de survie (équation 3.4, p.105) est approximée numériquement en sommant les valeurs obtenues sur les différents âges, à partir de n = 2.5 par intervalle de 2.5 ans. Les résultats sont transformés en proportions observées sur 5 ans ($_5S_n$) en agrégeant les valeurs calculées par saut de 2.5 ans, tout en tenant compte de la mortalité ainsi que de la croissance au sein du groupe de n à $n + 4$.

$$_{25+n}p_0 = \beta_0 + \beta_1 \times M + \beta_2 \times_5 S_{n-5} \times l_2$$

Ces coefficients sont repris dans le tableau 3.6. L'erreur relative se chiffre à 1.8 % sur la probabilité de survie de la naissance à 40 ans, et elle augmente avec l'âge, atteignant 3% sur la probabilité l_{75}.

Pour construire leur modèle, Hill et Trussell (1977) utilisent eux aussi le standard général de Brass [23] pour la mortalité et les modèles de fécondité de Coale et Trussell (1974) [24]. Ils génèrent 900 simulations différentes. Les populations sont stables et les taux de croissance sont définis par la combinaison des schémas de mortalité, de fécondité et des ISF [25]. A partir des mêmes séries de paramètres, nous obtenons des coefficients comparables (résultats non présentés ici).

n	β_0	β_1	β_2	R^2	CV
20	-0.3534	0.00553	1.1568	0.994	0.0183
25	-0.3768	0.00755	1.1360	0.998	0.0129
30	-0.4134	0.00997	1.1192	0.999	0.0101
35	-0.4620	0.01270	1.1091	0.998	0.0127
40	-0.5145	0.01541	1.1059	0.998	0.0173
45	-0.5504	0.01736	1.1037	0.998	0.0208
50	-0.5342	0.01736	1.0841	0.996	0.0305

TABLE 3.6 – *Coefficients utilisés pour convertir les proportions de mères survivantes en probabilités de survie - Hill et Trussell (1977)*

Le désavantage de cette méthode est qu'il faut disposer d'informations fiables sur la survie des enfants de moins de deux ans. Même quand ces informations sont disponibles, il devient délicat de déterminer le niveau à retenir dans les cas (qui sont les plus fréquents) où la mortalité des enfants a évolué.

Nations Unies (1983)

Dans le célèbre Manuel X, les Nations Unies (1983) fournissent des coefficients également. Ils s'intègrent dans une équation qui ne fait pas intervenir la mortalité des enfants et procurent des probabilités de survie *conditionnelles* :

$$_np_{25} = \beta_0 + \beta_1 \times M + \beta_2 \times_5 S_{n-5}$$

Les coefficients sont repris en annexe A.3.2 (p. 428). Les détails du calcul de ces coefficients ne sont pas fournis par les Nations Unies. Comme chez Hill et Trussell (1977), les coefficients ne s'appliquent qu'à des informations procurées par des enquêtés de plus de 15 ans. Dès lors, ces coefficients ne sont pas applicables à la plupart des enquêtes EDS, où seuls les

23. En retenant dix variantes ($\alpha = \{-0.4, -0.2, 0.0, 0.2, 0.4\}$ et $\beta = \{0.8, 1.1\}$), qui correspondent à des niveaux d'espérances de vie à la naissance variant de 29 à 57 ans, avec 43 ans de moyenne.

24. En retenant 45 variantes ($a_0 = \{13, 14, 15\}$, $S = \{18, 19, 20, 21, 22\}$ et $m = \{0, 0.2, 0.4\}$). Ces schémas sont combinés à deux niveaux d'ISF (5 et 7).

25. Ils peuvent être calculés à partir de la méthode itérative de Coale (1957) ; ils varient de 2.5 à 6.6% (4.6% en moyenne). Les âges moyens à la maternité, ainsi que les âges moyens à la naissance des enfants peuvent également être calculés ; les premiers varient de 28 à 31 ans, et les seconds de 24.9 à 29.6 ans.

enfants de moins de 15 ans sont interrogés sur la survie de leurs parents (ni à certains recensements).

Palloni et Heligman (1985)

Palloni et Heligman (1985) ont utilisé une équation équivalente à celle du Manuel X, mais ils fournissent des coefficients qui peuvent s'appliquer également chez les enfants (à partir des répondants de 5 à 9 ans). Leur recherche présente également la particularité d'utiliser les cinq familles de tables de mortalité des Nations Unies (1985) destinées aux pays en développement [26]. Les auteurs suggèrent d'utiliser l'une de ces cinq séries de coefficients, en fonction du contexte analysé, car ils constatent que *les estimations sont obtenues avec une plus grande précision lorsque les proportions générées par un schéma de mortalité donné (p.ex. le schéma latino-américain) sont converties en probabilités par le biais du même schéma*.

Cela étant, les erreurs introduites par le recours à une autre famille type (que celle qui sous-tend les simulations) restent faibles : elles peuvent facilement être calculées, à partir des valeurs des erreurs quadratiques fournies par Palloni et Heligman (1985). Par exemple, au sein de l'ensemble des simulations faites par les auteurs, dans l'hypothèse où la mortalité réelle se conforme au *schéma général*, l'estimation de la probabilité de survie ${}_{20}p_{25}$ (à partir des 15-19 ans) s'accompagne d'une erreur relative de l'ordre de 0.27% (cette erreur est liée au fait que les probabilités prédites sont calculées à partir des coefficients de la régression). Cette erreur relative augmente légèrement si d'autres coefficients sont utilisés ; elle va varier entre 0.32% (avec le schéma chilien) et 0.6% (avec le schéma extrême-oriental). Soit des variations négligeables au regard des erreurs induites par les problèmes de qualité des données (dont nous discuterons plus loin) ou par la violation d'autres hypothèses de base [27]. Des erreurs bien plus conséquentes seront introduites dans le cas de schémas de mortalité atypiques, tels que ceux perturbés par le VIH-sida.

Importance du choix des coefficients retenus

Après avoir présenté les différentes séries de coefficients disponibles, cette section s'intéresse à l'importance du choix entre ces derniers ; quel est l'impact de ce choix sur les estimations dégagées ?

Timaeus (1992) a mené une première analyse de la robustesse de la méthode des orphelins aux variations des paramètres utilisés. Il génère une série de proportions de *pères* survivants en modulant les valeurs de α, β et r, ainsi que les paramètres de fécondité (nous reviendrons sur l'application de la méthode aux pères). Il applique ensuite les facteurs de Brass

26. Pour chacun de ces cinq schémas de mortalité, les coefficients sont obtenus en combinant sept niveaux d'espérance de vie (40, 45,...70 ans) avec 96 schémas de fécondité issus du modèle de Coale et Trussell (1974). Les populations sont stables et leur structure est définie par la combinaison des taux de mortalité et de fécondité, ainsi que d'un taux brut de reproduction fixé à 3.

27. Ces erreurs sont toutefois légèrement supérieures pour les âges plus élevés. Ainsi, pour la probabilité ${}_{40}p_{25}$ (estimée à partir des 35-39 ans), les mêmes calculs procurent une erreur relative de 0.6% lorsque le schéma général est retenu. Ces erreurs varient entre 0.65% (pour le schéma chilien) et 1.28% (pour le schéma sud-asiatique) si d'autres coefficients sont utilisés. On reste globalement dans des marges très acceptables.

et Hill (1973), puis ses propres coefficients, sur ces proportions simulées. Il calcule à chaque fois les erreurs observées par rapport au standard de mortalité qui est à la base des simulations de départ. Outre le fait que ses propres coefficients sont associés à une plus grande précision, cette analyse lui permet de mettre en évidence les éléments suivants :

- même lorsque les schémas de mortalité ou de fécondité diffèrent des schémas qui sont à la base des coefficients, les erreurs estimées restent *faibles*.
- la méthode est très robuste aux variations de la *structure par âge* (définie par r).
- les variations de l'*étendue de la fécondité* ont par contre un impact plus important ; si la fécondité s'étend sur un intervalle *plus long* que ce qui a été prévu pour générer les coefficients, la mortalité va être *surestimée* à partir des plus jeunes répondants (5-20 ans) et *sous-estimée* à partir des répondants plus âgés (35 et plus).
- l'impact le plus important est celui des *niveaux* et *schémas* de mortalité : la mortalité sera sous-estimée si elle est plus élevée que celle qui a prévalu pour générer les coefficients. Elle sera surestimée si elle est plus faible que celle qui est à la base des simulations. Cela est problématique pour les coefficients de Brass et Hill (1973) générés à partir d'un seul schéma de mortalité.

Pour compléter cette analyse, nous avons appliqué les *différentes séries de coefficients disponibles* aux proportions de mères survivantes calculées à partir de la série de simulations proposée par Timaeus (1992).

n	10	15	20	25	30	35	40	45
Timaeus 1992	0.22	0.36	0.58	0.88	1.24	1.67	2.13	2.62
Nations Unies 1983	-	-	0.65	0.98	1.37	1.8	2.23	2.8
Brass-Hill 1973	0.22	0.36	0.57	0.88	1.4	2.13	3.73	6.22
Palloni-Heligman 1985	0.42	0.61	0.91	1.29	1.73	2.12	2.62	4.16
Hill-Trussell 1977 ($_{25+n}p_0$)	-	-	2.85	2.5	2.42	2.68	3.46	5.09

TABLE 3.7 – *Erreurs relatives (en %) sur les probabilités de survie féminines $_np_{25}$ avec différentes séries de coefficients*

Le tableau 3.7 détaille les erreurs relatives, en %, obtenues avec les différentes séries de coefficients [28]. On observe que les trois premières séries procurent des erreurs qui restent en-deça de 1% lorsque les répondants ont moins de 30 ans. Les erreurs sont par contre plus importantes pour les 20-35 ans au départ des coefficients de Hill et Trussell (1977), mais elles portent sur une probabilité de survie plus longue ($_{25+n}p_0$ plutôt que $_np_{25}$). Comme on l'a déjà indiqué, les erreurs augmentent rapidement avec l'âge. Outre les coefficients de Timaeus (1992) [29], les coefficients des Nations Unies procurent des résultats satisfaisants, ainsi que le système de pondération de Brass et Hill (1973), mais seulement jusqu'à 30 ans.

28. Les erreurs relatives sont calculées comme RMSE/moyenne, soit $\sqrt{E((\hat{\theta}-\theta)^2)}/\hat{\theta}$. Les coefficients de Palloni et Heligman (1985) sont ceux correspondant au modèle Général des Nations Unies.

29. Qui procurent logiquement les erreurs les plus petites puisqu'ils sont calculés à partir de cette série de simulations…

Afin d'y voir plus clair, les graphes 3.9 et 3.10 opposent (en abscisses) les probabilités de survie simulées et (en ordonnées) les mêmes probabilités recalculées avec différentes séries de coefficients. Ils présentent également, sous forme de boxplots, les écarts observés par âge entre les probabilités simulées et celles prédites par la méthode. Pour chaque groupe d'âge, les boxplots sont construits à partir de 96 valeurs simulées de ${}_np_{25}$.

Les coefficients de Timaeus (1992) procurent logiquement des valeurs de ${}_np_{25}$ très proches de celles qui sont à la base des simulations (figure 3.9, deux graphes du haut). Il en va de même pour les coefficients du Manuel X (mais les paramètres à la base de ces derniers nous sont inconnus).

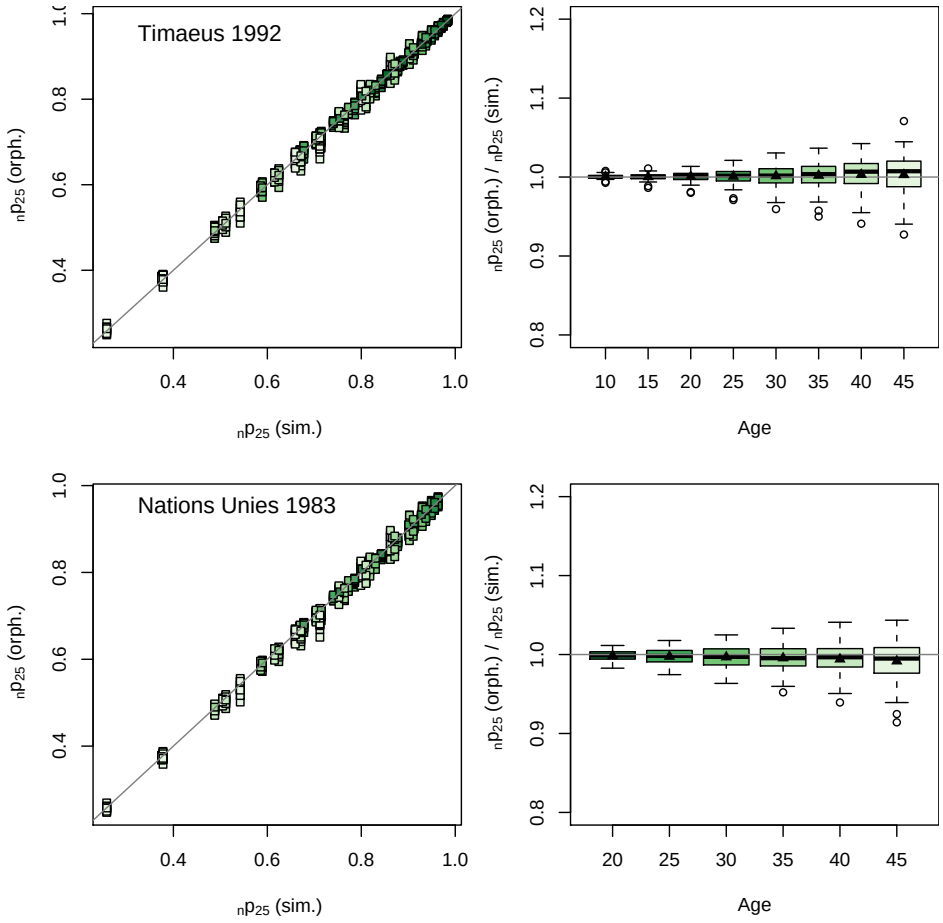

FIGURE 3.9 – *Ecarts entre les probabilités ${}_np_{25}$ simulées et recalculées, coefficients de Timaeus (1992) et Nations Unies (1983)*

Par contre, on constate que les coefficients de Palloni et Heligman (1985) et de Brass et Hill (1973) tendent à induire une *sous-estimation* de la probabilité de survie lorsque n s'élève au delà de 30 ans. Cela est particulièrement marqué pour les coefficients de Brass et Hill (1973). En fait, ces derniers sont calculés à partir d'un seul schéma dont l'espérance de vie vaut 43.5 ans. Or les simulations utilisées ici sont basées sur des niveaux d'espérances de vie variant entre 36 et 73 ans, avec une moyenne de 55 ans. La mortalité 'simulée' est donc *plus faible* que la mortalité utilisée pour calculer les coefficients, ce qui se traduit par une *sous-estimation des*

probabilités de survie. Aux jeunes âges (lorsque n est faible), les valeurs de ${}_np_{25}$ sont relativement comparables d'un niveau de mortalité à l'autre, mais elles se différencient davantage aux âges élevés. Cela explique que la sur-estimation de la mortalité est plus prononcée au-delà de 30-35 ans.

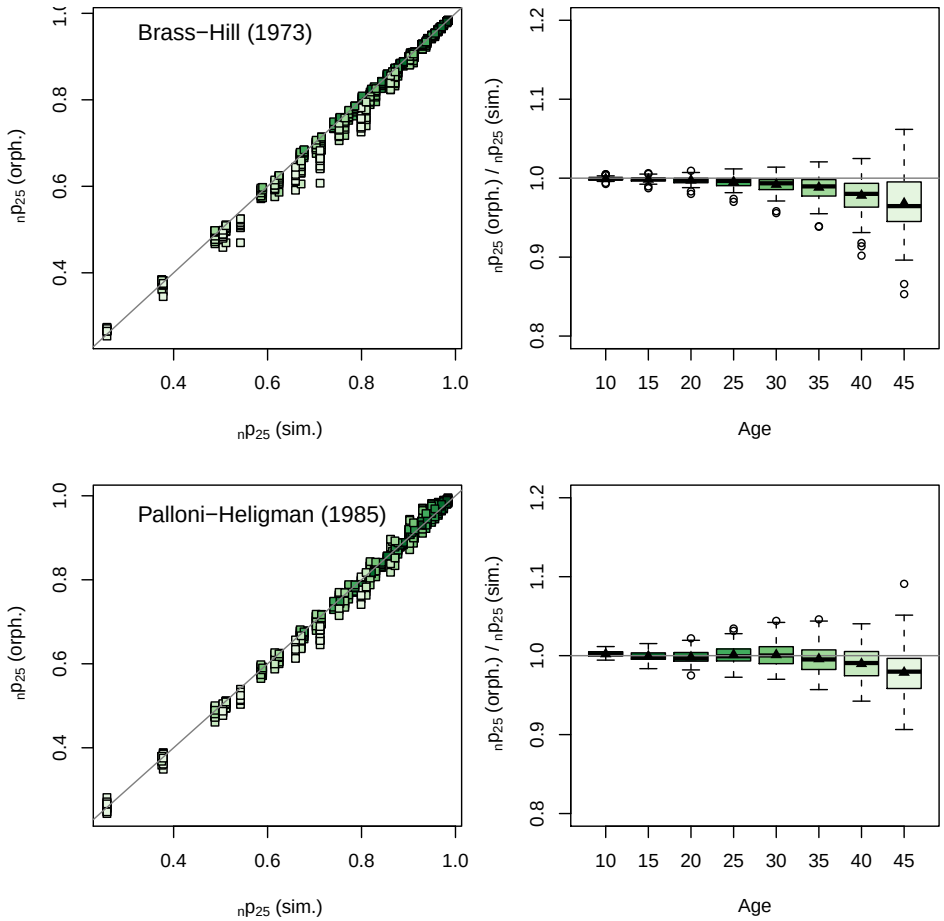

FIGURE 3.10 – *Ecarts entre les probabilités ${}_np_{25}$ simulées et recalculées, coefficients de Brass-Hill (1973) et Palloni-Heligman (1985)*

On observe un troisième type de configuration avec les coefficients de Hill et Trussell (1977) ; à tous les âges, les probabilités de survie recalculées sont en-deçà des probabilités de survie générées par simulation (graphe 3.11). A nouveau, cela peut s'expliquer partiellement par le fait que les coefficients de Hill et Trussell (1977) sont issus de séries de simulations utilisant des mortalités plus fortes : les niveaux d'espérances de vie varient entre 29 et 57 ans, avec 43 ans de moyenne.

En conclusion, on notera que pour les valeurs de n inférieures à 30-35 ans, la méthode des orphelins (quels que soient les coefficients retenus) paraît robuste aux variations des schémas de mortalité et de fécondité utilisés pour générer les coefficients. C'est moins le cas aux âges élevés, pour des raisons déjà évoquées.

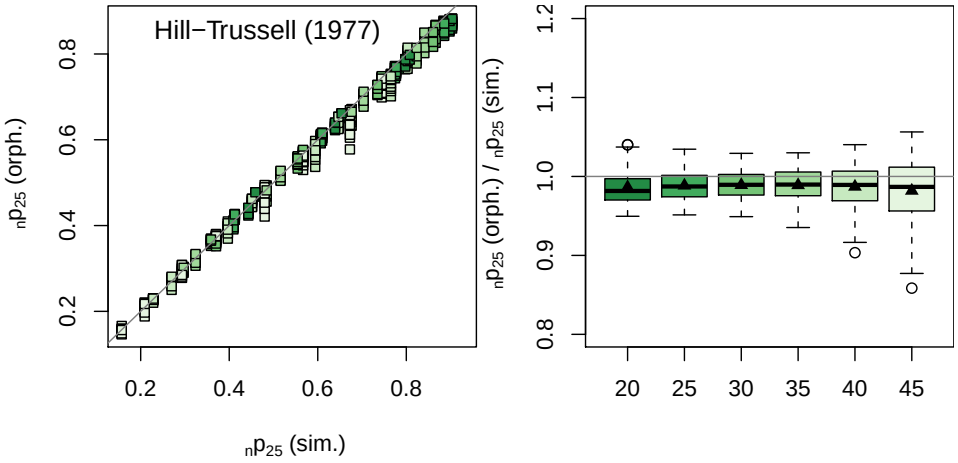

FIGURE 3.11 – *Ecarts entre les probabilités* $_{25+n}p_0$ *simulées et recalculées, coefficients de Hill-Trussell (1977)*

Parce qu'ils sont bien documentés et qu'ils peuvent permettre d'estimer la mortalité masculine, il nous paraît préférable d'utiliser les coefficients de Timaeus (1992).

3.2.4 Modèles construits pour la mortalité masculine

L'application de la méthode des orphelins à la mortalité *masculine* a fait l'objet de moins de développements. Seuls deux séries de coefficients sont disponibles : ceux fournis par Brass et Hill (1973) et ceux de Timaeus (1992).

Au moins deux éléments rendent l'application de la méthode à la survie des pères plus délicate :

- L'âge moyen à la naissance des enfants est difficile à estimer, parce que les données nécessaires à son calcul sont rarement collectées.
- L'intensité et le calendrier de la mortalité sont moins bien appréhendés. La fécondité connaît une grande variation d'un contexte à l'autre - notamment en fonction de l'importance de la polygamie. Le calendrier masculin démarre également plus tard que le calendrier féminin, et il s'étend davantage aux âges élevés, là où la mortalité est plus forte. Ceci augmente la sensitivité de la méthode aux paramètres introduits.

Estimation de l'âge moyen des pères à la naissance des enfants

La première difficulté relève de l'estimation de l'âge moyen des pères à la naissance des enfants ($M_{♂}$). Plusieurs options sont possibles pour procéder à ce calcul.

La *première approche* consiste à calculer cet âge moyen de façon directe, c'est-à-dire comme pour les mères, à partir des naissances survenues dans les douze derniers mois tabulées par l'âge des pères (en retranchant neuf mois, puisque la période d'exposition des pères débute dès la conception [30]). Cette première approche est la meilleure option, mais les

30. Raison pour laquelle les poids de (Brass et Hill 1973) ne sont pas appropriés pour

données ne sont pas toujours disponibles. Une source importante comme l'*Annuaire démographique* des Nations Unies ne fournit les tabulations nécessaires au calcul direct que pour quelques pays africains et quelques années (UN 1966; 1970; 1976; 1983a)[31]. Certaines enquêtes EDS menées auprès des hommes comprennent également des informations sur la date de naissance de leur dernier enfant. Ces données ont été collectées dans un peu plus d'une enquête sur quatre, parmi les 70 EDS africaines au cours desquels les hommes ont été interrogés.

La *seconde approche*, qui est utilisée par défaut, repose sur des approximations à partir de l'âge moyen à la naissance des enfants calculées chez les mères ($M_{♀}$). L'âge moyen des pères est estimé en ajoutant à $M_{♀}$ un *écart d'âge entre époux* .

Cet écart d'âge entre époux doit souvent lui-même être estimé. Dans ce cas, deux options sont possibles :

1. Soit cet écart d'âge est estimé à partir de la différence entre l'*âge moyen au premier mariage* des hommes et celui des femmes. Mais ce calcul ne considère que les premiers mariages. En conséquence, il *sous-estime* l'écart d'âge entre époux. En effet, les hommes ont tendance à se remarier davantage que les femmes (Blacker et Mukiza-Gapere 1988). Si l'écart d'âge entre époux est sous-estimé, l'âge moyen $M_{♂}$ l'est aussi, et la mortalité est donc surestimée.
2. Cet écart peut alternativement être calculé comme la différence entre l'*âge médian des hommes mariés* et l'*âge médian des femmes mariées*. Ceci intègre des informations sur les personnes remariées, mais aussi sur les personnes qui ne sont plus en âge de reproduction. Si, aux âges élevés, les différences d'âge entre époux sont plus prononcées qu'aux jeunes âges, ce calcul a pour effet de *surestimer* l'âge moyen des pères à la naissance de leurs enfants. Par ailleurs, notons que cette seconde option est applicable comme telle uniquement lorsque les proportions d'hommes polygames sont faibles (disons < 10%) ; là où la polygamie est plus fréquente, l'âge moyen des hommes devrait être pondéré par le nombre de femmes de ceux-ci[32].

Ces deux approches indirectes sont donc caractérisées par des biais allant dans un sens opposé. A partir d'une enquête menée au Pakistan en 1984, Blacker et Mukiza-Gapere (1988) fournissent une comparaison entre l'âge moyen tel qu'il est estimé *directement* (en tabulant les naissances par

un âge M donné, mais pour un âge $M + 0.75$, sans quoi les estimations sont légèrement biaisées à la hausse (Timaeus 1992).

31. Par rapport à ce calcul direct, Blacker (1977) fait remarquer qu'au Botswana et en Gambie, lorsque les enfants ont été classés en fonction de l'âge de leurs pères, une large proportion de naissances n'étaient pas associées à un âge, tout simplement parce les pères n'étaient pas présents dans les ménages au moment de l'enquête ou du recensement. Etant donné que ces pères absents sont généralement plus *jeunes* que les pères présents dans les ménages (en raison des migrations de travail), ce biais de sélection cause une sur-estimation de l'âge moyen des pères ($M_{♂}$), ce qui biaise les estimations de mortalité à la baisse.

32. Blacker (1977) (p.128) fournit un ajustement *adhoc*, basé sur une relation simple observée entre la prévalence de la polygamie et sa distribution par âge. Lorsque l'on dispose par âge du nombre moyen de femmes par homme, les données peuvent être directement pondérées.

l'âge des pères) et l'âge moyen tel qu'il est estimé *indirectement* par ces deux options. Ils constatent effectivement une *sous-estimation* par la première option (basée sur les âges au premier mariage), et une *sur-estimation* par la seconde (basée sur les âges médians des mariés), d'environ un an dans les deux cas.

Nous avons mené les mêmes calculs sur l'enquête tchadienne de 1964, qui fournit les éléments pour calculer cet âge moyen à la fois directement et indirectement (INSEE-SEDES 1966). Le tableau 3.8 en reprend les principaux résultats. Comme précédemment, on observe une sous-estimation de l'âge moyen des pères par la première méthode (de plus de deux ans et demi). La seconde option procure une valeur nettement plus proche de celle estimée directement, quand un ajustement est introduit pour tenir compte de la polygamie [33].

Calcul direct	
Age moyen des mères à la naissance des enfants	27.1
Age moyen des pères à la naissance des enfants	**36.3**
Approximation 1	
Age moyen au premier mariage (♀)	16.5
Age moyen au premier mariage (♂)	23
Ecart	6.5
$M_{♂}$ estimé	**33.5**
Approximation 2	
Age médian des femmes mariées	30.2
Age médian des hommes mariées	38.5
Ecart	8.3
Ecart ajusté pour la polygamie	9.7
$M_{♂}$ estimé	35.4
$M_{♂}$ estimé ajusté pour la polygamie	**36.8**

TABLE 3.8 – *Calculs directs et indirects de l'âge moyen des pères à la naissance de leurs enfants - Enquête démographique du Tchad, 1964*

Nous renvoyons à l'annexe A.3.4 (p.432) pour une discussion de ces deux approches indirectes à partir des enquêtes EDS. En définitive, nous pensons qu'il est préférable d'estimer $M_{♂}$ directement à partir des naissances survenues dans les douze derniers mois, quand cette information est disponible. Par défaut, cet âge peut être estimé à partir des âges médians, mais en ajustant les distributions par âge en fonction de l'importance de la polygamie et en cherchant, quand cela est possible, à tenir compte de la troncature des données dans les fichiers individuels (*cf.* annexe).

Modélisation de la fécondité masculine

Le calcul des coefficients de correction adaptés à la survie des pères diffère en trois points de celui appliqué à la survie des mères : (1) le stan-

33. Le nombre d'hommes mariés est simplement ici multiplié par le nombre moyen de femmes déclaré par mari pour chaque groupe d'âge.

dard de fécondité est légèrement modifié, (2) l'équation de base est revue, (3) les intervalles d'âges sur lesquels se font les calculs sont différents.

– Premièrement, un *modèle de fécondité masculine* doit être utilisé. Brass et Hill (1973) proposent une série de facteurs de pondération en modélisant la fécondité des hommes au moyen d'une fonction polynomiale comparable à celle utilisée pour les femmes (annexe A.3.3, p.431) - mais étendue aux âges élevés. L'âge de départ est relevé, car les hommes sont généralement plus vieux que les femmes à la naissance de leurs enfants. Les valeurs de $W(n)$ sont reprises dans le tableau A.3.2, en annexe (p. 428).

  Timaeus (1992) fournit une seconde série de coefficients, basée sur une modélisation plus fine de la fécondité masculine. Il utilise toujours le modèle relationnel de Booth (comme pour les mères), mais avec un autre standard ; celui développé par Paget et Timaeus (1994) pour la fécondité masculine. Il s'agit en fait du standard féminin, étendu de sorte qu'au lieu de varier de 10 à 50 ans, il varie de 10 à 80 ans. Ces deux schémas sont présentés dans le graphe 3.12.

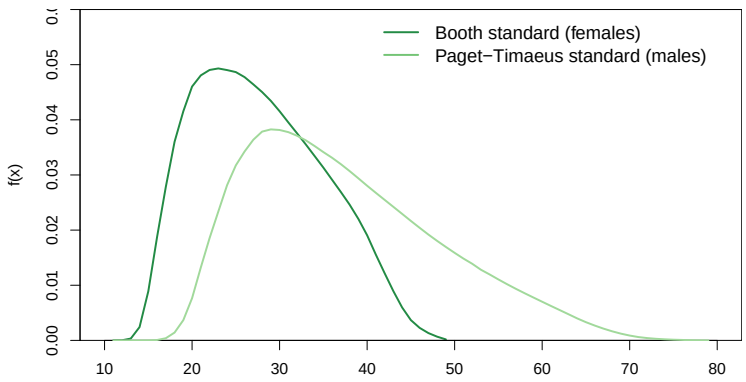

FIGURE 3.12 – *Standards de Booth (fécondité féminine) et de Paget-Timaeus (fécondité masculine) - $\alpha = 0$ & $\beta = 1$*

– L'*équation de base* utilisée pour calculer les proportions théoriques est également revue afin de tenir compte de la survie entre la conception et l'accouchement ($_ap_{a-0.75}$) :

$$ {}_5S_x = \frac{\int_x^{x+5} e^{-r(t)} \times_t p_0 \int_\alpha^\beta e^{-r(a)} \times_{a+t} p_0 \times_a p_{a-0.75} \times f(a)da\ dt}{\int_x^{x+5} e^{-r(t)} \times_t p_0 \int_\alpha^\beta e^{-r(a)} \times_a p_0 \times f(a)da\ dt} \quad (3.9) $$

– Enfin, l'équation masculine inclut les proportions déclarées par un *second groupe d'âge* pour affiner les estimations, car l'ajustement est moins bon pour la mortalité masculine, notamment en raison de la plus grande variance des âges paternels à la naissance de leurs enfants :

$$ {}_np_{35} = \beta_0 + \beta_1 \times M + \beta_2 \times_5 S_{n-5} + \beta_3 \times_5 S_n \quad (3.10) $$

Après avoir fait ces trois changements, les coefficients β peuvent se calculer comme précédemment, par régression. Ils sont présentés dans le tableau 3.9.

n	β_0	β_1	β_2	β_3	R^2	CV
10	-0.8251	0.00261	2.7269	-0.9953	0.978	0.0066
15	-0.4013	0.00576	1.5602	-0.3522	0.974	0.0117
20	-0.3329	0.01031	0.6656	0.3419	0.976	0.0172
25	-0.4726	0.01559	0.2161	0.7896	0.981	0.0216
30	-0.7056	0.02079	0.1997	0.9066	0.988	0.0239
35	-0.9153	0.02493	0.3484	0.8631	0.992	0.0261
40	-0.9950	0.02635	0.4269	0.8263	0.987	0.0466

TABLE 3.9 – *Coefficients utilisés pour convertir les proportions de pères survivants en probabilités de survie - Timaeus (1992)*

Notons que pour les pères, seule la valeur de la probabilité ${}_{10}p_{35}$ peut être calculée à partir des enquêtes EDS, qui ne collectent des données sur la survie des parents que pour les moins de 18 ans. Il faudrait idéalement pouvoir dater cette estimation, mais nous verrons que les méthodes disponibles actuellement ne le permettent pas (car elles supposent que l'on puisse calculer la proportion ${}_{10}S_{10}$).

3.2.5 Applications aux microsimulations

Une fois définis les taux de fécondité, les taux de mortalité, et les valeurs du taux de croissance, tous les éléments nécessaires pour porter l'analyse dans le cadre des microsimulations sont réunis. Dans un premier temps, nous répliquons simplement les calculs menés de façon analytique. L'objectif consiste à évaluer si nous pouvons retomber sur les mêmes coefficients lorsque nous nous situons au niveau individuel.

Pour traduire les simulations de Timaeus (1992) dans le cadre des microsimulations, il suffit de convertir les différentes séries de taux de fécondité et de mortalité de façon à ce que SOCSIM puisse les utiliser [34]. 96 populations différentes sont construites, car il y a 96 combinaisons différentes de taux.

Dans la pratique, nous fixons la taille des populations de départ de façon à obtenir environ 60 000 personnes au terme de 150 ans de simulations (le taux de croissance est fixé à 0.02). En recalculant les taux de croissance, nous récupérons des valeurs proches de 2% ; la moyenne des taux s'établit à 2.02% [35]. Les structures par âges sont celles de la population stable équivalente aux taux introduits. Les taux de mortalité récupérés sont comparés aux taux introduits dans le graphe 3.13, pour différentes valeurs du paramètre α.

A partir des populations simulées, nous pouvons calculer, par âge, les proportions d'orphelins de mère pour les différentes simulations. Les courbes correspondantes sont présentées dans le graphe 3.14.

Nous calculons ensuite les proportions d'orphelins de mère par

34. Pour distribuer les décès au cours des 12 premiers mois, nous utilisons la répartition "type" issue des EDS et détaillée page 409.

35. Les populations finales valent en moyenne 62 000 personnes, et pour atteindre ces populations, entre 80 000 et 149 000 individus interviennent dans les simulations.

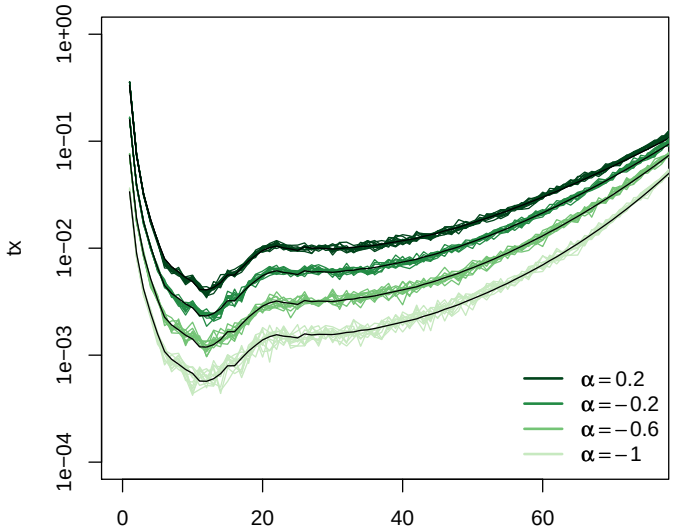

FIGURE 3.13 – *Taux de mortalité récupérés des microsimulations, pour différentes valeurs de α ($\beta = 1$)*

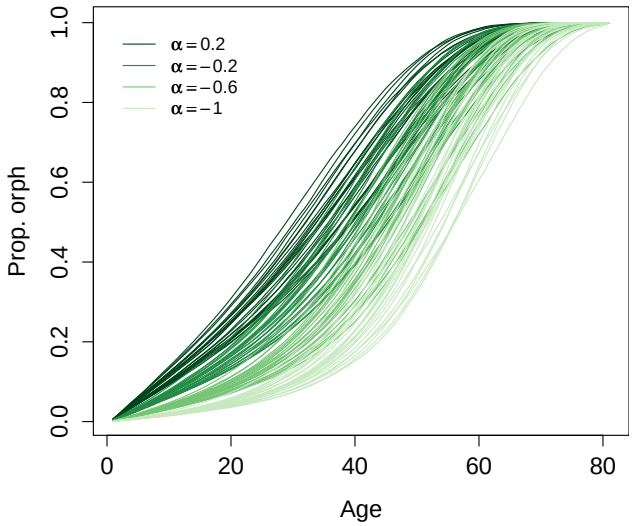

FIGURE 3.14 – *Proportions d'orphelins par années d'âge selon les paramètres retenues - 96 simulations*

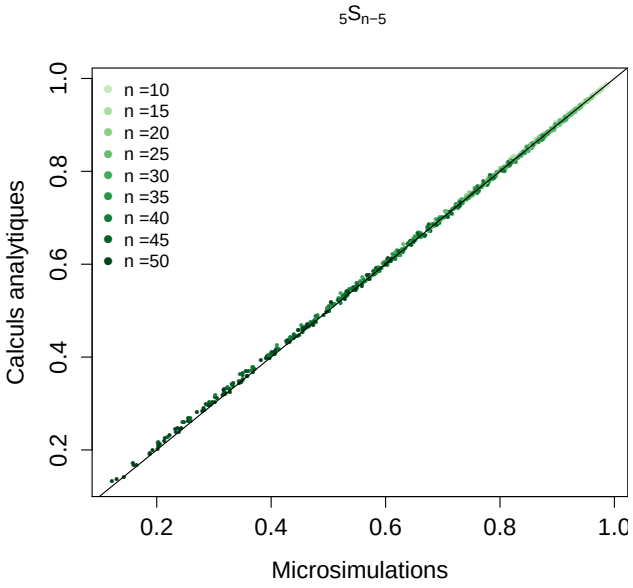

FIGURE 3.15 – *Proportions d'orphelins par groupes d'âge simulées et proportions estimées analytiquement*

groupe d'âge, et nous les comparons aux proportions calculées de façon analytique. Les deux séries de valeurs sont en fait quasiment identiques, comme on l'observe dans le graphe 3.15.

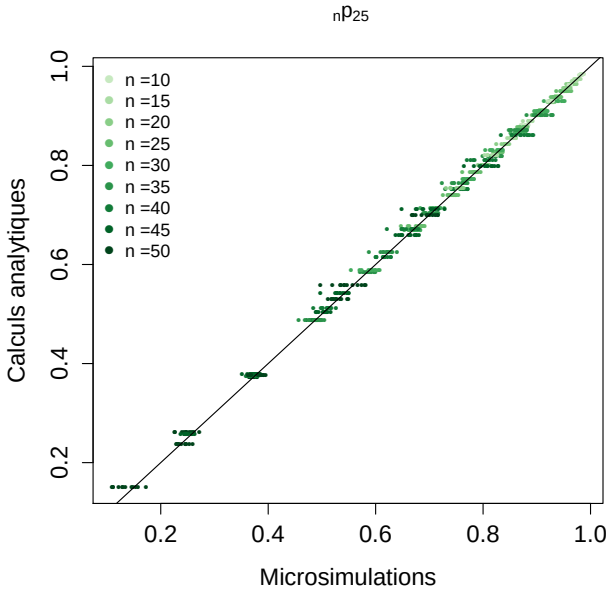

FIGURE 3.16 – *Probabilités de survie estimées à partir des orphelins et probabilités de survie de la table*

Dès lors, en appliquant les coefficients de Timaeus (1992) aux proportions d'orphelins par groupe d'âge, nous retrouvons les probabilités de survie de la table (graphe 3.16). En d'autres termes, la méthode des orphelins appliquée aux microsimulations nous permet de retomber exactement sur les probabilités de survie "réelles", lorsque les niveaux de mortalité et de fécondité qui sont introduits dans les simulations sont ceux qui sont à la base du calcul des coefficients de façon analytique.

3.3 Applications aux cas de fécondité et de mortalité non stables

Jusqu'alors, nous avons raisonné dans le cadre de populations stables, où les niveaux de mortalité et de fécondité restaient constants sur une longue période. Cette situation n'est pratiquement plus jamais rencontrée. Il faut donc ajuster la méthode en conséquence, ou bien recourir à des procédures permettant de dater les estimations obtenues.

Les changements du calendrier de la fécondité sont assez facilement pris en compte ; il suffit de faire varier l'âge moyen à la naissance des enfants selon le groupe d'âge considéré (section 3.3.1). Par contre, le cas d'une mortalité changeante est plus complexe, et plusieurs options sont possibles (section 3.3.2).

3.3.1 Fécondité non stable

A la base, la méthode de la survie des parents suppose que l'âge moyen à la naissance des enfants soit resté stable sur environ 35 - 40 ans. Or, au cours de la transition démographique, cet âge moyen peut connaître

différentes évolutions. A titre indicatif, les tendances de cet âge moyen sont présentées dans l'encadré 8.

Encadré 8 **Tendances de l'âge moyen à la naissance des enfants**

Durant le processus de transition démographique, l'âge moyen à la naissance des enfants tend généralement à diminuer (Bongaarts 1999). Mais il peut également augmenter, car son évolution dépend principalement de deux facteurs : le *retardement des naissances* chez les plus jeunes mères et la *limitation des naissances* chez les mères les plus âgées. L'âge moyen M augmente lorsque les naissances sont retardées (en particulier celles des premiers rangs, suite par exemple à une hausse de l'âge au mariage), il diminue lorsque les naissances de rangs élevés sont réduites.

Pour reconstituer les tendances de l'âge moyen à la naissance des enfants, nous avons rassemblé un peu plus de 120 estimations calculées à partir de différentes sources :

- 89 enquêtes EDS africaines où une question a été posée aux femmes sur le nombre de naissances survenues dans les 12 derniers mois[36],
- une vingtaine d'estimations issues de l'Annuaire statistique des Nations Unies, lorsque les naissances sont tabulées par âge des mères (UN 2009a; 1997),
- une dizaine d'estimations calculées à partir des fichiers individuels de recensements distribués par le projet IPUMS.

Ces différentes valeurs sont présentées dans le graphe 3.17. Les estimations sont représentées par des ronds si elles sont issues des EDS, et par des carrés si elles sont issues des registres de population, des recensements ou d'autres enquêtes.

Relevons d'abord les différences obtenues selon le type de données. Par exemple, au Kenya, la valeur de M dégagée du recensement de 1989 s'élève à 26.4 ans, alors que celle issue de l'enquête EDS de la même année se chiffre à 27.5 ans. De façon générale, les estimations issues des recensements apparaissent plus basses que celles dégagées des EDS ; c'est le cas du Kenya, de l'Ouganda, du Sénégal, du Mali... Ces écarts peuvent peut-être s'expliquer par une sous-déclaration plus importante des naissances par les femmes âgées dans les recensements.

Dans l'ensemble, on observe une série de parcours nationaux assez diversifiés. La valeur de M s'établit en moyenne à 27.2 ans, et elle varie de 25 ans (au Kenya en 1999) à 29.4 (au Rwanda en 2000).

- Dans un premier groupe de pays, l'âge moyen a connu *une baisse assez nette sur les 20-25 dernières années*. Ce groupe rassemble majoritairement des pays d'Afrique australe (Malawi, Namibie, Zimbabwe, Mozambique), ainsi que le Kenya et l'Ouganda (où les tendances sont plus erratiques car elles combinent des estimations issues des recensements et des EDS). Au Malawi, la baisse a été amorcée dès les années 1970 et elle s'est poursuivie de façon régulière jusqu'en 2010.
- Un second groupe reprend une dizaine de pays qui ont connu *une hausse de l'âge moyen à la naissance des enfants*. Ce groupe est constitué essentiellement de pays d'Afrique de l'Ouest (Bénin, Burkina Faso, Guinée, Niger, Sénégal), de quelques pays d'Afrique de l'Est (Ethiopie, Tanzanie), ainsi que de la Zambie.
- Enfin, dans un troisième groupe de pays, les tendances sont moins nettes, soit parce qu'une seule estimation est disponible (Lesotho, Gabon, Comores, etc.), soit parce que la valeur de M est restée relativement stable sur la période. Avant 1980, les estimations relatives à Madagascar, au Togo, au Liberia et au Rwanda témoignent de cette apparente stabilité.

36. Il s'agit de toutes les enquêtes accessibles au public à la date de juillet 2010, en excluant l'enquête Angola 2007 (AIS), Sénégal 1999 (fichier non standard), Soudan 1990 (classé en Afrique du Nord) et Tanzanie 2003 (AIS).

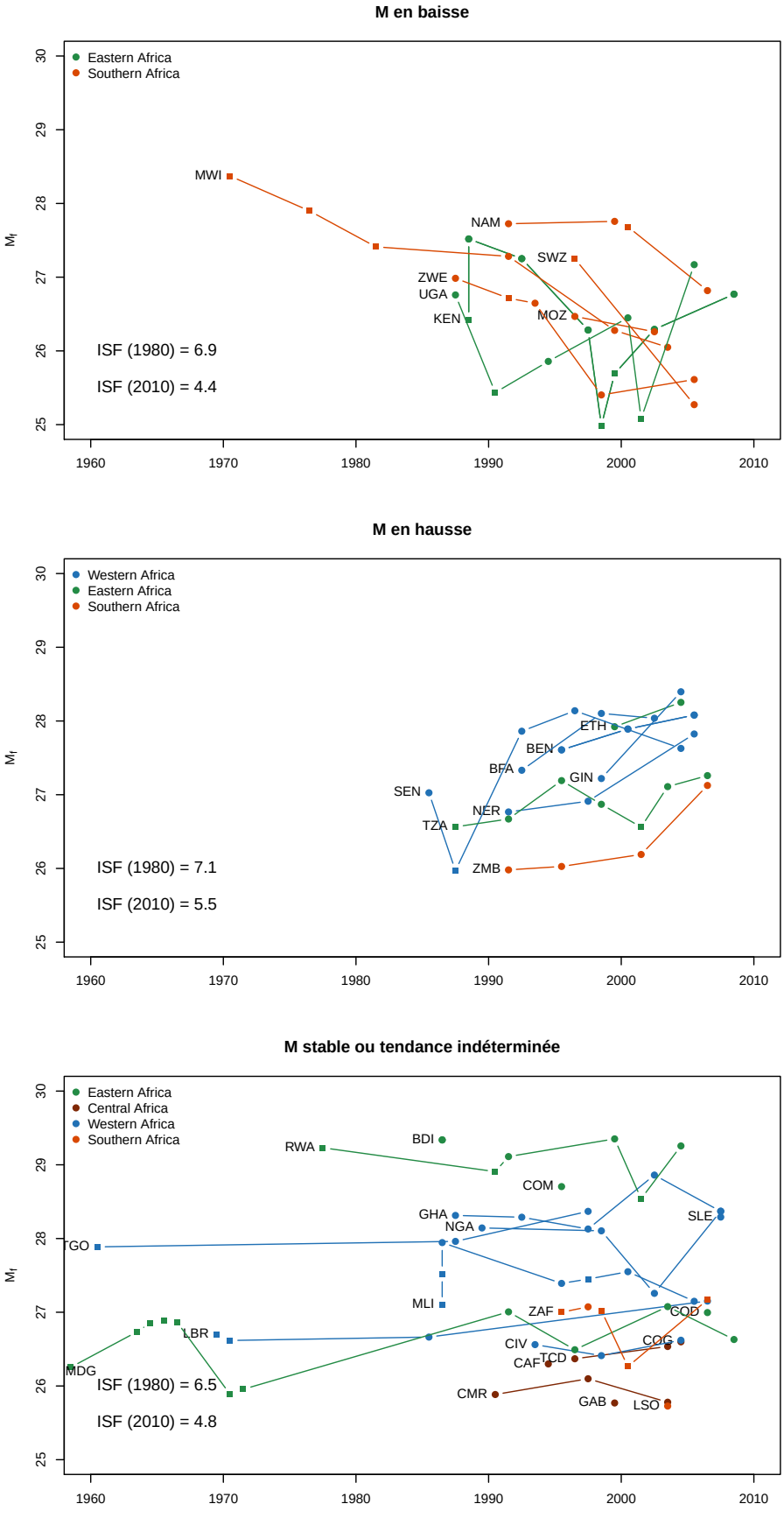

FIGURE 3.17 – *Age moyen à la naissance des enfants calculé à partir des enquêtes EDS et des recensements (n = 121)*

Les pays où l'âge moyen à la naissance des enfants est plutôt orienté à la baisse sont ceux qui ont entamé le plus tôt leur déclin de fécondité. Le Kenya et le Zimbabwe, notamment, ont enregistré une baisse particulièrement rapide (Kirk et Pillet 1998). Le niveau de fécondité est légèrement plus bas dans ce premier groupe : en 1980, l'ISF moyen (non pondéré) des pays de ce groupe se chiffrait à 6.9 enfants par femme, et il a chuté à 4.4 enfants en 2010. Dans le second groupe, où l'âge moyen est plutôt à la hausse, la réduction de la fécondité est moins rapide : de 7.1 enfants par femme en moyenne en 1980, ces pays ont atteint 5.5 enfants en 2010 (UN 2009c).

Les différents groupes se distinguent peu au niveau de l'âge moyen au premier mariage des femmes. Ils connaissent globalement tous les trois une légère hausse de l'âge d'entrée en union. Ils se distinguent par contre davantage lorsque l'on examine l'évolution de la *proportion des naissances de rang élevé*. A partir des histoires génésiques des femmes enquêtées, le graphe 3.18 présente l'évolution de la proportion des naissances de rang 6 et plus parmi l'ensemble des naissances. Nous avons calculé cette proportion sur les 5 ans précédant chaque enquête. On constate une baisse quasiment généralisée de cette proportion, surtout là où l'âge moyen à la naissance diminue - ce qui traduit une tendance plus forte dans ces pays en faveur d'une limitation des naissances chez les mères de parité élevée. Cette baisse est moins prononcée dans les pays où l'âge moyen augmente, et elle est globalement peu apparente en Afrique de l'Ouest, à l'exception de la Côte d'Ivoire (ce qui devrait être confirmé).

Le graphe 3.19 présente ensuite l'âge moyen *à la première naissance*, calculé comme l'âge moyen toutes naissances confondues, mais en retenant uniquement les femmes dont c'est la première naissance. Cet âge est calculé à partir d'un fichier restructuré en biographique sur les 5 ans précédant chaque enquête EDS. Dans les trois groupes, on peut observer une légère hausse de cet âge moyen. C'est particulièrement le cas au Sénégal, au Ghana et au Kenya. Ceci va de pair avec une légère hausse de l'âge d'entrée en union.

En définitive, on peut mettre en évidence trois types d'évolutions. Au Kenya, au Malawi, et au Zimbabwe, notamment, le recul de l'âge à la première naissance est plus que compensé par une limitation des naissances dans les familles nombreuses. Ceci se traduit par une baisse de l'âge moyen à la naissance des enfants. Dans le second groupe, c'est plutôt le retardement des premières naissances qui domine, et les valeurs de M connaissent une légère évolution à la hausse. Enfin, dans le dernier groupe, la stabilité de l'âge moyen (notamment en Côte d'Ivoire, au Ghana, ou au Cameroun) pourrait s'expliquer par un recul de l'âge à la première naissance concomitant d'une réduction de la fécondité aux parités élevées.

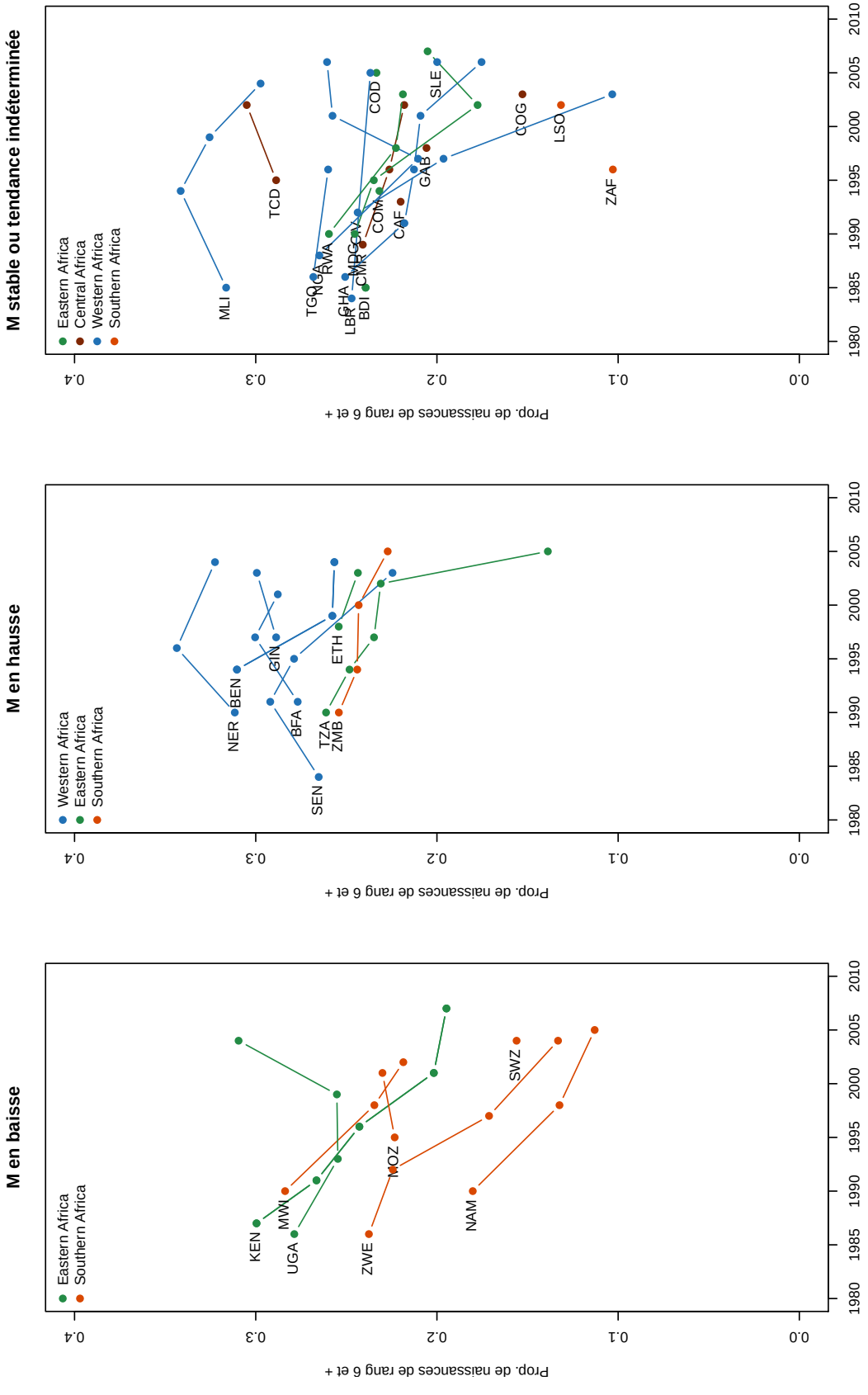

FIGURE 3.18 – *Proportions des naissances de rang 5 et + dans les 5 ans précédant les enquêtes EDS (n = 89) - calculs menés sur toutes les femmes de 15 à 49 ans*

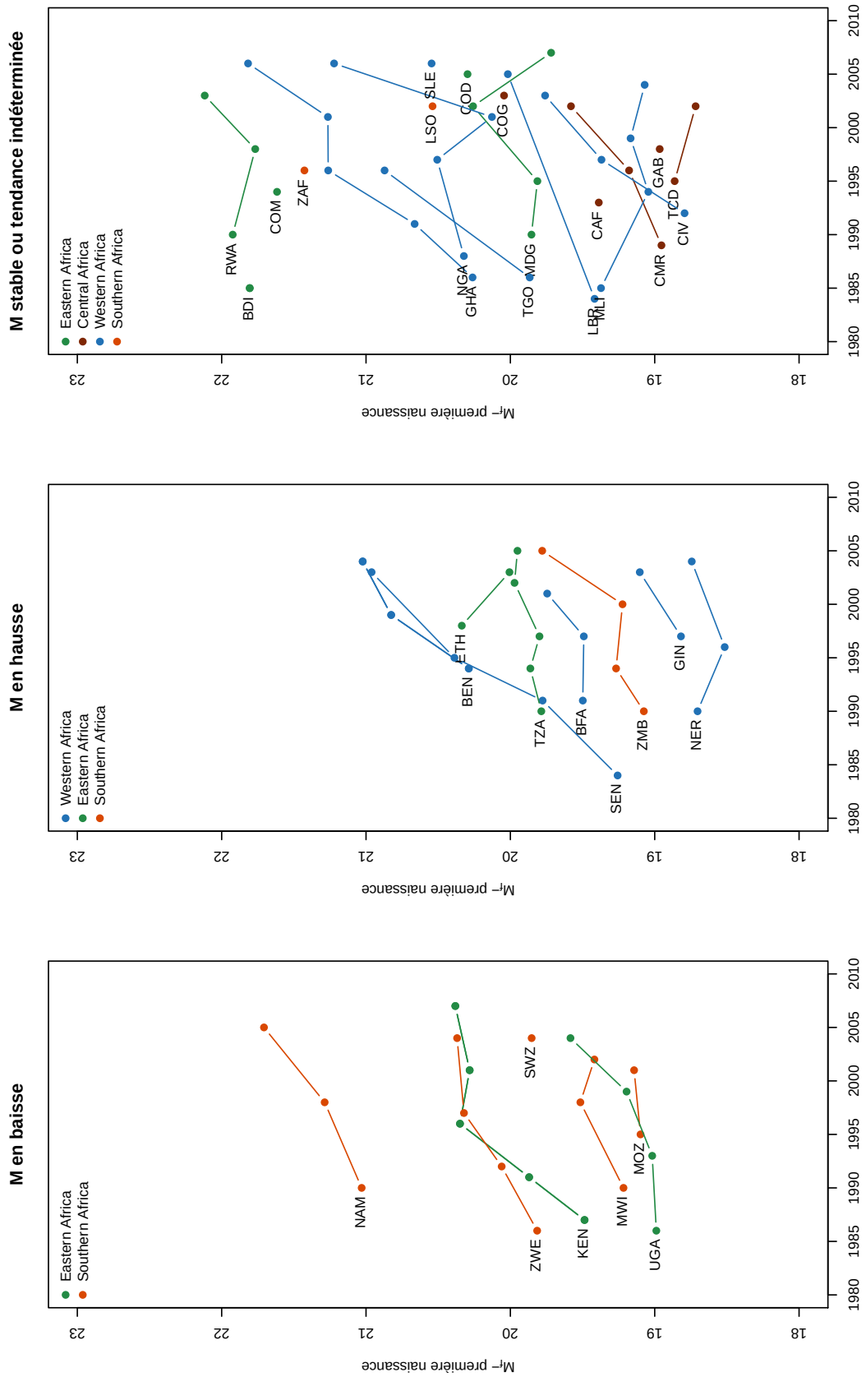

FIGURE 3.19 – *Age moyen à la naissance du premier enfant dans les 5 ans précédant les enquêtes EDS (n = 89) - calculs menés sur toutes les femmes de 15 à 49 ans*

Quel est l'impact de ces évolutions de l'âge moyen à la naissance des enfants sur les estimations de mortalité ?

- si l'âge moyen à la naissance a baissé, utiliser l'âge du moment va induire une sur-estimation de la mortalité.
- à l'inverse, si l'âge moyen a connu une évolution à la hausse, la mortalité sera sous-estimée : la période d'exposition observée aujourd'hui est trop longue par rapport à la durée d'exposition réelle [37].

Dans la pratique, appliquer la méthode des orphelins aux situations où l'âge moyen à la naissance des enfants a changé est relativement simple. Plutôt que faire intervenir l'âge moyen *du moment* dans le calcul des probabilités de survie, on préfèrera estimer la valeur de M *au moment de la naissance de ces enfants*. Une valeur spécifique de M sera donc retenue pour chaque groupe d'âge, en calculant cet âge au moment de la naissance de chaque cohorte.

Le problème, bien entendu, est celui de la disponibilité des données. Dans certains pays, plusieurs enquêtes sont disponibles et elles procurent des tendances qui paraissent plausibles (Zimbabwe, Tanzanie, Malawi, etc.). Par contre, dans d'autres, soit les estimations sont rares, soit les tendances sont peu claires (Sénégal, Ouganda), soit les estimations issues de différentes sources semblent peu compatibles (voir encadré précédent).

3.3.2 Mortalité non stable

Le cas d'une mortalité non stable est plus complexe que le cas d'une fécondité en transition. A partir d'une seule opération de collecte, les proportions de non-orphelins parmi les enquêtés plus âgés font référence à une mortalité *plus ancienne* que les mêmes proportions calculées à partir des répondants *plus jeunes*. Les parents de ces derniers ont en effet été moins longtemps exposés au risque. Dans ces conditions, si les risques de décès ont évolué dans le temps, les estimations obtenues indirectement mêlent des niveaux de mortalité assez variables.

Différentes méthodes ont donc été développées, soit pour dater les estimations, soit pour éliminer l'effet des tendances :

- La première option consiste à estimer une période de référence de façon indirecte.
- Une seconde option consiste simplement à dater "directement" les

37. Ceci peut s'expliquer autrement. Nous avons vu que la proportion de non-orphelins d'âge a peut se calculer comme une moyenne des probabilités de survie féminine, moyenne pondérée par la structure par âge des mères à la naissance il y a a années (*cf.* p.105) :

$$S(a) = \int_{\alpha}^{\beta} {}_{a}p_x \, W(x|t-a)dx$$

Si l'âge moyen à la naissance des enfants était plus élevé il y a a années, la distribution des poids $W(x|t-a)$ favorise davantage les femmes plus âgées, et les probabilités de survie ${}_{a}p_x$ correspondant à des valeurs de x élevées auront donc plus de poids. Ceci a pour effet de réduire la proportion de non-orphelins d'âge a.

estimations à partir des années de décès, quand cette information est disponible.
– Une troisième possibilité consiste à comparer les proportions de parents survivants déclarés dans *deux opérations de collecte successives* (idéalement de même nature et qualité), afin d'obtenir des estimations plus récentes. Cette troisième approche permet de "neutraliser" l'effet perturbateur des tendances.

Ces trois options sont présentées dans les quelques pages qui suivent.

Méthode indirecte de datation

Si la mortalité a baissé durant les années précédentes, le niveau observé aujourd'hui de façon *rétrospective* (c'est-à-dire au niveau des cohortes des parents) correspond à un niveau de mortalité qui a pu être observé en *transversal* quelque part entre les niveaux élevés du passé lointain et les niveaux faibles du présent. L'idée consiste alors à trouver le moment du passé pour lequel il existe une égalité entre, d'une part, la mortalité transversale à ce moment et, d'autre part, la mesure rétrospective de la mortalité dégagée de la survie des parents (avec la même exposition) :

> "The idea, following Feeney, is to determine for each measure of retrospective, cohort type survivorship, the equivalent time point such that the survivorship with the period life table of that time and corresponding exposure to risk would be equal." (Brass et Bamgboye 1981)

Ce moment est qualifié de "période de référence", et il est exprimé sous la forme d'une durée. Il renvoie à la durée écoulée depuis que le niveau de mortalité calculé à partir des cohortes ait pu être observé en transversal. Par définition, cette période de référence se situe entre la date de l'enquête (ou du recensement) et l'âge des enfants qui fournissent des informations sur leurs parents. Elle va augmenter avec l'âge des enquêtés.

Comment estimer cette période de référence ? Prenons le cas le plus simple, celui où la fonction de survie l_x est linéaire sur les âges des parents. Dans ce cas, les décès des mères des enfants d'âge n seront distribués de façon uniforme sur les n dernières années. La période de référence s'établit alors à $n/2$. Comme on l'a relevé précédemment, la fonction de survie est non linéaire, surtout pour des niveaux de mortalité intermédiaires et aux âges plus élevés (*cf.* graphe 3.1, p.107). Sa courbure vers le bas implique que les décès surviennent davantage aux âges élevés. Ils sont donc davantage répartis *dans les années proches de l'enquête ou du recensement*. La période de référence devrait être plus proche elle aussi. Elle est alors exprimée en faisant intervenir, en plus de $n/2$, un facteur de correction $C(n)$ (UN 2002b) :

$$T(n) = (n/2) \times (1 - C(n)) \tag{3.11}$$

Brass et Bamgboye (1981) montrent, de façon analytique, que ce facteur de correction $C(n)$ peut être calculé comme [38] :

$$C(n) = ln({}_{10}S_{n-5})/3 + Z(n+M) + 0.0037(27-M) \qquad (3.12)$$

Dans cette équation, M renvoie toujours à l'âge moyen à la naissance des enfants, tandis que ${}_{10}S_{n-5}$ correspond à la proportion de non-orphelins pour une classe d'âge de 10 ans centrée autour de n. Le terme $Z(n+M)$ renvoie à des valeurs types entre lesquelles il faut interpoler pour obtenir la valeur correspondante à $n+M$. Ces valeurs sont reprises en annexe A.3.2 (p.430) [39].

Remarquons que dans le cas des enquêtes EDS - où les informations sur la survie des parents sont généralement collectées à partir des moins de 18 ans - une seule estimation peut être datée ; celle qui est issue des enfants de 5 à 9 ans. Les proportions de non-orphelins de 10 à 14 ans peuvent être converties en probabilités de survie, mais elles ne peuvent être *datées* ; en effet, la valeur de ${}_{10}S_{10}$ ne peut pas être calculée à partir des moins de 18 ans. On l'a déjà relevé, mais répétons-le : étendre les questions sur la survie des parents à tous les âges permet de reconstituer de plus longues tendances et de mieux juger de la plausibilité des estimations. On peut espérer que cette option soit envisagée pour les prochains rounds d'enquêtes EDS.

Dans les recensements, les questions sur la survie des parents sont plus fréquemment étendues à tous les âges. Nous pouvons donc estimer la mortalité sur une plus longue période. A titre illustratif, nous avons appliqué cette méthode de datation aux proportions de non-orphelins issues des recensements est-africains présentés précédemment : Kenya (1989/1999), Rwanda (1991/2002), Tanzanie (1988/2002), et Ouganda (1991/2002) [40]. Nous y avons ajouté les estimations issues de trois recensements menés en Afrique du Sud (1996, 2001 et 2007) [41], ainsi que de deux recensements menés au Mali (en 1987 et 1998) [42].

38. Le développement nécessaire pour atteindre cette équation est relativement complexe, et nous renvoyons à leur document pour plus de détails.

39. Appliquées à la mortalité masculine, les équations 3.11 et 3.12 deviennent respectivement :

$$\begin{aligned} T(n) &= (\frac{(n+0.75)}{2})(1-C(n)) \\ C(n) &= ln({}_{10}S_{n-5})/3 + Z(n+M) + 0.0037(27-M+0.75) \end{aligned} \qquad (3.13)$$

40. Les coefficients utilisés pour convertir les proportions de non-orphelins en probabilités de survie sont ceux de Timaeus (1992), et la procédure de datation est celle de Brass et Bamgboye (1981).

41. Ces données sont également disponibles via le projet IPUMS. En 1996, le fichier de données censitaires ne reprend pas les informations nécessaires pour calculer l'âge moyen à la naissance des enfants. Nous avons donc supposé que la valeur de M s'établissait à 27 ans, ce qui correspond à la valeur observée dans l'enquête EDS de 1998.

42. Pour le Mali, les données [43] ne permettent pas de calculer directement l'âge moyen à la naissance des enfants. Pour 1987, nous avons retenu une valeur de 27.5, ce qui correspond à une moyenne entre la valeur calculée à partir du Demographic Yearbook pour 1987 (à partir du recensement) (UN 1997) et l'enquête EDS de 1987. Pour 1998, nous avons interpolé entre les valeurs disponibles en 1996 et 2001 pour fixer la valeur de M à 27.4.

Les différentes valeurs de ${}_{45}q_{15}$ sont reprises dans le graphe 3.20. Il ne s'agit pas d'estimations définitives, notamment parce que les coefficients utilisés ne tiennent pas compte de l'impact du VIH-sida, et parce que les valeurs de ${}_{45}q_{15}$ sont obtenues par interpolation au sein du standard général de Brass (un schéma peu adapté aux contextes perturbés par le VIH-sida.). Néanmoins, à ce stade, nous pouvons déjà observer :

- une bonne concordance entre les estimations issues des différents recensements, surtout au Mali, au Kenya et en Afrique du Sud (panel du bas),
- une hausse de mortalité très nette au Rwanda (dès 1980), au Kenya (à partir de 1985) et en Afrique du Sud (après 1990),
- une baisse régulière de la mortalité adulte au Mali entre 1970 et 1990,
- des tendances un peu moins cohérentes en Ouganda et en Tanzanie, où la baisse apparente au cours des années 1980 est suivie d'une série d'estimations en forme de U renversé autour de 1990-1995. Il est probable que ceci soit lié à des distorsions introduites par le VIH-sida.

A partir du cas du Kenya, il est possible de prolonger la tendance avant 1970. Les informations sur la survie des parents ont en effet été collectées tous les dix ans, entre 1969 et 1999. A nouveau, ces données sont converties en probabilités de survie à l'aide des coefficients de Timaeus (1992), traduits en probabilité de décès entre 15 et 60 ans à partir du standard de Brass, puis datées par la méthode de Brass et Bamgboye (1981) [44]. Les estimations obtenues sont présentées dans le graphe 3.21.

Les niveaux dégagés des *trois premiers recensements* constituent une bonne illustration d'une tendance tout à fait implausible. La baisse de la mortalité semble sur-estimée à partir de chacun de ces trois recensements, ce qui procure une série de trois courbes assez parallèles. Timaeus (1986) a déjà relevé cette incohérence, et il l'attribue au *biais d'adoption*. Ce terme renvoie au fait que le chef de ménage (qui est souvent celui qui répond aux questions sur la survie des parents) peut procurer des informations sur d'autres personnes que les parents biologiques des membres de son ménage. Cela advient surtout lorsque des adultes considèrent leurs enfants confiés ou les enfants adoptifs comme leurs propres enfants. Dans ce cas, si les parents biologiques de ces enfants sont décédés, les proportions d'orphelins seront sous-estimées. Ce biais d'adoption est discuté dans le chapitre 6.

Comme le biais d'adoption caractérise surtout les informations relatives aux plus jeunes, *les niveaux de mortalité les plus récents sont davantage sous-estimés que les niveaux plus anciens*. A partir d'une seule opération de collecte, le biais d'adoption conduit donc à surestimer les progrès, ce qui peut expliquer la discordance entre les estimations kenyanes de 1969, 1979 et 1989. Ce problème est partiellement résolu avec le recours aux cohortes synthétiques, une approche présentée dans la section 3.3.3.

44. La tabulation des individus selon la survie de leur mère est reprise de Timaeus (1986). L'âge moyen à la naissance des enfants est estimé en 1969 à 27.4 ans par Blacker (1977). Nous supposons qu'il est resté constant juqu'à 1979.

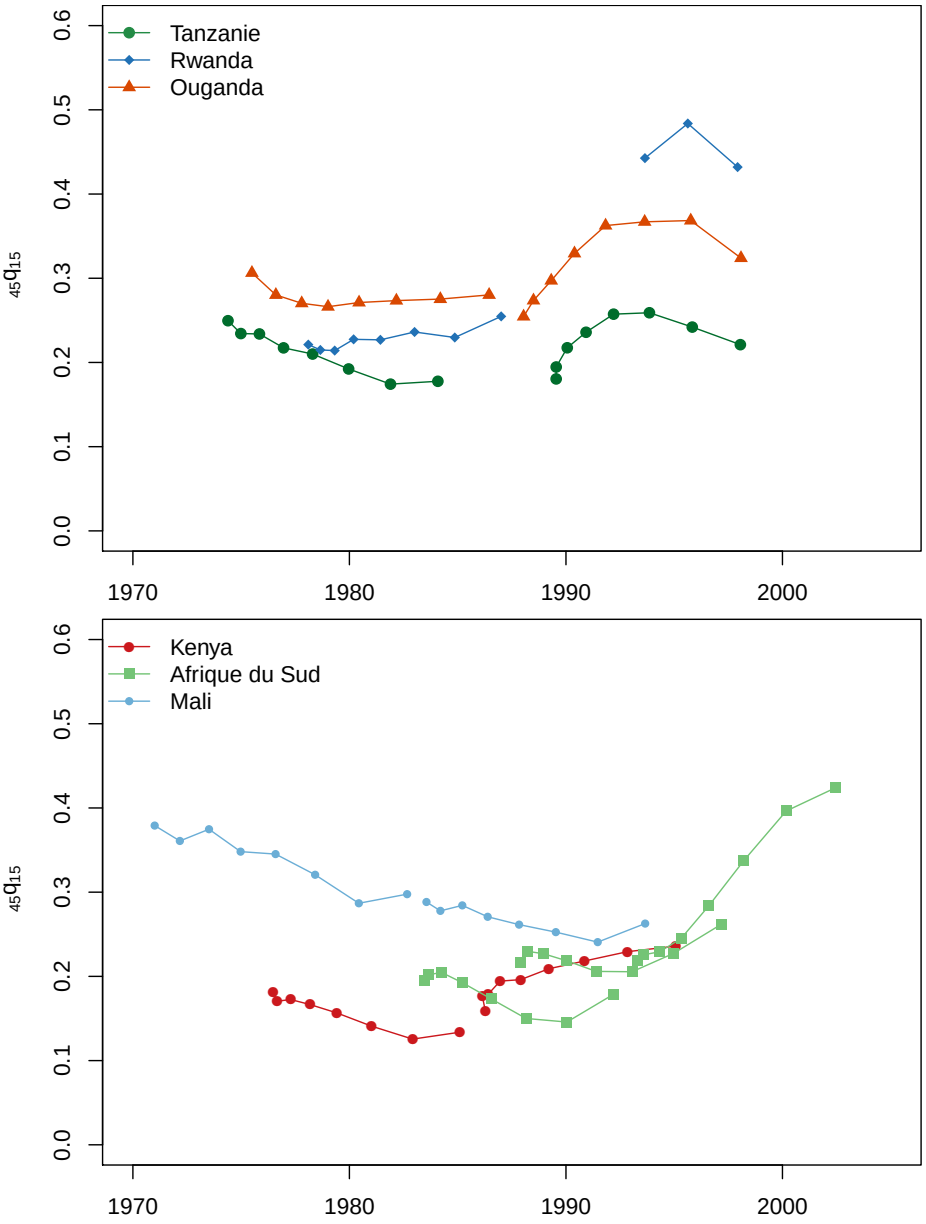

FIGURE 3.20 – *Estimations de ${}_{45}q_{15}$ à partir des proportions de non-orphelins de mères observées au Kenya (1989-99), Rwanda (1991-2002), Tanzanie (1988-2002), Ouganda (1991-2002), Afrique du Sud (1996, 2001, 2007) et Mali (1987, 1998)*

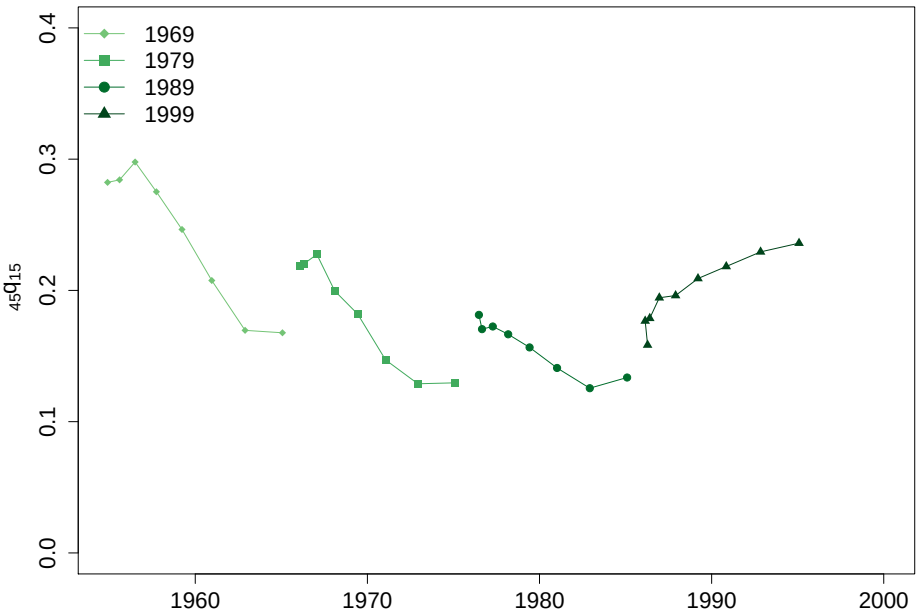

FIGURE 3.21 – *Estimations de ${}_{45}q_{15}$ à partir des proportions de non-orphelins de mères observées au Kenya dans les recensements de 1969, 1979, 1989 et 1999*

Avant de conclure cette section et de présenter d'autres méthodes de datation, il faut préciser ici que Brass et Bamgboye (1981) parviennent à l'équation 3.13 et aux coefficients $Z(n+M)$ en faisant deux hypothèses :

1. la mortalité est modélisée par le biais d'un système logit, et les tendances se traduisent par une évolution linéaire du paramètre de niveau α,
2. l'intensité de la mortalité peut être exprimée comme une fonction linéaire du temps écoulé depuis la période de référence $T(n)$ $(n < T(n) < 0)$.

Palloni et al. (1984) ont généralisé le calcul de cette période de référence en procédant par régression plutôt qu'en travaillant avec une formule de calcul unique, comme le font Brass et Bamgboye (1981). Pour chaque simulation, ils calculent des probabilités de survie à partir des orphelins, puis estiment le temps écoulé depuis que le niveau de mortalité obtenu en rétrospectif ait pu être observé en transversal. Pour construire leurs simulations, ils recourent aux tables des Nations Unies (UN 1982) plutôt qu'au système logit. Au niveau des tendances, ils font uniquement l'hypothèse que l'espérance de vie a augmenté de façon *linéaire* dans les 30 ans qui ont précédé l'opération de collecte, et prévoient six rythmes de

changements [45].

Leur analyse permet de cerner l'importance respective de *trois sources de variation dans les périodes de référence* : (1) le rythme de la baisse de la mortalité, (2) le calendrier de la fécondité et (3) l'interaction entre ces deux éléments. Dans l'ensemble, ils constatent que les variations dans les périodes de référence selon le *rythme de baisse de la mortalité* sont relativement faibles, tout comme celles imputables à l'interaction entre la fécondité et le rythme de cette baisse. Les variations dépendent essentiellement du *calendrier de la fécondité*. Ainsi, pour un groupe d'âge donné, plus la fécondité est tardive, plus la période de référence est lointaine [46].

Palloni et al. (1984) proposent donc des coefficients de régression qui contrôlent l'âge moyen à la naissance des enfants, mais également deux rapports de parités atteintes : $\frac{P_{15-19}}{P_{20-24}}$ et $\frac{P_{25-29}}{P_{30-34}}$. L'effet du rythme de la baisse de mortalité est appréhendé en introduisant la proportion d'enfants non orphelins [47]. La formule de calcul prend la forme suivante :

$$T(n) = \beta_0 + \beta_1 M + \beta_2 \frac{P_{15-19}}{P_{20-24}} + \beta_3 \frac{P_{20-24}}{P_{25-29}} + \beta_4 \, {}_5S_n + \beta_5 \, {}_5S_n \times M \quad (3.14)$$

Les coefficients β sont repris en annexe A.3.2. Notons que seule la mortalité féminine est considérée ici ; la seule option disponible pour dater les estimations issues de la survie des pères est donc la méthode établie par Brass et Bamgboye (1981). Par contre, le mode de calcul de Palloni et al. (1984) permet de dater les estimations issues des proportions d'orphelins de 10 à 14 ans, ce qui n'est pas le cas de Brass et Bamgboye (1981).

Méthode directe

Chackiel et Orellana (1985) proposent une méthode alternative pour dater les estimations : il s'agit de collecter des informations supplémentaires sur l'*année du décès* des parents. A partir de là, pour les différents groupes d'âges des répondants, la moyenne des années des décès peut servir de période de référence. Leurs analyses, menées sur le Honduras, indiquent que cette datation "directe" donne des résultats plus plausibles que ceux obtenus indirectement. Selon les auteurs, la datation indirecte procure une période de référence *trop proche de l'enquête*, surtout à partir des répondants plus âgés (les deux types d'estimations sont quasiment identiques pour les premiers groupes d'âge).

45. D'un niveau de départ de 35 ans, l'espérance de vie peut atteindre après 30 ans un niveau minimal de 45 ans (soit une hausse annuelle constante de 0.33 ans) et un niveau maximal de 70 ans (soit un gain annuel constant de 1.17 ans). La fécondité est supposée constante. 96 schémas de fécondité, issus du modèle de Coale et Trussell (1974), sont utilisés.

46. En effet, les parents qui donnent naissance à un âge élevé (lorsque les risques de mortalité sont comparativement plus élevés) bénéficient davantage de la baisse de la mortalité que les parents qui donnent naissance à des âges plus jeunes (aux risques de mortalité plus faibles). Or, si les décès surviennent comparativement moins aux âges élevés, la période de référence sera plus éloignée de l'enquête.

47. L'interaction entre le rythme de la baisse de mortalité et le schéma de fécondité est contrôlée en introduisant le produit entre la proportion d'enfants non-orphelins et l'âge moyen à la maternité.

Les applications de cette approche directe sont rares en Afrique, car la question sur la date du décès des parents a rarement été intégrée dans les enquêtes nationales. On peut toutefois mentionner le cas de l'enquête EDS menée au Burundi en 1987, dont les données sont exploitées par Makinson (1993) [48].

Aucune enquête EDS récente ne permet de mener une analyse comparable [49]. C'est certainement une question à considérer pour les phases ultérieures du programme EDS (et pour toute autre enquête sur la mortalité adulte). Cette approche est particulièrement intéressante dans les contextes perturbés par le VIH-sida, où, comme on le verra, les hypothèses qui sous-tendent la datation "indirecte" des estimations ne sont plus rencontrées.

Un second avantage de la connaissance des dates de décès est qu'il est possible de reconstituer, pour les différents groupes d'âge, les proportions d'orphelins il y a 5 ans, ou il y a 10 ans. De cette façon, les données peuvent être exploitées comme si elles provenaient *de trois sources différentes*. Elles vont procurer des estimations de mortalité pour une période bien plus longue que celles obtenues à partir d'une seule enquête (Chackiel et Orellana 1985). Il est également possible de calculer des probabilités de survie sur les intervalles de 0-4 ans ou 5-9 ans avant l'enquête, en appliquant les méthodes basées sur les cohortes synthétiques, qui sont détaillées dans la section suivante.

3.3.3 Recours aux cohortes synthétiques

En l'absence de données sur les dates de décès, nous avons vu qu'il existait deux méthodes indirectes de datation. Celles-ci reposent toutefois sur l'hypothèse d'une mortalité évoluant linéairement dans le temps. Cette tendance linéaire est soit appréhendée à partir du paramètre α du système logit, soit à partir des espérances de vie. Ceci est très restrictif, et rend ces méthodes peu utiles dans les contextes perturbés par le VIH-sida. Par ailleurs, dans les deux cas, les périodes de référence vont s'étendre typiquement de 3 à 15 ans. Les estimations obtenues font donc référence à des périodes relativement anciennes. Enfin, nous avons vu que les biais d'adoption pouvaient se traduire par des perturbations dans les tendances.

Afin de disposer d'estimations plus récentes, et de neutraliser l'effet des tendances, il est possible, à partir de deux enquêtes ou de deux recensements, de reconstruire des *cohortes synthétiques*. Cette technique d'analyse, régulièrement utilisé en démographie, consiste dans ce cas-ci à re-

48. Makinson compare la datation directe et indirecte, mais elle observe la situation inverse de celle obtenue au Honduras. Les périodes de référence directes sont situées *dans un passé plus récent* que les périodes calculées indirectement. Ceci est lié aux réponses manquantes. L'auteur précise qu'environ 25% des femmes enquêtées dont la mère est décédée n'ont pas pu dater ce décès, et les réponses manquantes ont été bien plus importantes chez les femmes plus âgées. En conséquence, sans procéder à une imputation, la datation directe va procurer des périodes de référence trop récentes.

49. Notons que cette question sur la date du décès a été posée dans l'enquête LSMS du Malawi en 2004.

construire l'évolution de la proportion d'orphelins par âge qui correspond à celle que l'on pourrait observer *si la mortalité qui prévaut entre les deux opérations de collecte était maintenue constante sur toute la vie des individus* [50]. Ceci présente trois avantages :

1. la *cohérence des estimations obtenues* à partir des différents groupes d'âge peut être plus facilement examinée, puisqu'elle n'est pas influencée par les changements de la mortalité dans le temps.
2. Ces méthodes procurent des estimations *récentes*, et *aucune datation* n'est nécessaire, puisque les estimations dégagées sont relatives à la période qui sépare les deux opérations de collecte.
3. Les estimations sont par ailleurs plus *précises*. En effet, si chacune des enquêtes est biaisée dans une proportion équivalente par l'omission de décès survenus dans un passé lointain, ce biais s'annule lorsque les enquêtes sont combinées. De même, le biais d'adoption est atténué : les estimations issues des adultes (qui ne sont plus susceptibles d'être adoptés) en sont exemptes (si les deux sources de données sont entachées de ce biais dans la même proportion). Seules les proportions d'orphelins calculées sur les moins de quinze ans sont affectées.

Le principal inconvénient de ces approches basées sur les cohortes synthétiques est qu'elles sont particulièrement sensibles *aux différences de précision* entre les deux séries de données (surtout en ce qui concerne les âges des répondants). Comme souvent pour ce genre d'approche comparative, il faut donc faire l'hypothèse que l'ampleur et la structure par âge des erreurs de déclarations reste relativement constante. Par ailleurs, comme nous le verrons, le biais d'adoption est atténué, mais il n'est pas complètement éliminé.

Deux méthodes sont disponibles, selon que les comparaisons entre les deux enquêtes (ou recensements) sont menées *par cohorte* ou *par groupe d'âge*. Dans le premier cas, les proportions de non-orphelins séparés d'un multiple de cinq ans sont simplement assemblées d'une classe d'âge à l'autre pour constituer une cohorte. Dans l'autre cas, on travaillera directement avec les taux de croissance de la population de non-orphelins.

Enchaînement des proportions séparées d'un multiple de cinq ans

Dans le cas de deux opérations de collecte séparées par un multiple de cinq ans, les différentes proportions de survivants peuvent être assemblées d'une classe d'âge à l'autre pour estimer la mortalité dans la période qui sépare les deux opérations (Zlotnik et Hill 1981, Timaeus 1986). La proportion de parents survivants d'un groupe d'âge donné est comparée à la proportion de parents survivants du groupe d'âge supérieur, cinq (ou dix) ans plus tard. En supposant que la migration et la mortalité ne sont pas sélectives, le rapport entre les deux proportions reflète la mortalité

50. Plus généralement, on parle de cohorte hypothétique ou synthétique en démographie pour faire référence à une "section transversale de la population dont les diverses expériences démographiques (telles que les risques de mortalité) sont consolidées en indicateurs récapitulatifs, comme pour une véritable cohorte (naissances)" (Stover 2007).

des parents entre les deux opérations de collecte.

Par exemple, le Kenya a conduit une enquête EDS en 2003 (entre avril et septembre), quasiment cinq ans après l'enquête conduite en 1998 (entre février et juillet). La proportion de mères survivantes de la première EDS, calculé sur les enfants de 5 à 9 ans, peut être mise en relation avec la même proportion déclarée dans l'enquête qui a suivi. Notons ${}_5S_n(1)$ la première proportion [51], et ${}_5S_n(2)$ la seconde proportion. La proportion de mères qui survivraient au sein d'une cohorte synthétique pour ce même groupe d'âge est notée ${}_5\bar{S}_n$. C'est cette proportion "synthétique" que l'on cherche à calculer. Elle sera transformée comme précédemment en probabilité de survie en contrôlant la période d'exposition [52]. Rappelons-le, les probabilités de survie qui seront alors obtenues seront celles qu'une cohorte synthétique aurait connues si elle avait été exposée durant toute la vie adulte à la mortalité observée entre les deux opérations de collecte.

Pratiquement, cette proportion ${}_5\bar{S}_n$ s'obtient en calculant le rapport entre ${}_5S_n(2)$ et ${}_5S_{n-5}(1)$, puis en appliquant ce rapport à la proportion inférieure ${}_5\bar{S}_{n-5}$ (qui a été calculée auparavant) :

$$ {}_5\bar{S}_n =_5 \bar{S}_{n-5} \times \frac{{}_5S_n(2)}{{}_5S_{n-5}(1)} \tag{3.15} $$

La cohorte synthétique est donc reconstruite par *enchaînement des différentes proportions*, en commençant par ${}_5\bar{S}_0$ (qui s'obtient simplement en retenant ${}_5S_0(2)$). Le graphe 3.22 en présente une illustration en supposant deux opérations de collecte en 2004 et 1999.

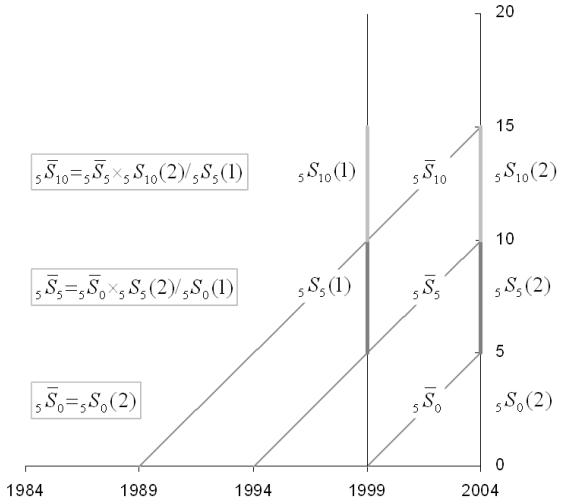

FIGURE 3.22 – *Exemple de l'enchaînement des proportions de non-orphelins par âge*

51. Qui est obtenue à partir des enquêtés de n à $n+4$

52. Par le biais d'une moyenne des âges moyens à la naissance des enfants enregistrés dans chacune des deux opérations de collecte.

Estimations basées sur les taux de croissance des populations de non-orphelins

La méthode précédente, basée sur l'enchaînement des proportions séparées d'un multiple de cinq ans, est assez simple mais pas des plus pratiques. En effet, elle recquiert que les opérations de collecte soient séparées par un multiple de cinq ans ou, à défaut, nécessite de recourir à des interpolations.

Une méthode alternative a été développée pour analyser de façon plus souple les variations que connaît la proportion d'enfants orphelins de chaque groupe d'âge entre deux opérations de collecte. C'est alors le *taux d'accroissement de la proportion de parents survivants* entre les deux opérations qui est à la base des calculs (Timaeus 1986, Preston et al. 2001). Comme précédemment, les estimations obtenues sont exemptes de l'effet des tendances, elles sont plus récentes, et les problèmes posés par les erreurs de déclaration sont réduits *si ces erreurs sont présentes dans la même proportion dans les deux opérations de collecte*.

Cette méthode alternative - que l'on appelera pour faire bref la méthode de Preston - repose sur le principe selon lequel, pour convertir la proportion de non-orphelins d'âge a en probabilité de survie, il est possible d'utiliser un facteur de correction, qui s'obtient en cumulant les taux de croissance des proportions de non-orphelins :

> "Stationary measures of parental survival can be obtained from lifetime data by adjusting the latter to remove the impact of past trends in mortality as reflected in the age-specific growth rates in the proportions of respondents with living parents" - Timaeus (1991b), p.214.

Pour en faire la démonstration, il faut préciser la relation qui existe entre (1) la proportion de non-orphelins, (2) le taux de croissance de la proportion de non-orphelins, et (3) la mortalité (Preston et al. 2001).

1. Dans une population stable, la population d'âge a, notée $n(a)$, peut se calculer à partir de la formule suivante : $n(a) = n(0)e^{-ra}{}_a p_0$. Si le taux d'accroissement r varie par âge, la population d'âge a observée il y a t années s'obtient comme :

$$n(a,t) = n(0,t) \times exp\left[-\int_0^a r(x,t)dx\right] \times {}_a p_0 \tag{3.16}$$

2. La probabilité ${}_a p_0$ peut être exprimée par le biais du taux instantané de mortalité $\mu(x,t)$, ce qui donne :

$$\begin{aligned} n(a,t) &= n(0,t) \times exp\left[-\int_0^a r(x,t)dx\right] \times exp\left[-\int_0^a \mu(x,t)dx\right] \\ n(a,t) &= n(0,t) \times exp\left[-\int_0^a (r(x,t) + \mu(x,t))dx\right] \end{aligned} \tag{3.17}$$

3. Toute forme de sortie (sortie par mortalité, migration, etc.) peut être introduite dans cette équation. Il faut simplement qu'une fonction

puisse saisir l'ampleur du phénomène qui provoque la sortie selon l'âge, et que les causes de sortie soient mutuellement exclusives (Preston et Coale 1982). Ainsi, si $\mu_M(x,t)$ renvoie au taux instantané de mortalité des mères des individus d'âge x, le nombre de non-orphelins d'âge a (au temps t) peut s'exprimer comme ceci :

$$n_{NO}(a,t) = n(0,t) \times exp\left[-\int_0^a (r_{NO}(x,t) + \mu(x,t) + \mu_M(x,t))dx\right] \tag{3.18}$$

Dans cette équation, la sortie de la population des non-orphelins est donc décomposée en deux facteurs : la *mortalité des parents* et la *mortalité des enfants*. Le taux de croissance par âge $r_{NO}(x,t)$ est celui des non-orphelins ; il est distinct de $r(x,t)$. Par contre, le taux instantané de mortalité $\mu(x,t)$ s'applique à tous les enfants, orphelins ou non. Sans cela, si les orphelins font face à une surmortalité, un biais de sélection est introduit (et biaise les estimations de mortalité à la baisse).

4. Rapporter le nombre de non-orphelins au nombre total de la population revient à rapporter les différents termes de l'équation 3.18 sur ceux de l'équation 3.17. La proportion de non-orphelins dans la population devient donc :

$$\begin{aligned}\frac{n_{NO}(a,t)}{n(a,t)} &= \frac{n(0,t) \times exp\left[-\int_0^a (r_{NO}(x,t) + \mu(x,t) + \mu_M(x,t))dx\right]}{n(0,t) \times exp\left[-\int_0^a (r(x,t) + \mu(x,t))dx\right]} \\ \frac{n_{NO}(a,t)}{n(a,t)} &= exp\left[-\int_0^a (r_{NO}(x,t) + \mu_M(x,t) - r(x,t))dx\right]\end{aligned} \tag{3.19}$$

5. Comme la différence entre le taux d'accroissement de la population de personnes dont les parents survivent et le taux d'accroissement de la population totale ($r_{NO}(x,t) - r(x,t)$) correspond au taux d'accroissement de la proportion d'orphelins (noté $r_\pi(x,t)$), l'équation qu'il s'agit d'exprimer sous forme discrète est :

$$\frac{n_{NO}(a,t)}{n(a,t)} exp\left[\int_0^a r_\pi(x,t)dx\right] = exp\left[-\int_0^a \mu_M(x,t)dx\right] \tag{3.20}$$

Cette expression met donc bien en relation (1) la *proportion de non orphelins*, (2) le *taux de croissance* de cette proportion, et (3) l'*intensité de la mortalité* des parents (sur une période qui s'étend de la naissance de leurs enfants au moment où ils atteignent l'âge a). Ce que nous apprend cette expression, c'est qu'une proportion "ajustée" de non-orphelins peut être obtenue en multipliant la proportion observée entre deux opérations de collecte[53], par un facteur de correction qui s'obtient en cumulant les taux

53. La proportion de non-orphelins pour un âge donné entre les deux opérations (notée ${}_5\bar{S}_x$) peut être obtenue en prenant la moyenne géométrique des proportions correspondante des deux opérations de collecte : ${}_5\bar{S}_x) = \sqrt{{}_5S_x(t) \times {}_5S_x(t+h)}$, si ${}_5S_x(t)$ est la proportion estimée à partir de la première et ${}_5S_x(t+h)$ à partir de la seconde.

de croissance des proportions de non-orphelins [54]. Sous forme discrète, ce facteur s'obtient comme ceci :

$$exp\left[\int_0^a r_\pi(x,t)dx\right] \approx exp\left[5\sum_0^{x+2.5}(r_\pi)\right] \tag{3.21}$$

La proportion ainsi ajustée peut être transformée comme précédemment en indices de mortalité. A la différence de la méthode précédente, les données sont combinées par groupes d'âge, et non par cohortes, ce qui est un avantage dans les pays où les déclarations sur les âges ne sont pas précises (Timaeus 1986).

Timaeus (1986) a appliqué cette méthode aux recensements du Kenya de 1969 et 1979, ce qui procure un niveau de mortalité plus élevé que ceux obtenus classiquement à partir des deux recensements pris individuellement. On peut l'observer sur le graphe 3.23 (panel du haut, point A). Nous y avons ajouté la combinaison des recensements kenyans de 1979 et 1989 (point B). Le dernier point (C), issu de la combinaison des recensements de 1989 et 1999, procure un estimation légèrement supérieure aux deux autres.

En Ouganda, en Tanzanie et au Rwanda, l'application de cette méthode aux données censitaires fournit des niveaux de mortalité qui cadrent avec ceux estimés rétrospectivement (résultats non présentés ici). Au Mali, l'estimation obtenue en combinant les recensements de 1987 et de 1998 poursuit la tendance à la baisse de façon cohérente (graphe 3.23 - panel du bas, point D). Par contre, la surprise vient de l'Afrique du Sud : le niveau estimé en combinant les recensements de 2001 et 2007 (point F) cadre avec les tendances précédentes, mais celui obtenu à partir des recensements de 1996 et 2001 est nettement inférieur (point E). Il pourrait s'agir d'une distorsion introduite par le VIH-sida.

Le chapitre 7, consacré à l'examen des niveaux de mortalité observés au niveau national, fournit d'autres illustrations de cette méthode, notamment à partir des enquêtes EDS. Retenons ici qu'elle est selon nous la méthode qui va fournir les estimations les plus plausibles, pour les différentes raisons évoquées plus haut.

3.4 Survie des parents aux âges adultes

Les deux méthodes basées sur la construction de cohortes synthétique présentent l'intérêt de fournir des estimations *récentes*, qui ne doivent pas être *datées*, puisqu'elles font référence à la période qui sépare les deux opérations de collecte. Elles annulent également en partie les erreurs de déclarations, si celles-ci sont présentes dans la même proportion d'une opération de collecte à l'autre. Il reste qu'elles font toutes deux intervenir des *proportions d'orphelins calculées à partir des premiers groupes d'âge*. Or ces informations sont celles qui sont les plus affectées par l'effet d'adoption.

54. Le taux se calcule comme : $r_\pi(\bar{t}) = ln(\frac{{}_5S_x(t)}{{}_5S_x(t+h)})/h$, où h est l'intervalle intercensitaire.

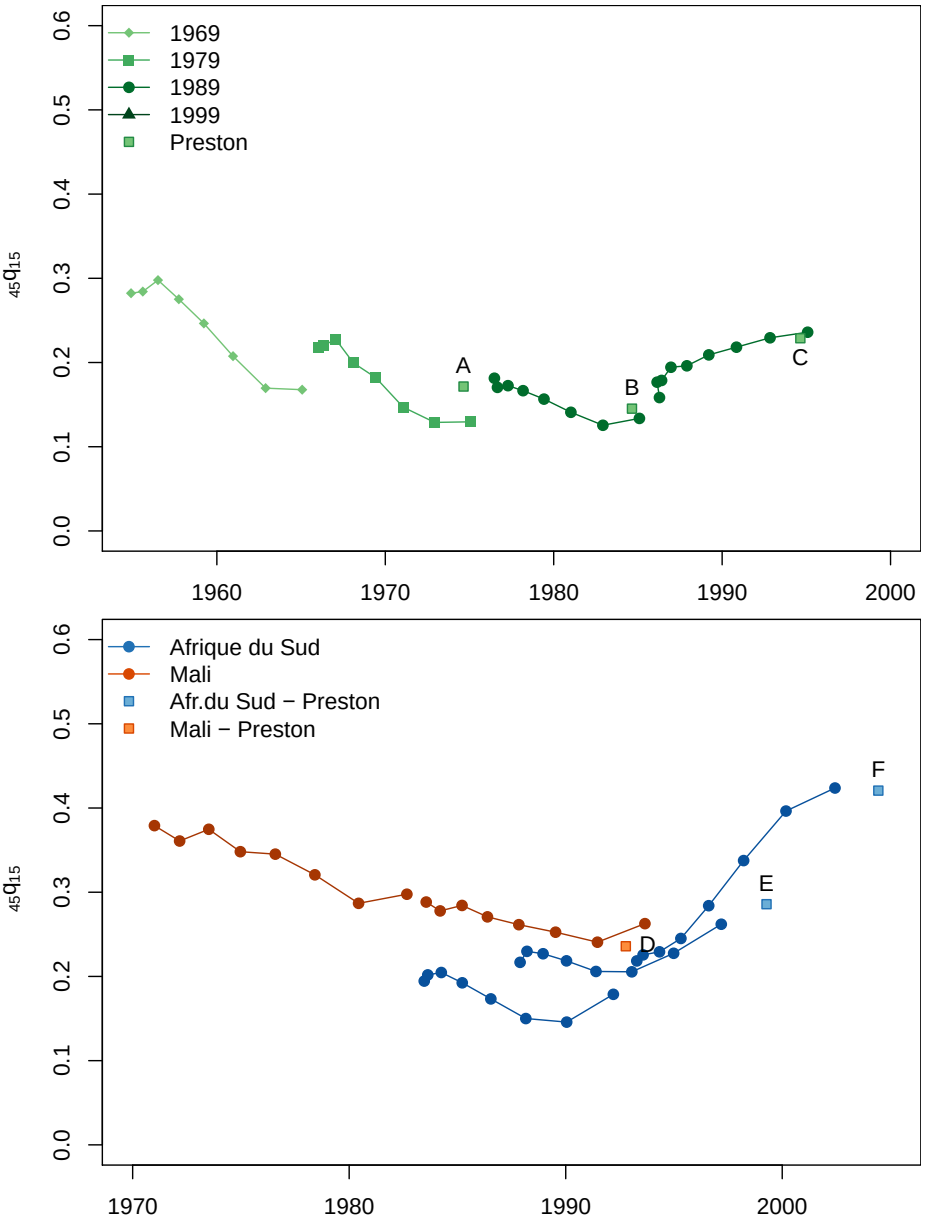

FIGURE 3.23 – *Estimations de $_{45}q_{15}$: nos calculs à partir des proportions de non-orphelins de mères observées au Kenya (1969-1979-1989-1999), Afrique du Sud (1996, 2001, 2007) et Mali (1987, 1998), méthode classique de datation (les traits) et estimations basées sur les taux de croissance des populations de non-orphelins (les points).*

En combinant les proportions de parents survivants d'un groupe d'âge à l'autre [55], ces méthodes vont transmettre une partie de ce biais d'adoption vers les âges plus élevés.

Deux alternatives ont donc été développées pour estimer la mortalité uniquement à partir de la survie aux âges adultes. C'est à nouveau Ian Timaeus qui est à leur initiative.

3.4.1 Survie des parents à 20 ans

Plutôt que de faire intervenir tous les groupes d'âge dans les calculs, Timaeus (1991b) a développé une méthode qui consiste, comme précédemment, à construire une cohorte synthétique à partir de données de deux dates distinctes, mais uniquement en sélectionnant les répondants *qui avaient encore leurs parents survivants lorsqu'ils avaient vingt ans* [56]. L'objectif est d'obtenir des estimations récentes, mais sans les erreurs de déclarations faites par les plus jeunes enquêtés.

Les proportions à la base des estimations sont donc les proportions de non-orphelins *parmi ceux dont la mère (ou le père) vivait lorsqu'ils avaient vingt ans*. Il s'agit en somme d'une extension de la méthode des cohortes synthétiques depuis la naissance. Reprenons l'équation 3.18, qui met en relation (1) la proportion de non-orphelins, (2) le taux de croissance de cette proportion, et (3) l'intensité de la mortalité des parents :

$$\frac{n_{NO}(a,t)}{n(a,t)} exp\left[\int_0^a r_\pi(x,t)dx\right] = exp\left[-\int_0^a \mu_M(x,t)dx\right]$$

Ces trois éléments peuvent être agencés de la même façon pour les individus de plus de 20 ans. Si $a > 20$, on peut écrire :

$$\frac{n_{NO}(a,t)}{n(a,t)} exp\left[\int_{20}^a r_\pi(x,t)dx\right] = \frac{n_{NO}(20,t)}{n(20,t)} exp\left[-\int_{20}^a \mu_M(x,t)dx\right] \quad (3.22)$$

En divisant les deux termes de cette équation par la valeur de $S(20,t)$, qui correspond à la proportion de parents survivants à 20 ans, puis en traduisant cette équation sous forme discrète, on obtient :

$$\frac{{}_5S_x(\bar{t})}{S(20,\bar{t})} exp\left[5\sum_{20}^{x+2.5}(r_\pi)\right] = \frac{{}_5S_x(\tau)}{S(20,\tau)} \quad (3.23)$$

Dans cette expression, la proportion de non-orphelins ${}_5S_x(\bar{t})$ est celle qui est observée entre les deux recensements [57], tandis que la proportion

55. C'est-à-dire en enchaînant les changements survenus au sein des cohortes (avec la méthode de Zlotnik et Hill (1981)) ou en combinant les taux de croissance des non-orphelins (avec la méthode de Preston (1983)).

56. La limite de 20 ans est retenue pour deux raisons :

1. limiter la sous-déclaration de la perte des parents chez les plus jeunes,
2. pouvoir travailler avec les quelques enquêtes EDS où la survie des parents a été posée aux femmes de 15 à 59 ans.

57. ${}_5S_x(\bar{t}) = \sqrt{{}_5S_x(t) \times {}_5S_x(t+h)}$

${}_5S_x(\tau)$ est celle qui correspond aux taux de fécondité et de mortalité de la population stable équivalente. Comme précédemment, il est possible de passer de l'un à l'autre, par le biais d'un facteur d'ajustement calculé en cumulant les taux de croissance des proportions de non-orphelins (de plus de 20 ans). Ce facteur est multiplié par les proportions de non-orphelins aux âges adultes observées entre les deux recensements $({}_5S_x(\bar{t})/S(20,\bar{t}))$ [58] et ce produit est ensuite converti en probabilités de survie.

En recourant aux mêmes simulations que celles détaillées plus haut (p.111), Timaeus (1991b) a pu calculer des coefficients de conversion à partir d'un modèle de régression en contrôlant l'âge moyen à la naissance des enfants [59]. Dans ce cas-ci, l'exposition au risque commence environ 20 ans après l'âge moyen à la naissance des enfants. L'équation de base est donc revue en conséquence et les probabilités de survie sont calculées à partir de 45 ans :

$$_{25+n}p_{45} = \beta_0(n) + \beta_1(n) \times M + \beta_2(n) \times \frac{{}_5S_{n-5}}{S(20,t)} \tag{3.24}$$

Les coefficients β sont présentés dans le tableau 3.10. Les coefficients utilisables pour la mortalité masculine sont repris en annexe A.3.2 [60].

n	β_0	β_1	β_2	R^2	CV
25	-0.8623	0.00292	1.7861	0.982	0.0037
30	-0.3822	0.00679	1.2062	0.983	0.0077
35	-0.4355	0.01197	1.1310	0.984	0.0177
40	-0.5995	0.01847	1.1419	0.987	0.0144
45	-0.7984	0.02547	1.1866	0.991	0.0146
50	-0.9360	0.03039	1.2226	0.992	0.0149

TABLE 3.10 – *Coefficients utilisés pour convertir les proportions de non-orphelins de mères (parmi les non-orphelins à 20 ans) en probabilités de survie féminines - Timaeus (1991b)*

Comme elles se rapportent uniquement aux âges adultes, les estimations obtenues sont épurées de l'effet d'adoption. Par contre, la méthode est plus sensible aux variations importantes dans le schéma par âge de la fécondité (Timaeus 1991b).

58. La proportion $S(20,\bar{t})$ observée entre deux recensements peut se calculer comme : $S(20,\bar{t}) = \sqrt{{}_5S_{15}(t) \times_5 S_{20}(t)}$

59. L'équation à la base des calculs (p.104) est légèrement modifiée, afin de ne retenir que les individus qui avaient toujours leurs mères en vie à 20 ans :

$$\frac{{}_5S_y}{S(20)} = \frac{\int_y^{y+5} e^{-r(a)} \times_a p_0\ S(a)\ da}{\int_y^{y+5} e^{-r(a)} \times_a p_0\ da\ S(20)}$$

60. L'équation masculine est légèrement différente. D'une part, l'âge moyen à la naissance des enfants est plus tardif, d'autre part la qualité de l'ajustement est meilleur quand un second groupe d'âge intervient dans les calculs :

$$_{35+n}p_{55} = \beta_0(n) + \beta_1(n) \times M + \beta_2(n) \times \frac{{}_5S_{n-5}}{S(20,t)} + \beta_3(n) \times \frac{{}_5S_n}{S(20,t)}$$

Dans la pratique, les niveaux obtenus sont assez proches de ceux calculés à partir de tous les enquêtés. Le tableau 3.11 compare les valeurs de ${}_{45}q_{15}$ calculées à partir des recensements africains utilisés précédemment, en construisant une cohorte synthétique depuis la naissance (colonne *d*, ${}_{45}q_{15} > 0$), ou depuis vingt ans uniquement (colonne *e*, ${}_{45}q_{15} > 20$)[61]. Contrairement à ce qu'on aurait pu penser, les niveaux de mortalité calculés à partir des âges adultes ne s'avèrent pas systématiquement supérieurs aux estimations obtenues par la méthode de Preston. Cette méthode alternative, basée sur les adultes uniquement, procure donc des niveaux de mortalité à comparer avec ceux obtenus à partir de tous les enquêtés.

	a	*b*	*c*	*d*	*e*
	Pays	t(1)	t(2)	${}_{45}q_{15}$ (> 0 ans)	${}_{45}q_{15}$ (> 20 ans)
1	Kenya	1969	1979	0.171	0.182
2	Kenya	1979	1989	0.145	0.146
3	Kenya	1989	1999	0.229	0.212
4	Mali	1987	1998	0.236	0.204
5	Tanzanie	1988	2002	0.254	0.248
6	Rwanda	1991	2002	0.482	0.447
7	Uganda	1991	2002	0.363	0.356
8	Afrique du Sud	1996	2001	0.286	0.284
9	Afrique du Sud	2001	2007	0.421	0.337

TABLE 3.11 – *Valeurs de ${}_{45}q_{15}$ obtenues par le biais des cohortes synthétiques (depuis la naissance et depuis 20 ans)*

	a	*b*	*c*	*d*	*e*
	Age	t(1)	t(2)	${}_{45}q_{15}$ (>0 ans)	${}_{45}q_{15}$ (>20 ans)
1	20-24	1988	2002	0.273	0.287
2	25-29	1988	2002	0.264	0.256
3	30-34	1988	2002	0.253	0.241
4	35-39	1988	2002	0.226	0.206
5	40-44	1988	2002	0.206	0.187
6	45-49	1988	2002	0.158	0.138

TABLE 3.12 – *Valeurs de ${}_{45}q_{15}$ obtenues par le biais des cohortes synthétiques (depuis la naissance et depuis 20 ans), par groupe d'âge - Tanzanie 1988 et 2002*

En fait, comme elle se base sur les enquêtés plus âgés, cette approche alternative est davantage biaisée par les erreurs sur les âges. L'effet de ces erreurs s'observe en comparant les valeurs obtenues à partir des différents groupes d'âge. Dans la mesure où les niveaux de mortalité obtenus à partir de ces groupes d'âge font référence à une même période, ils peuvent en effet être directement comparés. Dans la plupart des pays considérés ici, on constate *une baisse de la mortalité à mesure que les enquêtés avancent en âge*, ce qui reflète la tendance qu'ont les adultes plus âgés à se vieillir. C'est le

61. Dans les deux cas, les estimations retenues sont celles fournies par les 20-39 ans, mais le calcul des taux de croissance est limité aux plus de vingt ans pour produire les estimations de la seconde option.

cas des deux approches inter-censitaires considérées ici, mais peut-être de façon plus prononcée lorsque les cohortes synthétiques sont construites uniquement aux âges adultes. A titre illustratif, le tableau 3.12 compare les valeurs de ${}_{45}q_{15}$ obtenues par le biais des cohortes synthétiques (depuis la naissance et depuis 20 ans), par groupe d'âge. Les estimations sont issues des recensements de 1988 et 2002 en Tanzanie. On observe effectivement une baisse des niveaux de mortalité chez les répondants plus âgés, légèrement plus rapide dans le cas de la seconde approche.

3.4.2 Survie des parents au mariage

Si les problèmes de mémoire sont fréquents autour des dates de décès, il est fort probable qu'une majorité des individus mariés se souviennent avec précision si leurs parents étaient toujours en vie au moment de leur mariage. Avant de terminer cette section, mentionnons donc une méthode alternative, proche des méthodes précédentes, mais basée sur la survie des parents *depuis* le mariage et *avant* le mariage (Timaeus 1991a). L'idée consiste toujours à éliminer les déclarations faites par les plus jeunes répondants, tout en procurant des estimations récentes.

Cette méthode suppose qu'une question additionnelle soit posée sur la survie des parents au moment du premier mariage. En Afrique subsaharienne, cela limite son application aux quelques enquêtes EDS qui ont intégré cette question dans le questionnaire individuel (Sénégal 1986, Burundi 1987, Ghana et Ouganda 1988). A l'origine, cette question devait permettre d'analyser l'impact de la proximité avec la famille et la belle-famille sur les comportements de fécondité. Remarquons que ces questions ont été posées aux femmes *mariées* uniquement (sauf au Burundi). Les estimations de mortalité seront donc biaisées vers le haut pour les premiers groupes d'âge si les femmes se marient plus jeunes dans les ménages à plus forte mortalité.

Pour les détails techniques de l'estimation, nous renvoyons à l'article de Timaeus (1991a). Précisons simplement que les coefficients sont obtenus comme précédemment, au moyen d'une série de simulations. Les proportions de parents survivants sont régressées sur les probabilités de survie, en contrôlant à la fois l'âge moyen à la naissance des enfants et l'âge moyen au premier mariage. La primo-nuptialité est modélisée par le biais du schéma de Coale et McNeil (1972).

Cette méthode repose sur des hypothèses relativement contraignantes. Notamment, *l'âge au mariage des individus est supposé ne pas être associé avec l'âge des parents au moment de leur naissance*. La *probabilité de se marier est également supposée indépendante de la survie des parents*. Ces hypothèses ne sont pas toujours rencontrées en pratique. Beegle et Krutikova (2008) ont ainsi montré, à partir de données longitudinales, que la mort du père *accélérait* l'entrée en union des filles en Tanzanie (quand ces filles n'étaient pas scolarisées au moment du décès). C'est particulièrement le cas dans les ménages agricoles et dans les ménages plus pauvres. Les auteurs ne trouvent par contre pas d'effet significatif de la perte d'un parent sur

l'entrée en union des garçons.

A partir des données de recensements, nous pouvons comparer les proportions de non-orphelins par âge selon le statut matrimonial, afin de cerner si les orphelins se marient effectivement plus tôt. Cette comparaison est plus robuste qu'à partir des données d'enquête, dans lesquelles les effectifs de jeunes mariés sont faibles en dessous de 20 ans.

Le graphe 3.24 présente simplement l'évolution par âge de la proportion de non-orphelins *de mère*, à partir du recensement mené en Tanzanie en 1988 (panel du haut) et en 2002 (panel du bas), selon le statut matrimonial[62]. Les graphes de gauche font référence aux garçons et les graphes de droite aux filles. On observe qu'à un âge donné, la proportion de non-orphelins est nettement plus faible chez les filles qui sont déjà mariées. C'est le cas en 1988 comme en 2002, et pour tous les âges entre 10 et 50 ans. Par contre, chez les garçons, les proportions de non-orphelins ne varient pas sensiblement avec le statut matrimonial, sauf en-deçà de 25 ans.

On peut faire les mêmes constats en Ouganda (graphe 3.25), à partir des recensements de 1991 et 2002 : le décès des mères ne semble pas avoir d'impact majeur sur la mise en union des garçons (sauf en-deçà de 25 ans chez les garçons en 1991), mais les filles qui sont orphelines de mère semblent entrer en union un peu plus tôt que les autres[63]

Les graphes 3.26 et 3.27 présentent les mêmes proportions, mais pour la survie des *pères*, à partir du recensement tanzanien de 2002 et des recensements ougandais de 1991 et 2002. A nouveau, on compte davantage de filles déjà mariées chez les orphelins de père. Chez les garçons de moins de 30 ans également, les proportions de non-orphelins de pères sont inférieures parmi ceux qui sont déjà mariés, bien que les écarts avec les proportions de non-orphelins mariés soient moins prononcés que chez les filles.

62. Le mariage précoce et la forte mortalité des parents sont vraisemblablement concentrées sur certaines populations, et il faudrait donc idéalement mener une analyse plus fine en intégrant les différentes variables individuelles collectées dans les recensements.

63. Notons que l'on pourrait penser que ces écarts reflètent aussi partiellement des différences en termes d'âge moyen à la naissance des enfants. Mais nous ne le pensons pas. En effet, si l'on suppose que les parents des enfants qui se marient plus tôt se sont eux aussi mariés tôt, l'écart d'âge entre les jeunes *déjà mariés* et leurs parents devrait être plus faible que le même écart d'âge entre les jeunes *non mariés* et leurs parents. Or, plus l'âge moyen est faible, plus la proportion de parents survivants est élevée, puisque ces parents ne sont pas exposés à une mortalité à des âges très avancés. En conséquence, les différences dans l'âge moyen à la naissance des enfants tendent à *réduire* les écarts entre les proportions d'orphelins mariés et non-mariés.

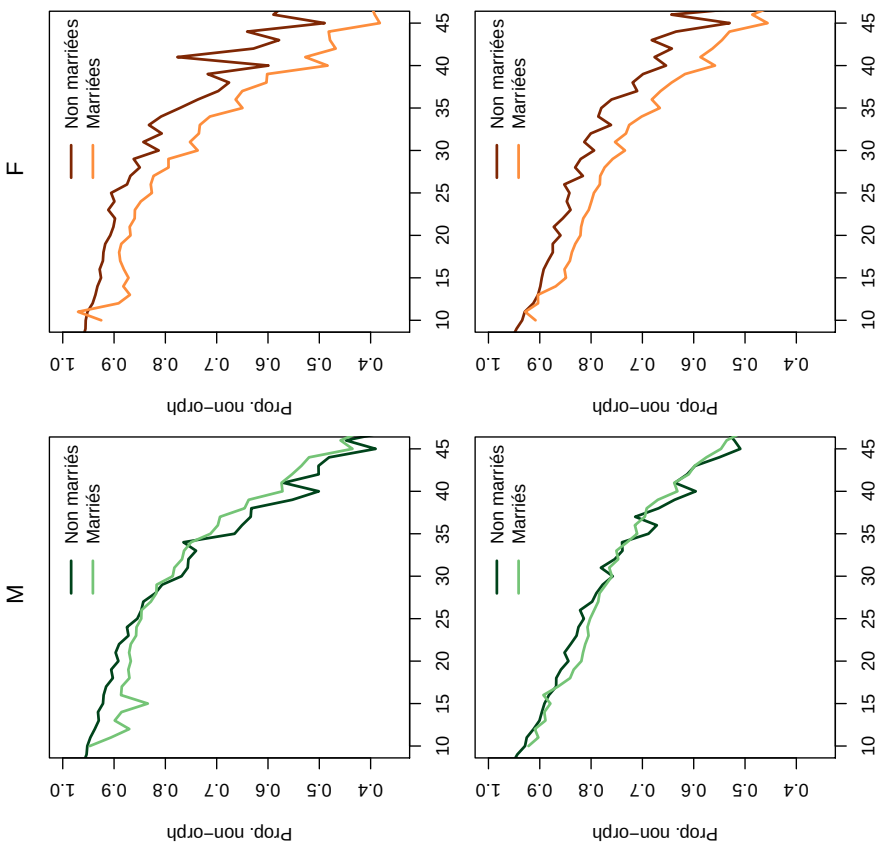

FIGURE 3.24 – *Proportion de non-orphelins de* ***mère*** *selon le statut matrimonial, Tanzanie 1988 et 2002*

FIGURE 3.25 – *Proportion de non-orphelins de* ***mère*** *selon le statut matrimonial, Ouganda 1991 et 2002*

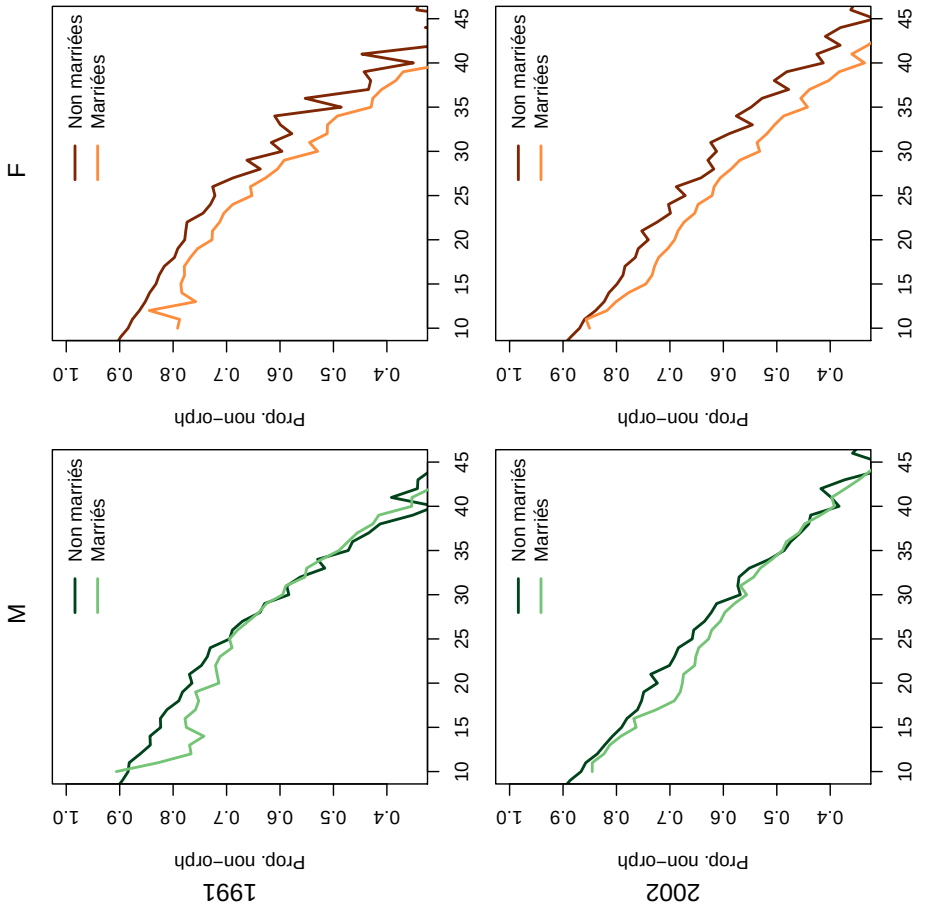

FIGURE 3.27 – *Proportion de non-orphelins de* ***père*** *selon le statut matrimonial, Ouganda 1991 et 2002*

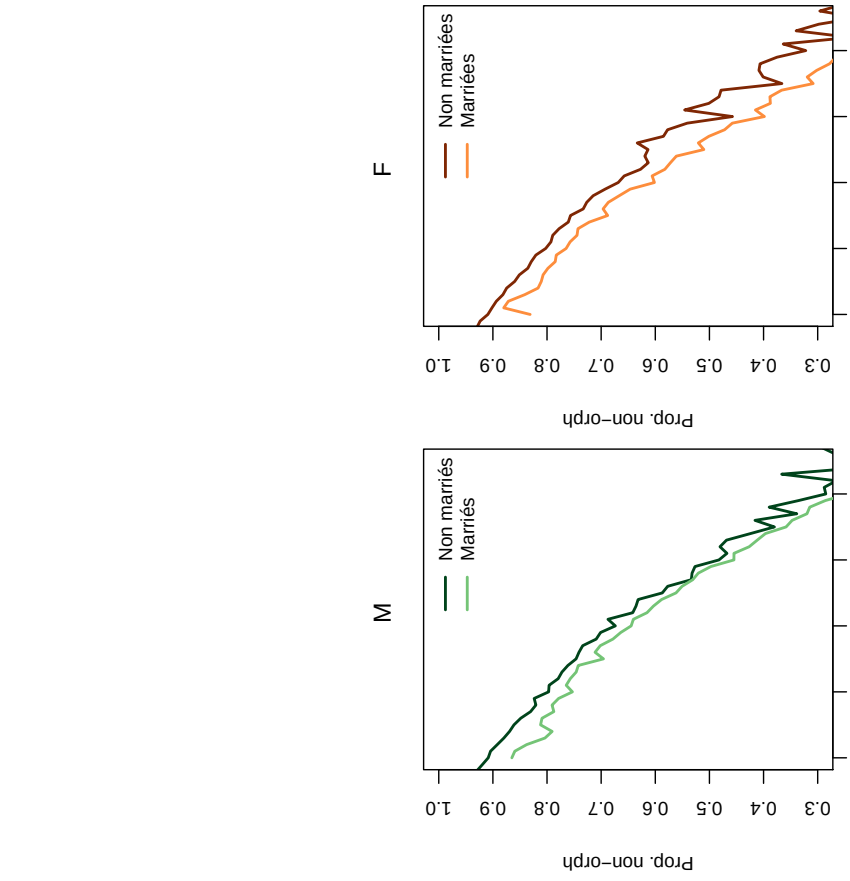

FIGURE 3.26 – *Proportion de non-orphelins de* ***père*** *selon le statut matrimonial, Tanzanie 2002*

3.5 Biais de sélection

Par rapport aux autres méthodes d'estimation de la mortalité adulte (survie intercensitaire, méthodes basées sur la distribution des décès par âge, etc.), les techniques basées sur la survie des proches sont particulièrement affectées par les biais de sélection. C'est par exemple le cas de la méthode basée sur la survie du premier conjoint, en raison de la forte corrélation entre les probabilités de décès des partenaires (notamment en raison du VIH). Mais c'est également le cas de la méthode des orphelins, en raison des différentes formes d'association entre la survie des enfants, la survie des mères, leur fécondité, etc.

Nous avons vu que la méthode des orphelins repose sur l'approximation de la *proportion de mères survivantes* par la *proportion d'enfants dont la mère est survivante*. Ceci suppose (1) que la survie des mères ne soit pas associée à la survie des enfants, (2) que la fécondité des mères ne soit pas associée à leur mortalité, et (3) que la taille des fratries ne soit pas associée à la mortalité des enfants.

Si ces conditions ne sont pas rencontrées, des biais sont introduits. Ceux-ci peuvent se traduire par des perturbations à la fois dans les *niveaux* de mortalité et dans les *tendances*. Nous examinons ici ces différents biais, en repartant du travail fait par Palloni et al. (1984). L'objectif est de présenter une certaine formalisation de ces biais, et de suggérer des ordres de grandeur. Nous abordons les problèmes liés au VIH-sida, mais ceux-ci font surtout l'objet du chapitre suivant.

3.5.1 Surmortalité des orphelins

Un premier biais de sélection survient lorsque les orphelins font face à une surmortalité. Ce biais est discuté dans les deux sections suivantes, à partir d'estimations empiriques puis en recourant à des microsimulations.

Estimations empiriques

Plusieurs recherches témoignent du fait que la situation des orphelins comporte des risques sanitaires importants, particulièrement en cas de décès de la mère. De nombreux facteurs de risques sont en cause. Les décès adultes entraînent souvent la déstabilisation du foyer familial, voire la dissolution du ménage (Urassa et al. 2001), ainsi qu'une plus grande mobilité des enfants (Ford et Hosegood 2005). La maladie et le décès d'un parent réduisent également le budget des ménages, en raison du coût des soins, des moindres rentrées, des coûts liés aux funérailles, et des rapports de dépendances plus élevés. Pour compenser, les orphelins peuvent être amenés à prendre une part plus active dans les tâches ménagères ou les activités génératrices de revenus [64], ou bien à prendre en charge leurs plus jeunes frères et sœurs, ou d'autres membres de la famille. Ces différents

64. Ceci n'est toutefois pas apparent dans les enquêtes EDS, du moins lorsque l'on examine les cas extrêmes de travail des enfants (plus de 40 heures semaines). Mais les enquêtes ménages sont sans doute peu appropriées pour mesurer ce phénomène (Monasch et Boerma 2004).

éléments sont associés à des coûts d'opportunité élevés, surtout en termes de scolarisation (Case et al. 2004, Ainsworth et Filmer 2006) [65].

Les différentes perturbations provoquées par la perte d'un parent peuvent également rendre les soins accordés aux enfants plus hasardeux, et compromettre leur nutrition. La perte d'une mère dans les premiers mois de la vie a des conséquences directes en termes de santé, puisqu'elle va provoquer un sevrage précoce, non exempt de risques infectieux et nutritionnels. Toutefois, lorsque l'on examine la prévalence de l'insuffisance pondérale dans les EDS, les orphelins ne semblent pas discriminés (Monasch et Boerma 2004). La morbidité et la malnutrition des orphelins n'apparaissent pas non plus comme leur étant systématiquement dommageables dans les observatoires de populations (Crampin et al. 2003).

Paradoxalement, plusieurs études menées dans des sites d'observation africains (en Ouganda, en Tanzanie) ont tout de même mis en évidence une *surmortalité significative des orphelins* (Nakiyingi et al. 2003, Ng'weshemi et al. 2003). Dans le cas de la mortalité des enfants de mères séropositives, il est toujours difficile de distinguer les cas de causes directes, liés à la transmission verticale, et indirectes. Il est toutefois clair qu'une surmortalité est observée même chez les orphelins de mère séronégative.

Pour fournir un ordre de grandeur de cette surmortalité des orphelins, nous pouvons mentionner trois analyses récentes, menées à partir de deux observatoires de population (Kisega et Masaka) et d'une enquête à passages répétées (Karongo) :

1. **Kisega** (Tanzanie) : L'analyse des données de cet observatoire révèle un gradient de mortalité important selon la survie et le statut sérologique de la mère des enfants. Entre 1994 et 2001, les probabilités de décès des enfants de moins de cinq ans s'élevaient à :
   - à 133 ‰ [119-148] chez les enfants de mère non infectée survivante,
   - à 248 ‰ [178-339] chez les enfants de mères infectées survivantes,
   - à 300 ‰ [185-464] chez les enfants de mères non infectées décédées,
   - et à 460 ‰ [234-759] chez les enfants de mères infectées décédées (Ng'weshemi et al. 2003).

   En contrôlant une série de facteurs (âge de la mère et de l'enfant, éducation, etc.), les auteurs concluent que l'excès de mortalité auquel font face les enfants dont la mère est *en phase terminale* (c'est-à-dire un an avant son décès) ou est *décédée depuis moins d'un an* se chiffre à 4.8 [2.7-8.4] chez les moins de 5 ans. Cette surmortalité due au décès de la mère est *indépendante de son statut sérologique*. La surmortalité des enfants de mères séropositives se chiffre à 2.3 [1.7-3.3].

65. Si l'on rapporte la proportion d'orphelins âgés de 10 à 14 ans qui vont à l'école à la proportion d'enfants du même âge qui vont à l'école (telles qu'observées dans les EDS), la médiane de ce ratio sur 34 enquêtes EDS récentes se chiffre à 0.87 (Monasch et Boerma 2004). Le désavantage des orphelins apparaît plus prononcé dans les pays où les taux de scolarisation sont plus bas, ainsi que chez les orphelins de mères et les doubles orphelins (Bicego et al. 2003).

2. **Masaka** (Ouganda) : Ces constats recoupent ceux faits dans un observatoire ougandais. Les deux ans qui entourent le décès de la mère sont également apparus comme étant associée à une forte surmortalité - avec un rapport de taux (Hazard ratio) estimé à 3.8. Cette surmortalité des orphelins est même supérieure à celle associée à la séropositivité des mères (HR : 3.2) (Nakiyingi et al. 2003).
3. **Karongo** (Malawi) : Par contre, une étude à passages répétés menée au Malawi n'a pas mis en évidence de surmortalité significative parmi les orphelins de mères séronégatives, ni de surmorbidité chez les enfants de mères séropositives. Les auteurs observent uniquement une surmortalité des enfants de mères séropositives, imputable à la transmission verticale (Crampin et al. 2003).

Zaba et al. (2005) ont analysé ces trois séries de données, et ils concluent que la surmortalité associée au fait d'avoir une mère séropositive s'élève à 2.9 [2.3-3.6] et que cet effet perdure durant l'enfance. La surmortalité associée au fait d'avoir une mère en phase terminale (un an avant son décès) ou décédée depuis moins d'un an se chiffre à 3.9 [2.8-5.5]. Ceci affecte tant les enfants de mères séropositives que de mères séronégatives. Les enfants de mères séropositives font donc face à deux facteurs de surmortalité.

Par ailleurs, Zaba et al. (2005) montrent que l'excès de mortalité lié au décès de la mère est sensiblement plus élevé aux tout premiers âges (moins de 3 ans). Il est plus élevé lors de l'année qui *suit* le décès maternel (par rapport à l'année qui le *précède*), mais la datation des décès reste sujette à caution.

Une enquête plus ancienne, menée en 1995 dans la région de Mwanza (au nord-ouest de la Tanzanie) a également documenté une surmortalité des orphelins (Bicego et al. 1997) [66]. Cette enquête assez exploratoire visait à estimer s'il était possible de corriger le biais qui est introduit par la surmortalité des orphelins dans l'estimation de la mortalité infanto-juvénile. Comme la mortalité des enfants est calculée classiquement à partir des histoires génésiques des femmes enquêtées, la mortalité des enfants de mère décédée échappe à l'échantillonnage. Cela biaise les estimations à la baisse. L'enquête, appelée SACM, a donc été conçue pour évaluer si l'histoire génésique des femmes décédées et la survie de leurs enfants pouvaient être collectées de façon fiable en interrogeant leurs sœurs.

Pratiquement, plus de deux milles femmes âgées de 15 à 50 ans ont été interrogées sur leur propre histoire génésique, sur la survie de leurs frères et sœurs ainsi que sur l'histoire génésique de chacune de leurs sœurs (ayant survécu à quinze ans). Les déclarations faites sur une même fratrie ont ensuite été partiellement appariées, ce qui a permis de juger de la qualité des données recueillies sur l'histoire génésique des sœurs. Cette enquête a également permis de calculer des taux de mortalité chez les enfants selon la survie des mères. Ce calcul met en évidence des différentiels importants de mortalité, et c'est ce qui nous intéresse ici.

Les niveaux de mortalité observés sont repris dans le tableau 3.13.

66. Il s'agit d'une enquête EDS spécifique : Sumve Survey on Adult and Childhood Mortality.

L'avant dernière colonne du tableau présente le rapport entre les probabilités de *décès* des orphelins et des non-orphelins (noté ${}_aq_0^{\text{OR}}/{}_aq_0^{\text{NON-OR}}$), et la dernière colonne présente la même information sous forme du rapport entre les probabilités de *survie*. Nous pouvons constater que la surmortalité des orphelins est particulièrement élevée, surtout avant un an (plus de trois fois la probabilité de décès des non-orphelins).

	Enfants des femmes enquêtées	Enfants des sœurs survivantes	Enfants des sœurs décédées	Ratio ${}_aq_0^{\text{OR}}/{}_aq_0^{\text{NON-OR}}$	Ratio ${}_ap_0^{\text{OR}}/{}_ap_0^{\text{NON-OR}}$
${}_1q_0$	89.6	85	260.2	3.1	0.81
${}_4q_1$	58.8	58.5	108.5	1.85	0.95
${}_5q_0$	143.2	138.5	340.4	0.41	0.77

TABLE 3.13 – *Mortalité des enfants des femmes enquêtées, des enfants de leurs sœurs survivantes et des enfants de leurs sœurs décédées (0 à 9 ans avant l'enquête) - Enquête SACM (1995)*

Ces chiffres doivent être interprétés avec prudence, car il semble que dans cette enquête, les décès des neveux et nièces ont été sensiblement sous-déclarés, surtout ceux survenus plus de cinq ans avant l'enquête. Ces décès sont également davantage entachés par des erreurs de déclaration (attirance sur les âges ronds, déplacements des décès dans le temps, etc.) [67].

Cela étant, même s'ils doivent être considérés comme des ordres de grandeur plutôt que comme des estimations précises, ces chiffres, ainsi que ceux issus plus récemment des observatoires de population en Afrique de l'Est, indiquent que les orphelins font face à une surmortalité prononcée. S'il est clair que cette surmortalité touche surtout les plus petits, il est encore difficile d'évaluer si elle varie en Afrique avec le sexe de l'enfant et celui du décédé [68].

67. Bicego et al. (1997) tirent une conclusion peu encourageante quant à la tentative d'estimer la mortalité des enfants à partir des neveux et nièces : "Proxy reporting of birth histories results in a significant underestimate in childhood mortality at all ages under 5 years. In this experiment, the bias is limited to the periods of more than 5 years before the survey, and is most problematic for the infant period, especially the postneonatal age segment (1-11 months). Consequently, estimates of both mortality trends and the age pattern of mortality are conspicuously distorted relative to the own-reported data. The fact that own and proxy estimates are nearly the same in the period 0-4 years before the survey is coincidental, being the result of a roughly proportionate underreporting of both surviving and non-surviving births"(p.52).

68. On peut citer à cet égard une analyse des risques de décès des enfants de moins de dix ans menée sur les données du site d'observation de Matlab au Bangladesh (Over et al. 1992). Il apparaît que le décès d'un parent affecte beaucoup plus les *jeunes filles* que les *jeunes garçons* ; les risques de décès d'un enfant orphelin de mère sont quasiment doublés lorsqu'il s'agit d'un fils et plus que triplés lorsqu'il s'agit d'une fille. C'est là une manifestation d'une forte préférence pour les garçons, combinée à la détérioration des soins accordés aux enfants en cas de décès parental. Il semble par ailleurs que la surmortalité des orphelins de mère soit plus importante que la surmortalité des orphelins de père. Les biais de sélection sont donc susceptibles de caractériser davantage les déclarations faites par les filles, mais également davantage les estimations de mortalité féminine.

Modélisation du biais de sélection

Palloni et al. (1984) montrent que le biais induit par la surmortalité des orphelins peut être estimé à partir de quelques hypothèses simples. Reprenons l'équation 3.2, qui exprime la proportion de non-orphelins d'âge a :

$$S(a) = \frac{\int_{\alpha}^{\beta} e^{-r(x+a)} \, {}_{x+a}p_0 \; m(x)dx}{\int_{\alpha}^{\beta} e^{-r(x+a)} \, {}_{x}p_0 \; m(x)dx}$$

Imaginons une première situation extrême où, en-deçà d'un certain âge v, les enfants décèdent systématiquement quand leurs mères décèdent. Ils connaissent par contre les mêmes risques de décès que les autres enfants si elle ne décède pas. Le dénominateur de l'équation 3.2 est modifié pour intégrer cette surmortalité. En exprimant la probabilité de survie ${}_{v}p_a$ sous forme continue [69] et en l'appliquant au dénominateur, on obtient :

$$S'(a) = \frac{\int_{\alpha}^{\beta} e^{-r(x+a)} \, {}_{x+a}p_0 \; m(x)dx}{\int_{\alpha}^{\beta} e^{-r(x+a)} \, {}_{x}p_0 \; \left[1 - \int_{0}^{v} {}_{z}p_a \; \mu(a+z)dz\right] m(x)dx} \tag{3.25}$$

En rapportant $S'(a)$ sur $S(a)$, on peut estimer l'ampleur du biais de sélection. Les deux numérateurs s'annulent, ce qui donne :

$$\frac{S'(a)}{S(a)} = \frac{\int_{\alpha}^{\beta} e^{-r(x+a)} \, {}_{x+a}p_0 \; m(x)dx}{\int_{\alpha}^{\beta} e^{-r(x+a)} \, {}_{x}p_0 \; \left[1 - \int_{0}^{v} {}_{z}p_a \; \mu(a+z)dz\right] m(x)dx} \tag{3.26}$$

Cette équation se simplifie comme ceci [70] :

$$\frac{S'(a)}{S(a)} = \frac{1}{1 - \mu(M + 0.5v) \times_v a_M} \tag{3.27}$$

Avec M, l'âge moyen à la naissance des enfants et
${}_{v}a_M$ le nombre attendu d'années passées dans l'intervalle $[M, M+v]$.

Cette expression peut être approximée par $1/{}_{v}p_M$. Cela signifie que le biais de sélection est directement proportionnel à l'intensité de la mortalité [71]. Comme tel, ce biais ne varie pas avec l'âge du répondant [72]. Il varie uniquement avec le niveau de mortalité, la valeur de M, et celle de v. Plus

69. Comme :

$${}_{v}p_a = 1 - \int_{0}^{v} {}_{z}p_a \; \mu(a+z)dz$$

70. Palloni et al. (1984) ne détaillent pas comment ils parviennent à cette simplification, mais nous pensons que la démarche est la suivante : en utilisant le théorème des accroissements finis, on sait qu'il existe une valeur (comprise entre les âges de reproduction α et β) telle que $\int_{\alpha}^{\beta} e^{-r(x+a)} \, {}_{x+a}p_0 \; m(x)dx$ puisse être factorisé au numérateur comme au dénominateur de l'équation 3.26. Ensuite, en exprimant ${}_{n}a_x$, le nombre d'années attendus passées dans l'intervalle $[x, x+n]$ sous forme continue (${}_{n}a_x = \int_{x}^{x+n} {}_{a}p_x da$), on parvient à l'équation 3.27.

71. A nouveau, Palloni et al. (1984) ne détaillent pas comment ils parviennent à cette approximation, mais nous pouvons établir les relations suivantes :

$$\mu(M+0.5) \times_v a_M = \mu(M+0.5) \int_{0}^{v} {}_{a}p_M da \approx \int_{0}^{v} {}_{a}p_M \; \mu(M+a)da =_v q_M$$

72. Pour des âges a inférieurs à v.

la mortalité est forte, plus le biais sera élevé. En outre, dans les situations où la mortalité a évolué dans le temps, non seulement les niveaux seront sous-estimés, mais les *tendances* seront également mal appréhendées. En cas de baisse de la mortalité, ce biais va induire une *sous-estimation plus conséquente dans les périodes plus anciennes* (puisqu'elles sont caractérisées par de plus fortes mortalités). L'ampleur des progrès est donc légèrement sous-estimée : la pente de la tendance est moins raide qu'en l'absence de ce biais de sélection. En cas de hausse de mortalité, c'est l'inverse : l'ampleur de la hausse est sous-estimée.

Dans l'ensemble, ce biais de sélection reste toutefois modeste. Palloni et al. (1984) estiment que l'erreur relative sur la proportion $S(a)$ s'élève au maximum à 4.3% environ, pour une valeur de v fixée à 5 ans et une mortalité élevée (correspondant à une espérance de vie de 45 ans).

Nous avons également procédé à l'estimation de ce biais à partir de la série de macro-simulations détaillée précédemment et nous obtenons des valeurs comparables. Le tableau 3.14 présente la moyenne des valeurs de ${}_1/_v p_M$ en fonction du niveau de mortalité (en faisant varier le paramètre α du système logit[73]). Le biais atteint 5.8% avec de fortes mortalités (e_0 = 36.4 et v = 5). Il est environ deux fois plus important quand v vaut 10.

α	e_0	$v = 5$	$v = 10$
-1	71.3	1.009	1.019
-0.6	62.0	1.018	1.038
-0.2	50.1	1.034	1.073
0.2	36.4	1.058	1.123

TABLE 3.14 – *Valeurs de ${}_1/_v p_M$ selon l'intensité de la mortalité - nos calculs à partir des simulations menées en retenant les paramètres du tableau 3.1, p.113*

L'ampleur de ce biais peut paraître importante, mais elle est surestimée ici, puisque nous avons supposé que *tous* les enfants orphelins décédaient s'ils perdaient leurs mères moins de v années après leur naissance. Palloni et al. (1984) montrent que cette hypothèse peut être facilement levée en intégrant dans l'équation 3.26 deux probabilités de survie différentes ; une pour les enfants qui perdent leur mère avant d'avoir atteint l'âge v (notée ${}_a p_0^{\text{OR}}$) et une pour les autres enfants (${}_a p_0^{\text{NON-OR}}$). Le biais de sélection s'en trouve sensiblement réduit. On peut le ré-écrire comme ceci :

$$\begin{aligned} \frac{S'(a)}{S(a)} &= \frac{1}{1 - \mu(M + 0.5v) \times_v a_M \times \left[1 - \frac{{}_a p_0^{\text{OR}}}{{}_a p_0^{\text{NON-OR}}}\right]} \\ \frac{S'(a)}{S(a)} &= \frac{1}{1 -_v q_M \times \left[1 - \frac{{}_a p_0^{\text{OR}}}{{}_a p_0^{\text{NON-OR}}}\right]} \end{aligned} \tag{3.28}$$

Dans ces conditions, le biais est logiquement plus faible que celui estimé précédemment ; le second terme du dénominateur est diminué par le facteur $[1 -_a p_0^{\text{OR}} /_a p_0^{\text{NON-OR}}]$. Plus les probabilités ${}_a p_0^{\text{OR}}$ et ${}_a p_0^{\text{NON-OR}}$ sont proches, plus le biais est faible. Il est nul si les deux probabilités sont égales. A la différence du cas précédent, ce biais-ci varie également avec

73. Les moyennes de ${}_1/_v p_M$ sont prises sur 24 valeurs différentes.

l'âge des enfants, et il sera plus faible pour les premiers âges [74].

a	α^{NON-OR}	α^{OR}	${}_ap_0^{NON-OR}$	${}_ap_0^{OR}$	${}_ap_0^{OR}/{}_ap_0^{NON-OR}$	$S(a)'/S(a)$
	(1)	(2)	(3)	(4)	(4)/(5)	
5	-1.0	-0.6	0.958	0.911	0.951	1.000
5	-1.0	-0.2	0.958	0.822	0.858	1.001
5	-1.0	0.2	0.958	0.676	0.706	1.003
5	-0.6	-0.2	0.911	0.822	0.902	1.002
5	-0.6	0.2	0.911	0.676	0.742	1.005
5	-0.2	0.2	0.822	0.676	0.822	1.006
10	-1.0	-0.6	0.954	0.903	0.947	1.001
10	-1.0	-0.2	0.954	0.808	0.847	1.003
10	-1.0	0.2	0.954	0.655	0.686	1.006
10	-0.6	-0.2	0.903	0.808	0.894	1.004
10	-0.6	0.2	0.903	0.655	0.725	1.010
10	**-0.2**	**0.2**	**0.808**	**0.655**	**0.810**	**1.013**

TABLE 3.15 – *Valeurs de $S(a)'/S(a)$ selon l'intensité de la sur-mortalité des orphelins - nos calculs à partir des simulations menées en retenant les paramètres du tableau 3.1, p.113*

Le tableau 3.15 détaille les valeurs du biais $S'(a)/S(a)$ que nous avons obtenues en prenant l'exemple d'une surmortalité des orphelins qui se traduit par une hausse du paramètre α. Les probabilités de survie de 0 à 10 ans et de 0 à 5 ans sont réduites en conséquence (*cf.* colonnes (3) et (4) du tableau). Pour différentes valeurs de α^{OR} et $\alpha^{\text{NON-OR}}$, le ratio entre les probabilités de survie varie entre 0.706 et 0.951 pour des enfants de 5 ans, et entre 0.686 et 0.947 pour des enfants de 10 ans. Ces valeurs sont à comparer avec l'estimation issue de l'enquête menée en Tanzanie, qui se chiffrait à 0.77 pour la survie de 0 à 5 ans (tableau 3.13).

Même pour des hausses de mortalité importantes, les biais restent modestes. Par exemple, une hausse du paramètre α de -0.2 à 0.2 se traduit par une baisse de la probabilité de survie jusqu'à 10 ans de 0.808 à 0.655, soit une baisse d'environ 20% (dernière ligne du tableau). Dans ce cas, le biais sur la proportion de mères survivantes $S(10)$ s'avère minime, puisqu'il ne s'élève qu'à 1.3% seulement. Cette valeur est à comparer avec celle de 7.3% obtenue dans le cas où tous les orphelins décèdaient lorsqu'ils perdaient leurs mères avant 10 ans (tableau 3.14). On voit donc ici qu'une modélisation plus réaliste de la surmortalité des orphelins se traduit par des valeurs nettement plus faibles du biais de sélection introduit.

74. Le biais dépend du ratio ${}_ap_0^{\text{OR}}/{}_ap_0^{\text{NON-OR}}$. Or ce ratio sera plus élevé aux premiers âges, et il va progressivement baisser, car il traduit le désavantage cumulatif des orphelins. En conséquence, le biais sera plus élevé pour les répondants plus âgés. Palloni et al. (1984) parviennent à la conclusion inverse à partir du même constat : "The upward errors [in $S(a)$] will decrease as the length of exposure increases provided that mortality has remained constant" (p. 273). Nous pensons pourtant que c'est l'inverse. Si le ratio ${}_ap_0^{\text{OR}}/{}_ap_0^{\text{NON-OR}}$ baisse effectivement quand augmente l'âge a, le biais $S'(a)/S(a)$ va augmenter avec l'âge, et non baisser. C'est ce qui ressort de l'équation 3.28. Ceci est d'ailleurs assez intuitif : si les orphelins font face à un cumul des désavantages, on trouvera relativement moins d'orphelins aux âges plus élevés en raison de cette sélection.

3.5.2 Association entre la survie des mères et des enfants

Les équations de la section précédente peuvent être généralisées en supposant que la population est divisée en différents sous-groupes, et que chaque sous-groupe i fait face à des risques de mortalité spécifiques. Les probabilités de survie par sous-groupes sont notées ${}_ap_0(i)$.

Dans ce cas, on peut exprimer la proportion de mères survivantes $S(a)$ (qui se réfère au niveau de la population d'ensemble) comme une somme pondérée des différentes valeurs de $S_i(a)$ (qui, elles, se réfèrent au niveau des sous-groupes). Les poids, notés $w_i(a)$, correspondent simplement à la distribution entre les sous-groupes des naissances survenues il y a a années :

$$\begin{aligned} w_i(a) &= \frac{\int_\alpha^\beta n_i(x,a)\ m_i(x)dx}{\sum_{i=1}^n \int_\alpha^\beta n_i(x,a) m_i(x)dx} \\ S(a) &= \frac{\sum_{i=1}^n S_i(a) \times w_i(a)}{\sum_{i=1}^n w_i(a)} \end{aligned} \tag{3.29}$$

Dans cette équation, $n_i(x,a)$ correspond au nombre de femmes du sous-groupe i qui ont donné naissance à l'âge x il y a a années (*cf.* p.103).

Lorsque les sous-groupes font face à des risques de mortalité spécifiques, ce qui sera observé, c'est une somme des valeurs de $S_i(a)$, pondérée à la fois par la répartition des naissances $w_i(a)$ et par les probabilités de survie ${}_ap_0(i)$ (Palloni et al. 1984) :

$$S'(a) = \frac{\sum_{i=1}^n S_i(a) \times w_i(a) \times_a p_0(i)}{\sum_{i=1}^n w_i(a) \times_a p_0(i)} \tag{3.30}$$

Généralement, les groupes où la mortalité des mères est plus élevée sont également ceux où la mortalité des enfants est plus élevée. Dès lors, là où la proportion de mères survivantes ($S_i(a)$) est élevée, la probabilité de survie ${}_ap_0(i)$ est élevée. En conséquence, la valeur de $S'(a)$ va excéder celle de $S(a)$. En d'autres mots, *les groupes à faible mortalité sont sur-représentés, ce qui biaise à la hausse la proportion de non-orphelins*.

Une application pratique sur le Kenya

Palloni et al. (1984) indiquent que cette sélection des enquêtés introduit un biais négligeable, mais ils ne fournissent pas d'ordre de grandeur du biais introduit. Blacker (1984) discute également de ce biais et propose une approche pour estimer l'ampleur de la sous-estimation de la mortalité. Il prend l'exemple des proportions de parents survivants observées dans le recensement kenyan de 1979, qui varient avec le niveau d'éducation des répondants. A partir du même recensement, il montre que les proportions d'enfants survivants varient également significativement avec le niveau d'éducation des mères. Il y a donc potentiellement une sous-estimation de la mortalité. Le biais introduit est toutefois minime. En effet, Blacker (1984) fait le constat suivant :

> "Large differentials are clearly apparent in both sets of data [les proportions de mères et d'enfants survivants]. However, it is comforting to note that the implied bias in the overall figures is, in this case at any rate, negligible. If one assumes that the respondents' birth cohort has been subject to a level of mortality commensurate with that indicated by the child survival data, it is possible to reconstruct the size of the original birth cohorts for each education group ; the proportions with mothers alive for each education group can then be reweighted by the number of births of which the respondents represent the survivors, rather than by the number of respondents themselves. The overall proportion with mother alive so obtained differs from the unweighted proportion by 0.2 per cent"' - Blacker (1984) , p. 86.

Son approche consiste donc à redonner davantage de poids aux groupes où la mortalité des enfants est la plus forte. Elle peut être appliquée aux enquêtes EDS. Nous en proposons ici une application à partir du cas du Kenya, qui connaît de larges inégalités ethniques de mortalité, aux âges adultes comme dans l'enfance (Brockerhoff et Hewett 2000, Masquelier 2008).

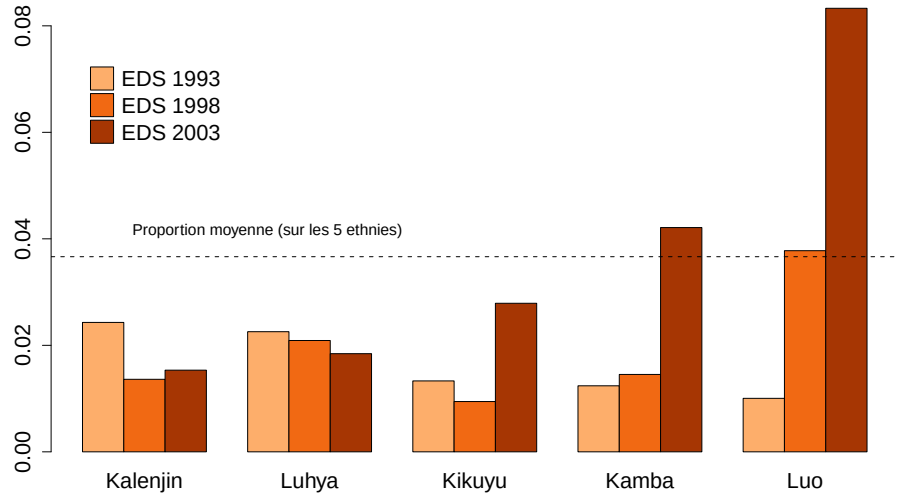

FIGURE 3.28 – *Proportions d'orphelins selon le groupe ethnique, EDS Kenya 1993, 1998 et 2003*

Le graphe 3.28 détaille les proportions d'orphelins de 5 à 9 ans selon le groupe ethnique des enfants[75], pour les cinq principaux groupes ethniques du Kenya (Kalenjins, Luhyas, Kikuyus, Kambas et Luos). Trois enquêtes EDS sont présentées : celles menée en 1993, en 1998 et en 2003[76]. En 2003, on observe une gradient très prononcé dans les proportions d'orphelins selon l'appartenance ethnique. Ces inégalités ethniques sont telles

75. Comme l'information sur l'appartenance ethnique n'est collectée que dans les questionnaires individuels, il s'agit plus précisément du groupe ethnique des adultes éligibles du ménage où se trouvent les enfants.

76. L'enquête EDS Kenya 2009 n'a pas collecté d'information sur les orphelins.

que la proportion d'orphelins de 5 à 9 ans observée parmi les Luos en 2003 est *plus de 5 fois supérieure à celle observée parmi les Kalenjins*. De façon un peu inattendue, on note que le gradient de mortalité des parents semble s'être inversé entre 1993 et 2003 ; les groupes les plus défavorisés en 1993 sont ceux où la proportion d'orphelins est la plus faible en 2003. Ceci est vraisemblablement lié à la crise du VIH-sida, dont l'impact a varié selon les différents groupes ethniques et les différentes zones géographiques. D'une enquête à l'autre, la hausse de mortalité adulte chez les Luos est particulièrement frappante. Les Kambas connaissent également une hausse de mortalité importante, tout comme les Kikuyus. Les Kalenjins et Luhyas se maintiennent globalement au même niveau.

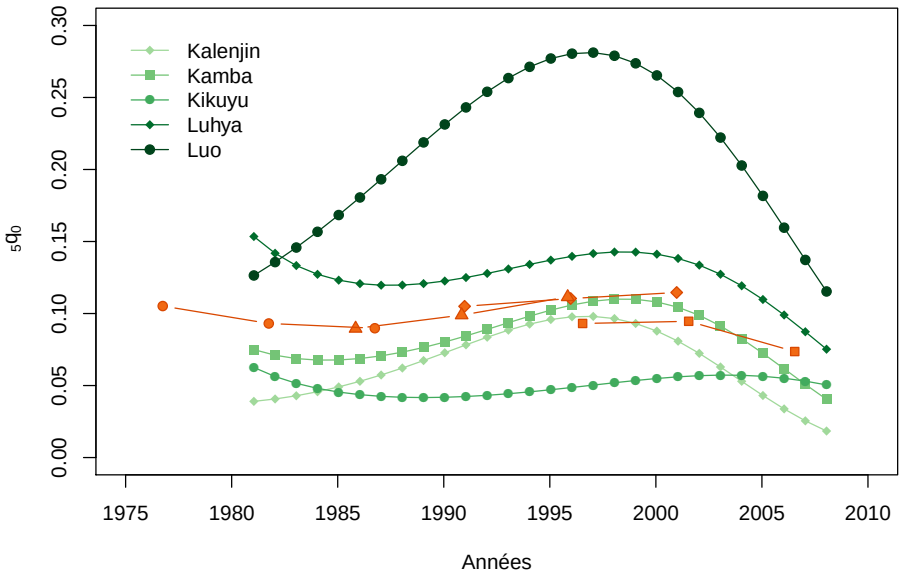

FIGURE 3.29 – *Tendances de la mortalité des moins de 5 ans selon le groupe ethnique, EDS Kenya 1989, 1993, 1998, 2003 et 2009*

Ces inégalités ethniques de mortalité adulte recoupent celles observées chez les enfants. Le graphe 3.29 présente les tendances de la mortalité des *moins de 5 ans* selon l'appartenance ethnique. Pour réduire l'importance de la variation aléatoire, nous avons estimé ces probabilités de décès en combinant cinq enquêtes EDS du Kenya (1989, 1993, 1998, 2003 et 2009). Nous travaillons en biographique (sur des fichiers personnes-périodes) en estimant les taux à partir d'une régression de Poisson (*cf.* chap. 5). Les tendances sont lissées au moyen de fonctions spline introduites comme variables explicatives dans la régression. Les différentes courbes en vert renvoient aux quotients de mortalité infanto-juvénile par ethnie, tandis que les points oranges correspondent aux estimations reprises dans les rapports d'enquête EDS, pour la population d'ensemble[77]. On note que les enfants Kikuyus, Kambas et Kalenjins bénéficient de la situation la plus favorable, sur globalement toute la période. Les Luhyas et les Luos font face à des risques de mortalité nettement plus élevés. C'est particu-

77. Pour une période de 0-4, 5-9 et 10-14 ans avant chaque enquête.

lièrement le cas des Luos, dont les risques de décéder avant l'âge de 5 ans ont connu une forte hausse jusqu'au milieu des années 1990 (ils ont quasiment triplé), avant de baisser. Ces probabilités de décès ${}_5q_0$ ont été plus de 5 fois supérieurs à ceux des Kikuyus durant toute la décennie 1990.

On est donc là face à des inégalités ethniques de mortalité chez les enfants qui défavorisent surtout les groupes ethniques où les proportions d'orphelins sont les plus élevées dans la seconde partie des années 1990 (les Luos, les Kambas, et dans une moindre mesure, les Luhyas). Ceci introduit un biais de sélection, mais son effet est négligeable sur les estimations de mortalité (au niveau de la population générale). Pour le calculer, nous avons opté pour la démarche de Blacker (1984). Nous retenons les proportions d'orphelins observés en 2003, puis nous repondérons le nombre d'enfants de chaque groupe ethnique par l'inverse de la probabilité de survie de la naissance à 10 ans[78]. Nous supposons que les enfants de 5 à 9 ans ont été exposés à une mortalité qui est inversément proportionnelle à cette probabilité de survie.

Ethnie	${}_{10}q_0$	$S_{5-9}(i)$	$w(i)_{10}p_0(i)$	$w(i)$	$S'(5-9)$	$S(5-9)$
Kalenjin	0.104	0.985	0.988	1241.684		
Kamba	0.114	0.958	0.980	1241.684		
Kikuyu	0.051	0.972	0.993	1241.684		
Luhya	0.151	0.982	0.992	1241.684		
Luo	0.284	0.917	0.989	1241.684		
Ensemble		**0.963**			**0.963**	**0.961**

TABLE 3.16 – *Proportions d'orphelins de 5 à 9 ans observées par ethnie au Kenya en 2003 et estimation du biais de sélection*

Le tableau 3.16 reprend les valeurs de la probabilité de décès de la naissance à 10 ans estimée pour 1996, les proportions d'orphelins de 5 à 9 ans calculées par ethnie pour 2003, ainsi que la répartition proportionnelle des enfants parmi ces cinq groupes - telle qu'elle est observée dans l'enquête ($w(i) \times_{10} p_0(i)$). A partir des probabilités de décès, nous recalculons la répartition proportionnelle des enfants de 5 à 9 ans *à leur naissance* ($w(i)$). Sur ces cinq groupes ethniques, la proportion de non-orphelins observée dans l'enquête se chiffre à 0.963. Cette valeur est à comparer avec celle "corrigée" pour tenir compte de la mortalité des enfants, soit 0.961. On le voit, cette correction ne change quasiment rien. La proportion non corrigée ne diffère de la proportion corrigée que de 0.2%, en dépit des différentiels de mortalité très prononcés.

Nous retombons donc sur une valeur de 0.2%, identique à celle estimée par Blacker (1984) (à partir du recensement kenyan de 1979). Cette valeur de 0.2% fournit-elle un ordre de grandeur du biais de sélection causé par les différentiels de mortalité ? Pouvons-nous généraliser le calcul de ce biais à partir des simulations ?

78. Estimée à partir du même modèle que les quotients repris dans le graphe 3.29. Nous retenons la valeur de ${}_{10}p_0$ estimée en 1996 (7 ans avant l'enquête), car c'est l'année où la variance des probabilités de survie est la plus forte (ce qui devrait maximiser le biais de sélection).

Estimation du biais de sélection à partir des simulations

Nous repartons ici de la série de simulations détaillée précédemment (*cf.* tableau 3.1, p.115). Pour rappel, cette série est construite à partir de 12 schémas de fécondité et de 8 schémas de mortalité. Pour chacun des 12 schémas de fécondité, nous supposons que les différents schémas de mortalité qui leur sont associés (n=8) renvoient à différents sous-groupes de population. Nous pouvons alors combiner ces différents sous-groupes entre eux comme s'ils formaient une population et estimer le biais introduit par les différentiels de mortalité entre ces groupes.

Pour chaque schéma de fécondité, si nous retenons toutes les combinaisons possibles[79], nous disposons de 247 combinaisons possibles ($C_8^2 + C_8^3 + ... + C_8^8$). Au total, il y a donc près de 3000 combinaisons différentes (12 schémas de mortalité $\times$ 247). Pour chacun des différents sous-groupes, la valeur de $w_i(a)$ est approximée numériquement (à partir de l'équation 3.1, p.103). Tous les groupes connaissent une population stable et un taux de croissance de 0.02. A partir de là, nous pouvons calculer les valeurs de la proportion de non-orphelins caractérisée par le biais de sélection. Nous retenons trois âges différents : $S'(5)$, $S'(20)$ et $S'(35)$. Ces différentes proportions peuvent ensuite être rapportées aux proportions $S(a)$ correspondantes. Le tableau 3.17 détaille les biais estimés. A nouveau, ils apparaissent assez minimes. A 5 ans, la médiane des biais s'élève à 0.1%, à 20 ans à 1% et à 35 ans à 2.9%.

	1er Qu.	Médiane	3e Qu.
5	1.001	1.001	1.002
20	1.006	1.010	1.013
35	1.018	1.029	1.039

TABLE 3.17 – *Valeurs de $S(a)'/S(a)$ obtenues en combinant différents schémas de mortalité (n = 2964)*

En définitive, *ces biais de sélection se traduisent par une sous-estimation de la mortalité qui reste modeste*. Ils augmentent - logiquement - avec la variance des probabilités ${}_ap_0(i)$, et ils augmentent également rapidement avec l'âge. *Cela signifie que les méthodes qui se concentrent sur les proportions de non-orphelins aux âges adultes sont davantage affectées par ce biais de sélection*.

3.5.3 Association entre mortalité des mères et fécondité

Enfin, le dernier type de biais de sélection envisagé ici est causé par l'association entre la mortalité et la fécondité. C'est encore à Palloni et al. (1984) qu'il revient d'avoir creusé cette question. Ils montrent notamment que les biais introduits ne vont pas systématiquement dans le sens d'une sous-estimation ou d'une sur-estimation de la mortalité. Les choses sont plus complexes, car deux phénomènes sont à l'oeuvre :

1. D'une part, *les mères qui décèdent plus tard auront plus d'enfants* : simplement parce que la durée d'exposition est prolongée. Ceci va causer une sous-estimation de la mortalité.

79. Avec un nombre de groupes variant de 2 à 8.

2. D'autre part, *l'intensité de la fécondité peut être associée à l'intensité de la mortalité*. C'est le cas par exemple si les populations plus favorisées contrôlent davantage leur fécondité. Ce point est bien documenté dans la littérature (Tabutin 2007). Si l'association est négative (si les femmes dont la mortalité est plus élevée ont moins d'enfants), la mortalité sera sur-estimée. Si elle est positive, c'est l'inverse.

Les biais introduits peuvent donc aller dans des sens différents. Ceci signifie qu'il peut y avoir une compensation partielle des différents biais en jeu.

Période d'exposition prolongée

Les mères qui survivent plus longtemps sont exposées plus longtemps au risque de grossesse, et elles auront donc en moyenne davantage d'enfants. Ces mères auront plus de chances d'être mentionnées à plusieurs reprises : par exemple, une première fois par un enquêté âgé de 5 à 9 ans, et une seconde fois par un enquêté âgé de 10 à 14 ans. Ce phénomène va induire une *sous-estimation* de la mortalité qui va augmenter à mesure qu'augmente l'âge des enquêtés [80]. En effet, parmi les mères d'une même cohorte, celles qui connaissent une forte mortalité sont de moins en moins représentées dans les déclarations faites par leurs enfants lorsque ceux-ci avancent en âge. En conséquence, à partir d'une seule enquête, si la mortalité a baissé, l'ampleur de cette baisse va être légèrement sous-estimée (et vice-versa si la mortalité a augmenté).

A nouveau, les biais introduits n'ont que des effets négligeables sur les estimations de mortalité. A partir de simulations, Palloni et al. (1984) indiquent que les erreurs introduites sur l'espérance de vie ne dépassent pas un an. De plus, ils font remarquer que cette sélection ne s'opère que sur une *fraction* seulement des mères. En effet, si certaines mères sont mentionnées à plusieurs reprises, c'est surtout par des enfants dont les groupes d'âges sont proches. Par exemple, il y a peu de mères qui sont susceptibles d'être mentionnées à la fois par des enfants de 25 à 29 ans et par des enfants de 5 à 9 ans. Donc, même si une sélection s'opère sur les mères ayant des enfants dans ces deux groupes d'âge, elle a un effet minime, car elle ne concerne qu'un petit nombre de mères. Et si l'on retient des groupes d'âge plus proches, par exemple 5-9 ans et 10-14 ans, certes il y a beaucoup plus de femmes qui sont susceptibles d'avoir des enfants dans ces deux groupes d'âge, mais la sélection n'a que peu de temps pour opérer. La représentation proportionnelle des mères selon leur niveau de mortalité n'est donc que peu modifiée entre les mères des enfants de 5 à 9 ans et les mères des enfants de 10 à 14 ans.

Intensité de la fécondité

La littérature sur les relations entre pauvreté et fécondité met généralement en évidence une plus forte fécondité parmi les populations moins

80. Notons toutefois que la sexualité et la fécondité sont généralement plus précoces dans les populations moins favorisées, où la mortalité est plus forte. Ceci va augmenter la période d'exposition des mères qui font face à de fortes mortalités, et donc atténuer ce biais à la baisse.

favorisées (Tabutin 2007). Ceci s'explique notamment par une pratique moins fréquente de la contraception, par des intervalles inter-génésiques plus courts (en raison notamment de plus fortes mortalité des enfants), ainsi que par une activité sexuelle plus précoce. Comme les mères les plus fécondes, qui sont souvent les plus pauvres, font également face à de plus fortes mortalités, elles ont proportionnellement plus d'enfants pour déclarer leur décès. Ceci induit donc une *sur-estimation* des risques de décès.

Cette sur-estimation est toutefois compensée par deux éléments déjà mentionnés : (1) la moins longue exposition au risque de grossesse chez les femmes à forte mortalité, et (2) la plus forte mortalité des enfants des mères qui sont soumis eux-mêmes à une plus forte mortalité. S'il est clair qu'il y a une compensation partielle entre ces différents phénomènes, il est toutefois difficile d'estimer l'effet "net" de ces différents biais.

3.6 Application aux populations simulées pour l'Afrique subsaharienne

Nous avons jusqu'ici détaillé la méthode des orphelins, ses variantes et ses biais de sélection de façon théorique. Nous cherchons maintenant à nous rapprocher du contexte réel de l'application de la méthode, dans les cas où de nombreuses perturbations entrent en jeu : des tendances perturbées en matière de mortalité, une fécondité en baisse, des schémas atypiques de mortalité en raison du VIH-sida, etc. Pour ce faire, nous appliquons la méthode des orphelins aux populations fictives issues de la série de microsimulations détaillée dans le chapitre précédent.

Dans un premier temps, nous présenterons les proportions d'orphelins par âge telles qu'elles peuvent être calculées à partir des populations simulées (pour les enfants de moins de 15 ans). Nous comparerons ces proportions issues des microsimulations avec celles observées dans les enquêtes EDS. Nous reviendrons ensuite sur la méthode des orphelins en elle-même.

3.6.1 Comparaison des proportions d'orphelins issues des simulations avec celles issues des enquêtes EDS

Pour calculer les proportions d'orphelins à partir des populations fictives, nous avons restructuré les fichiers de résultats sous forme de personnes-périodes en fonction des dates de naissances, du décès des enfants et du décès des mères. Les proportions d'orphelins sont ensuite calculées en rapportant, pour chaque année et chaque groupe d'âge, les durées d'exposition des enfants orphelins aux durées d'exposition vécues par tous les enfants. Les graphes 3.30 et 3.31 présentent, en traits pleins, l'évolution dans le temps *des proportions d'orphelins de mère* pour les groupes d'âge 0-4, 5-9 et 10-14 ans et dans 16 pays différents. Pour chacun des pays repris sur ces graphes, les proportions d'orphelins observées dans les enquêtes EDS sont présentées également (sous forme de carrés).

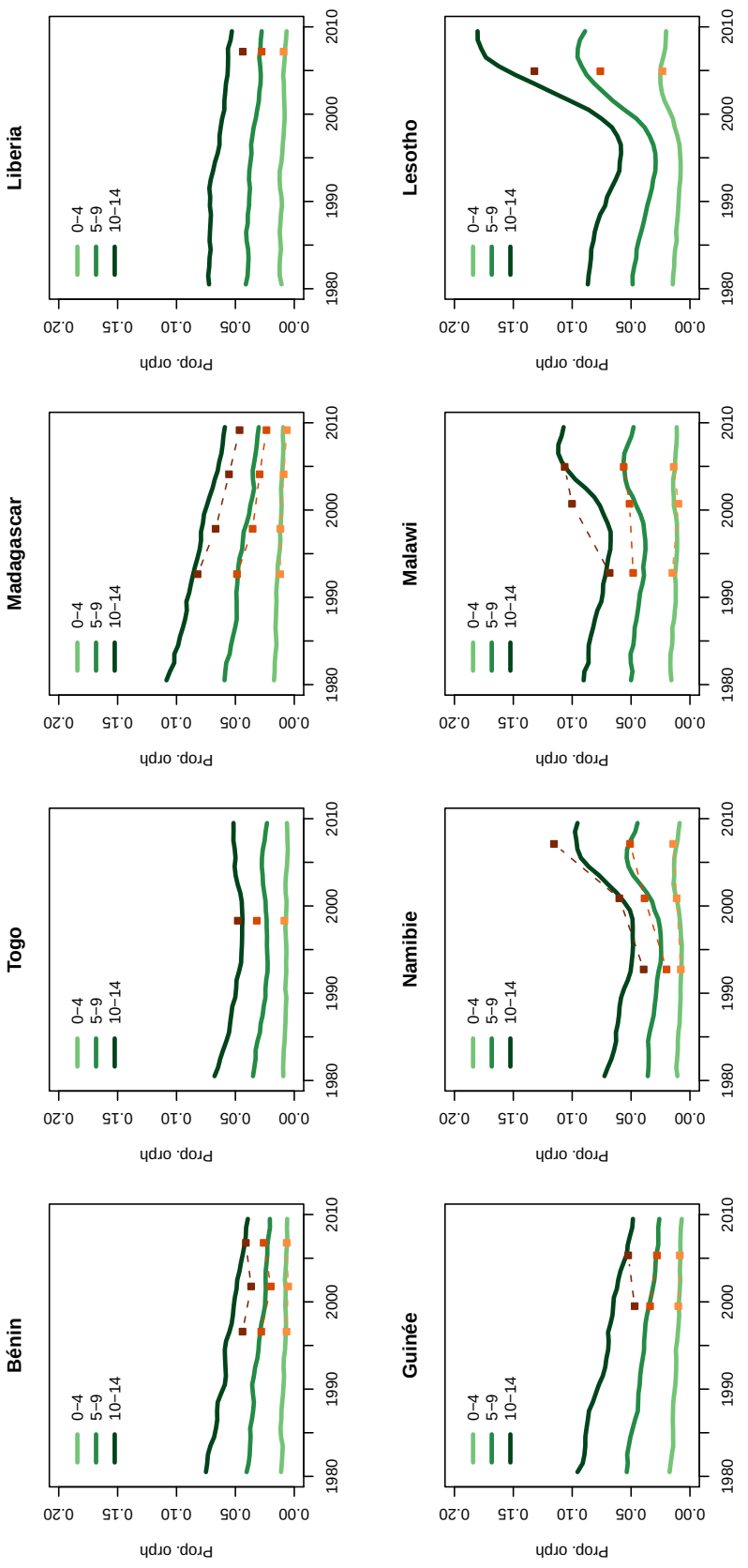

FIGURE 3.30 – *Proportions d'orphelins issues des microsimulations et observées dans les enquêtes EDS : Bénin, Togo, Madagascar, Liberia, Guinée, Namibie, Malawi, Lesotho.*
Note : les points correspondent aux proportions observées dans les enquêtes EDS et les courbes représentent les proportions d'orphelins calculées à partir de nos microsimulations.

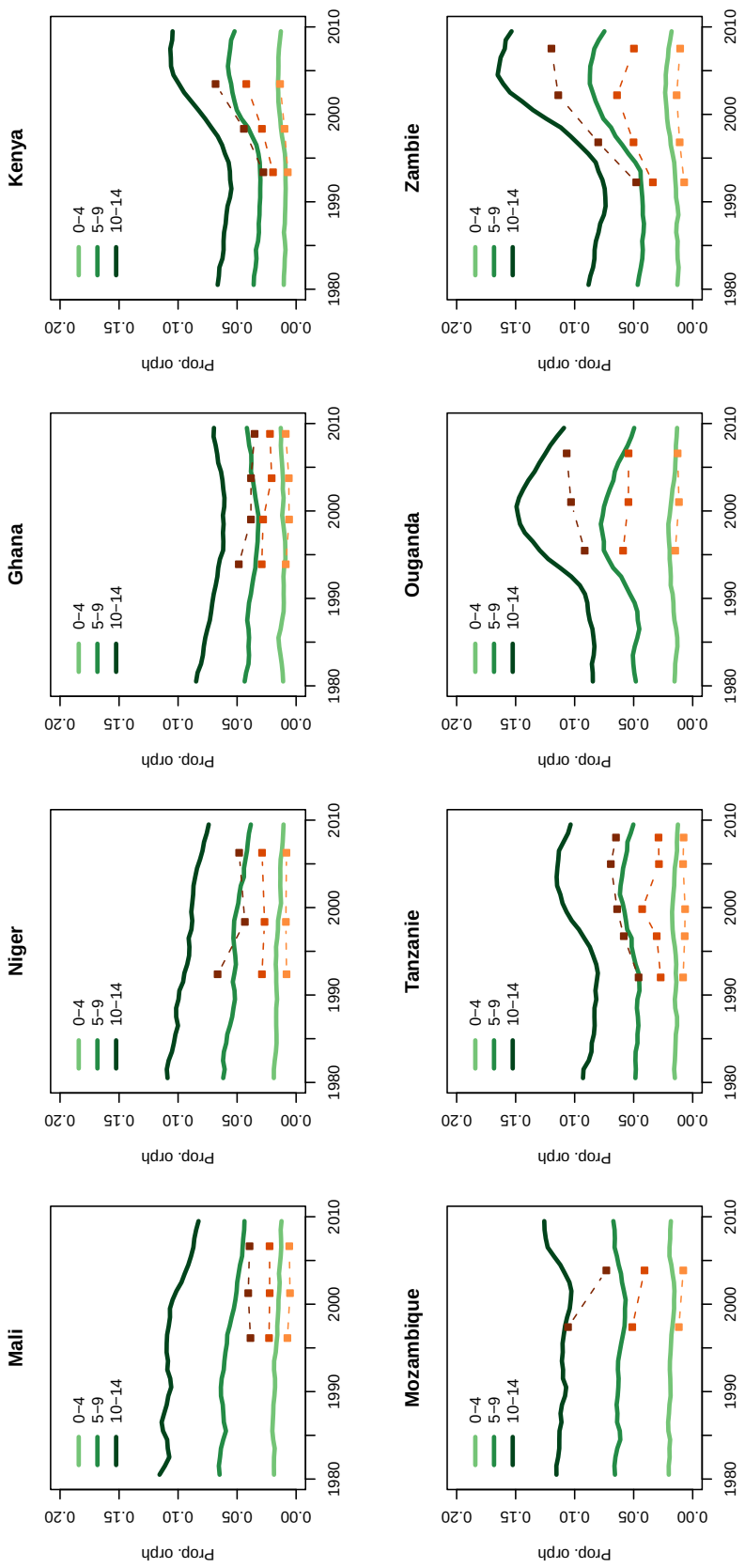

FIGURE 3.31 – *Proportions d'orphelins issues des microsimulations et observées dans les enquêtes EDS : Mali, Niger, Ghana, Kenya, Mozambique, Tanzanie, Ouganda, Zambie.*
Note : les points correspondent aux proportions observées dans les enquêtes EDS et les courbes représentent les proportions d'orphelins calculées à partir de nos microsimulations.

Dans la première série de pays (le graphe 3.30), les proportions d'orphelins 'empiriques' et 'simulées' sont relativement proches. Les tendances dégagées des différentes enquêtes cadrent également avec celles issues des microsimulations. C'est le cas notamment de la baisse de la proportion d'orphelins observée à Madagascar, ainsi que de la hausse de cette proportion en Namibie.

Sur l'ensemble des pays africains pour lesquels nous avons procédé à des simulations, les 8 pays présentés dans le graphe 3.30 font toutefois exception. Dans la majorité des cas (environ 4 pays sur 5), les proportions issues des enquêtes sont *moins élevées* que celles obtenues par simulation. On l'observe dans le graphe 3.31 pour huit pays : le Mali, le Niger, le Ghana, le Kenya, le Mozambique, la Tanzanie, l'Ouganda et la Zambie [81]. Si il y a désaccord sur les niveaux, il y a également parfois des divergences au niveau des tendances. Au Mozambique et en Ouganda, par exemple, la tendance esquissée par les différentes enquêtes ne cadre pas avec celle dégagée des simulations. L'apparente baisse observée à partir des deux enquêtées menées au Mozambique ne se retrouve pas dans les microsimulations. C'est même l'inverse en Ouganda ; les simulations indiquent une baisse des proportions d'orphelins dès la fin des années 1990, tandis que ces proportions semblent se maintenir, si l'on en croit les enquêtes. En revanche, en Zambie, si les niveaux sont différents, les tendances concordent.

Comment expliquer ces différences ? A priori, plusieurs éléments entre en jeu :

1. Il se peut que la *mortalité "sans sida"* prévue par les Nations Unies (qui est à la base de nos simulations) surestime les risques de décès aux âges adultes. Ceci se traduirait par des proportions d'orphelins 'simulées' trop élevées par rapport à celles observées dans les enquêtes.
2. Il se peut également que la façon avec laquelle nous avons *modélisé le VIH-sida* pose problème : comme la survie dépend de l'âge à l'infection dans les simulations (et pas dans les estimations de l'ONU) et que, par ailleurs, les femmes sont infectées plus tard, nous obtenons des risques de décès féminins souvent plus élevés que ceux estimés par les Nations Unies (*cf.* chapitre 2). Le rapport de l'incidence féminine sur l'incidence masculine que nous avons retenu pourrait également être trop élevé, et se traduire par un nombre trop élevé d'orphelins de mère.
3. Une troisième source de différences entre les deux séries de proportions, et qui est sans doute la plus importante, tient au fait que les proportions 'simulées' ne sont pas entachées d'*erreurs de déclarations*, ce qui est le cas des proportions empiriques. En particulier, le biais d'adoption va réduire les proportions d'orphelins dans les EDS, alors que les simulations en sont exemptes. Il se peut également qu'une attirance sur les âges ronds (5 et 10) déplace vers le

81. Dans plusieurs pays, les écarts entre les deux séries de valeurs sont tels que les proportions empiriques pour les 10-14 ans sont proches des proportions simulées pour le groupe des 5-9 ans (Tanzanie, Ghana, Kenya, etc.).

haut certains enfants, ce qui a pour effet de réduire encore davantage les proportions d'orphelins (*cf.* chapitre 6).

Afin de chercher à expliquer ces écarts, les graphes 3.32, 3.33 et 3.34 reprennent les proportions d'orphelins observées dans 75 enquêtes EDS et les comparent, pour chaque groupe d'âge, avec celles issues des microsimulations (pour la même année). Les estimations sont présentées selon la prévalence du VIH observée cinq ans avant l'enquête, car il se pourrait que les écarts se situent surtout au niveau de la modélisation du VIH-sida dans les simulations.

Que constate-t-on ? Dans les trois groupes d'âge, les *proportions simulées sont supérieures à celles issues des enquêtes*. La médiane du rapport entre les proportions simulées et les proportions empiriques se chiffre à 1.45 pour les enfants de 0 à 4 ans. Elle s'établit à 1.3 pour les 5-9 ans et à 1.42 pour les 10-14 ans.

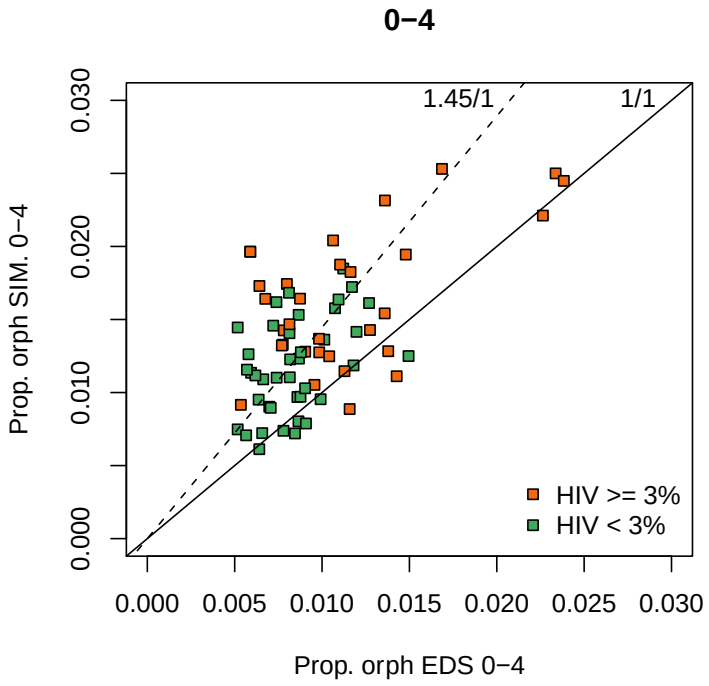

FIGURE 3.32 – *Proportions d'orphelins issues des microsimulations et observées dans les enquêtes EDS : 0-4 ans (n=75)*

Nous retrouvons ici une observation déjà faite par Grassly et al. (2004), qui ont comparé les proportions d'orphelins issues des enquêtes MICS et EDS avec celles calculées par l'UNICEF et l'UNAIDS. Ces dernières sont basées sur un modèle macro dans lequel interviennent les projections de population des Nations Unies, comme c'est le cas de nos microsimulations. Nous reviendrons sur ces projections de l'UNICEF dans le chapitre suivant.

Grassly et al. (2004) notent que les proportions empiriques s'avèrent inférieures aux proportions issues de la modélisation. Sur l'ensemble des enquêtes analysées par ces auteurs (40 EDS et MICS), les proportions

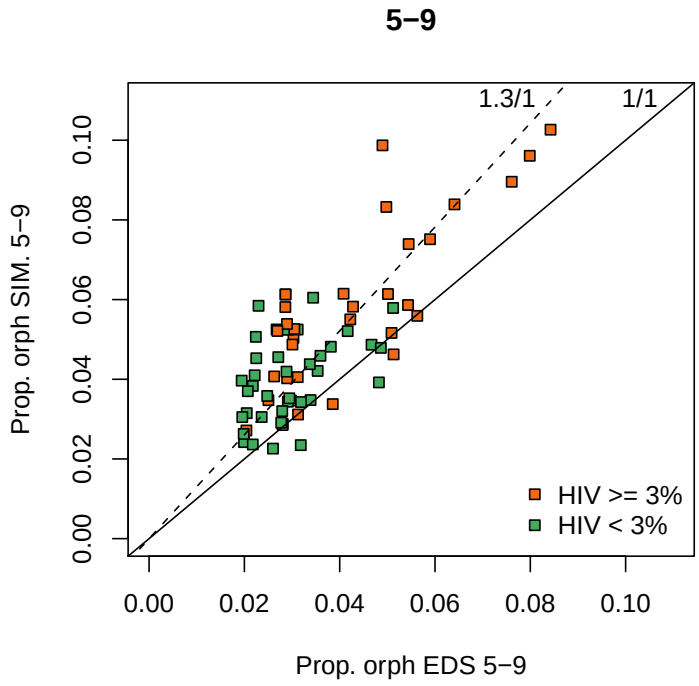

FIGURE 3.33 – *Proportions d'orphelins issues des microsimulations et observées dans les enquêtes EDS : 5-9 ans (n= 75)*

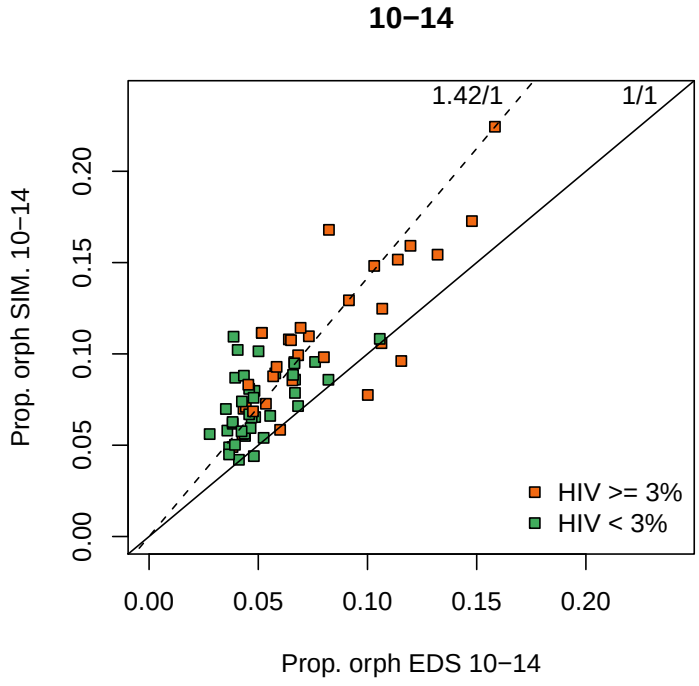

FIGURE 3.34 – *Proportions d'orphelins issues des microsimulations et observées dans les enquêtes EDS : 10-14 ans (n = 75)*

empiriques d'enfants orphelins de mères sont en moyenne inférieures de 41% aux proportions issues de la modélisation. Pour la survie des mères, cette différence varie peu avec le niveau de la prévalence du VIH ; un écart comparable est observé dans les pays à faible prévalence [82]. Pour les pères, les proportions empiriques cadrent mieux avec celles issues des modèles : elles sont inférieures de 17% en retenant toutes les enquêtes, et de 27% en retenant uniquement les enquêtes où la prévalence était faible cinq ans avant l'enquête [83]. Selon Grassly et al. (2004), ces observations impliquent que les divergences entre les proportions empiriques et modélisées ne peuvent pas s'expliquer par une surestimation de la mortalité due au VIH-sida dans les projections de population des Nations Unies. Cette hypothèse est également invalidée pour nos microsimulations, car nous pouvons faire les mêmes observations.

Selon Grassly et al. (2004), les écarts entre les proportions empiriques et celles issues de la modélisation sont plutôt dus à deux phénomènes :

1. Premièrement au *biais d'adoption*, qui tend à diminuer les proportions d'orphelins observées dans les enquêtes. Le fait que les biais d'adoption soient plus importants pour les mères que pour les pères pourrait expliquer le fait que les proportions d'orphelins *de père* cadrent davantage avec les estimations issues de modèles que les proportions d'orphelins *de mère* [84]. Une plus grande fréquence du remariage chez les hommes, et une proportion peut-être élevée de belles-mères déclarées comme mères biologiques, pourraient induire une sous-déclaration plus importante des décès des mères (Bicego et al. 2003, Blacker 1984).
2. Mais l'hypothèse principale relevée par Grassly et al. (2004) pour expliquer les divergences entre les deux séries de proportions est la *surestimation de la mortalité adulte "sans sida"* par les Nations Unies. Ces auteurs pointent notamment le recours aux tables de mortalité utilisées pour générer les proportions d'orphelins ; il s'agit des familles "Ouest" et "Nord" de Princeton (Coale et al. 1983). Or ces deux familles de tables se caractérisent par des niveaux de mortalité adulte élevés par rapport à la mortalité des enfants, ce qui pourrait expliquer la surestimation de la mortalité adulte "sans sida". En réduisant la mortalité "sans sida" de 45% chez les femmes et de 20% chez les hommes, Grassly et al. (2004) obtiennent des proportions d'orphelins qui ne sont plus significativement différentes de celles observées dans les enquêtes.

Pour expliquer les différences entre les estimations issues des microsimulations et les proportions d'orphelins déduites des enquêtes, nous pouvons faire les mêmes hypothèses que Grassly et al. (2004). Sans doute le biais d'adoption joue-t-il un rôle important, bien qu'il soit possible également que la mortalité adulte soit surestimée par les Nations Unies. Ces

82. En retenant uniquement les enquêtes pour lesquelles la prévalence du VIH était inférieure à 5% cinq ans avant l'enquête, les proportions empiriques sont inférieures de 45% aux proportions issues de la modélisation.

83. Inférieure à 5% cinq ans avant l'enquête.

84. Ceci n'est pas mentionné par Grassly et al. (2004).

deux éléments procurerait des proportions d'orphelins trop élevées à partir des microsimulations. Cela étant, il faut apporter deux précisions :

- Grassly et al. (2004) se basent sur des proportions d'orphelins établies pour 2002 (UNAIDS 2002b), qui dépendent des espérances de vie issues de la Révision 2002 des projections de population des Nations Unies. Depuis, la mortalité adulte a été revue à la baisse pour un certain nombre de pays (suite notamment à la révision des prévalence - *cf.* Reniers, Masquelier, et Gerland (forthcoming)). Or nos microsimulations se basent sur la dernière Révision (2008), avec des niveaux de mortalité plus faibles.
- Il faut noter également que les proportions empiriques cadrent davantage avec celles issues de nos microsimulations que ce n'était le cas avec les estimations de l'UNAIDS utilisées par Grassly et al. (2004). En moyenne, les proportions d'orphelins de mères à 5-9 ans sont inférieures à celles des microsimulations de 24% seulement (en retenant toutes les enquêtes)[85], alors que Grassly et al. (2004) indiquent qu'elles étaient en moyenne inférieures de plus de 45% aux estimations de l'UNAIDS/UNICEF de l'époque (sur les 0-14 ans).

Pour ces deux raisons, on est donc plutôt amené à penser que c'est le biais d'adoption qui joue ici le rôle le plus important.

3.6.2 Application de la méthode des orphelins aux microsimulations

Après avoir présenté les proportions d'orphelins issues des microsimulations, nous en déduisons des estimations de mortalité de façon indirecte. A ce stade, le fait que les proportions issues des enquêtes et celles issues des simulations soient différentes importe peu ici. L'objectif consiste plutôt à évaluer si les estimations indirectes sont être proches des niveaux de mortalité réellement à l'oeuvre dans les microsimulations.

A partir des populations fictives, nous procédons à des estimations pour chaque année, comme si des enquêtes ou des recensements étaient disponibles sur une base annuelle. Ces estimations peuvent ensuite être datées, afin de reconstruire une tendance de la mortalité féminine à partir de la survie des mères. Cette tendance, obtenue indirectement, peut alors être comparée à celle de la mortalité qui l'a produite.

Méthode "classique"

Dans un premier temps, nous utilisons les coefficients de Timaeus (1992). L'âge moyen à la naissance des enfants est calculé annuellement (à partir des naissances survenues dans l'année). Pour tenir compte des évolutions de la fécondité, les valeurs de l'âge moyen qui sont introduites dans les calculs ne sont pas celles *du moment*, mais celles qui correspondent *au moment de la naissance des enfants*. Par exemple, pour convertir la proportion d'orphelins de 5 à 9 ans observée en 2000 en probabilité de survie (${}_{10}p_{25}$), nous retenons la valeur de l'âge moyen à la naissance des enfants de mi-1992 (soit 2000 - 7.5).

85. Et de 22% en ne retenant que celles où la prévalence du VIH était inférieure à 3% cinq ans avant l'enquête. Ces moyennes valent respectivement 0.28 et 0.26 pour les 10-14 ans et 0.26 et 0.28 pour les 0-4 ans.

Les différentes probabilités de survie peuvent ensuite être datées par la méthode de Brass et Bamgboye (1981)[86]. Cette datation va permettre de comparer les estimations indirectes avec celles qui peuvent être calculées directement, de façon biographique.

Le graphe 3.35 présente, pour les différentes valeurs de n (c'est-à-dire pour les différentes classes d'âge des descendants), les probabilités de décès féminines ${}_nq_{25}$, telles qu'elles peuvent être estimées *directement* et *indirectement*, à partir des microsimulations menées au départ des paramètres démographiques de *Madagascar*. Comme les valeurs annuelles des probabilités de décès obtenues directement sont assez erratiques, elles sont lissées au moyen d'une fonction spline cubique (présentée en traits continus clairs). Les probabilités obtenues indirectement à partir de la survie des mères sont présentées en traits plus foncés.

Nous avons vu que la période de référence correspond à un passé plus lointain pour des valeurs de n plus élevés, ce qui explique que les estimations indirectes soient tronquées de plus en plus tôt à mesure qu'augmente l'âge des descendants. Les proportions d'orphelins de mères en 2009 sont en effet les dernières observées dans les microsimulations. A partir des proportions calculées sur les enfants de 5 à 9 ans (n= 10), nous obtenons une estimation pour septembre 2004. A partir des descendants de 40-44 ans (n = 45), la dernière estimation indirecte est datée de août 1994.

En l'absence d'erreurs de déclarations, la méthode des orphelins procure ici des probabilités de décès féminines pour Madagascar qui sont *très proches* des valeurs qui peuvent être recalculées directement. On relèvera uniquement une légère sous-estimation des risques de décès à partir des plus jeunes enfants, qui va se transformer progressivement en une légère surestimation lorsque n atteint 25 ans. Nous n'avons pas trouvé d'explication évidente à ces écarts entre les deux séries d'estimations[87]. Quoiqu'il en soit, ces écarts restent très légers. Lorsque n vaut 10, les estimations indirectes sont en moyenne inférieures de 3.5% aux estimations directes [-7.19 - 0%][88]. Si n vaut 45, elles leurs sont supérieures de 3.72% [2.14 - 5.23%] en moyenne. Les erreurs dans les données (sous-déclaration, attirances sur les âges ronds, etc.) sont susceptibles d'introduire des variations beaucoup plus importantes.

86. La méthode alternative de Palloni et al. (1984) ne donne pas des résultats très différents. Selon notre expérience, le calcul des périodes de référence par la méthode de Palloni et al. (1984) tend seulement à concentrer davantage les estimations sur un intervalle de temps plus restreint et plus éloigné de l'enquête.

87. Une hypothèse est que ces écarts sont liés au fait que nous modélisons la fécondité par parité dans les simulations, alors que le calcul des coefficients utilisés par la méthode des orphelins repose sur des schémas de fécondité tous rangs.

88. Les chiffres entre crochets correspondent aux 1er et 10ème déciles de la différence relative entre les deux séries de valeurs.

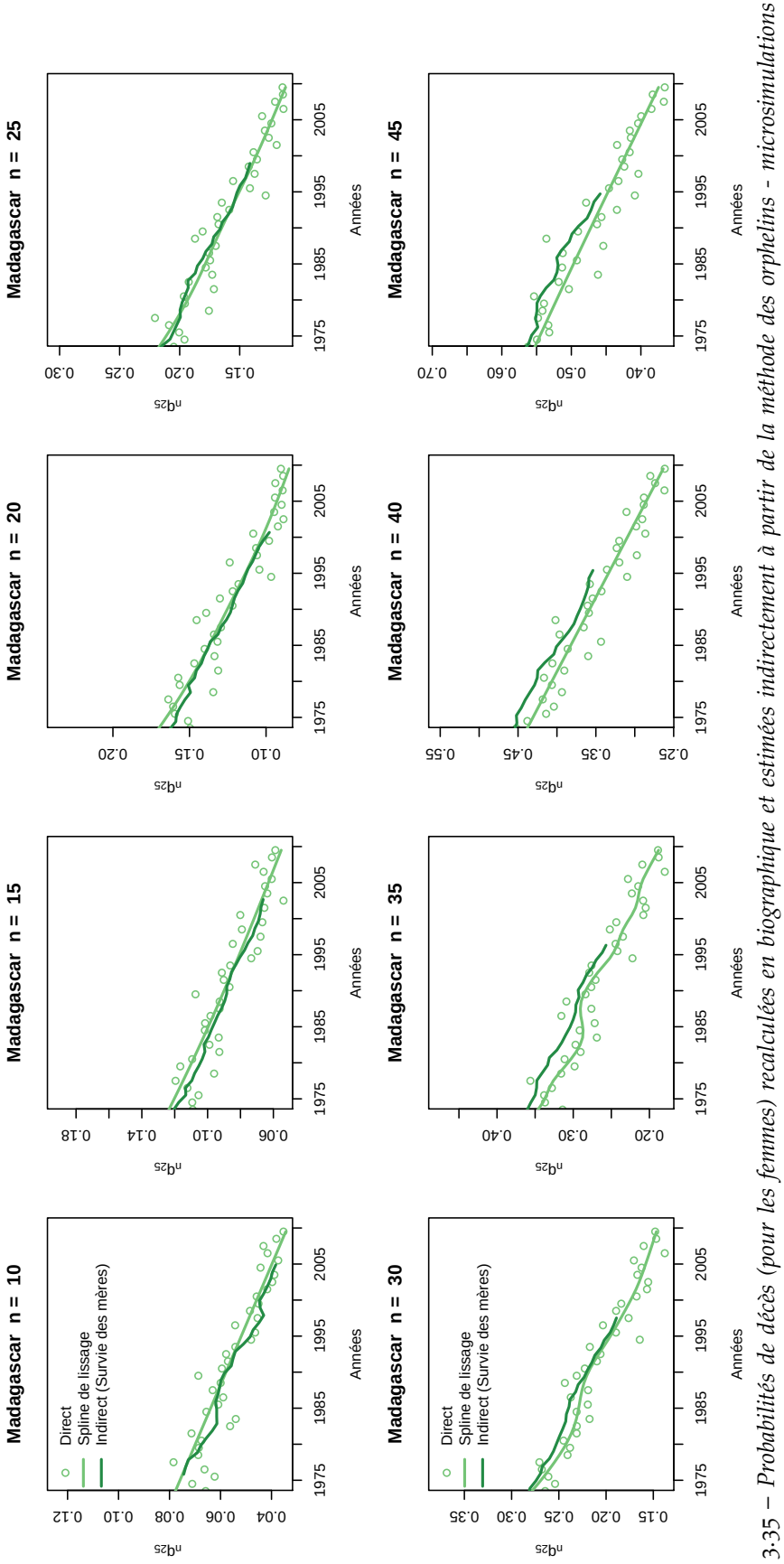

FIGURE 3.35 – *Probabilités de décès (pour les femmes) recalculées en biographique et estimées indirectement à partir de la méthode des orphelins - microsimulations menées pour Madagascar.*

Selon les estimations de l'ONU, qui sont à la base de nos simulations, la mortalité adulte a baissé de façon relativement linéaire à Madagascar, comme on peut l'observer sur le graphe 3.35. La principale condition pour pouvoir dater avec précision les estimations est donc rencontrée. Ce n'est plus le cas des pays où le VIH-sida est venu perturber les tendances. En outre, dans ces pays, les coefficients utilisés pour convertir les proportions de non-orphelins en probabilités de survie sont inappropriés, car leurs schémas de mortalité sont atypiques. Une sélection plus importante des orphelins s'opère également, en raison de la transmission verticale du VIH. En conséquence, les probabilités de décès calculées indirectement sont nettement moins proches des probabilités calculées directement. Ces problèmes sont présentés sur le graphe 3.36, à partir du cas du *Zimbabwe*.

- A partir des 5-9 ans (n = 10), les deux séries de valeurs sont relativement proches jusqu'au début des années 1990, mais on observe un décrochage très net à partir de 1990. Les estimations basées sur la survie des mères atteignent leur maximum à des niveaux bien inférieurs à ceux atteints par les probabilités calculées directement. Ainsi, en 2000, les probabilités de décès de 25 à 35 ans calculées directement s'élèvent à 0.24, tandis que les probabilités indirectes se chiffre à 0.13, soit une différence de plus de 45%. *A partir des enquêtes EDS, même en l'absence d'erreurs de déclarations, l'application "classique" de la méthode des orphelins aux pays perturbés par le VIH procure des niveaux de mortalité sérieusement biaisés à la baisse.*
- C'est le cas également des valeurs intermédiaires de n (15, 20 ou 25), pour lesquelles la datation va également poser problème : les estimations indirectes ont tendance à *anticiper* la reprise de mortalité de quelques années. A partir de ces différents groupes d'âge, on a donc à la fois une surestimation de la mortalité dans les premiers moments de l'épidémie (et même un peu avant), puis une sous-estimation de la mortalité (à partir de dix ans environ avant que celle-ci n'amorce une baisse).
- Ensuite, pour des valeurs de n supérieures à 25, c'est plutôt une surestimation qui prévaut. Les tendances esquissées par les estimations indirectes indiquent bien une reprise de mortalité, mais la localisation dans le temps de cette reprise est inexacte.

Le cas du Zimbabwe est typique des pays où la mortalité a connu une forte hausse en raison de la crise du VIH-sida : Botswana, Kenya, Malawi, Namibie, Ouganda, Swaziland, Zambie, Tanzanie, etc. On peut y faire exactement les mêmes constats en comparant les estimations directes et indirectes. Mais une configuration comparable s'observe dans les pays où la mortalité connaît une hausse importante, malgré une prévalence du VIH-sida plus faible. Les niveaux de mortalité atteints y sont en conséquence moins élevés (Cameroun, Congo, Côte d'Ivoire, etc.). Les simulations menées à partir des paramètres du *Congo-Brazza*, par exemple, sont présentées dans le graphe 3.37. Là encore, on note une sous-estimation importante de la mortalité à partir des premiers groupes d'âge, puis, une surestimation croissante avec l'âge.

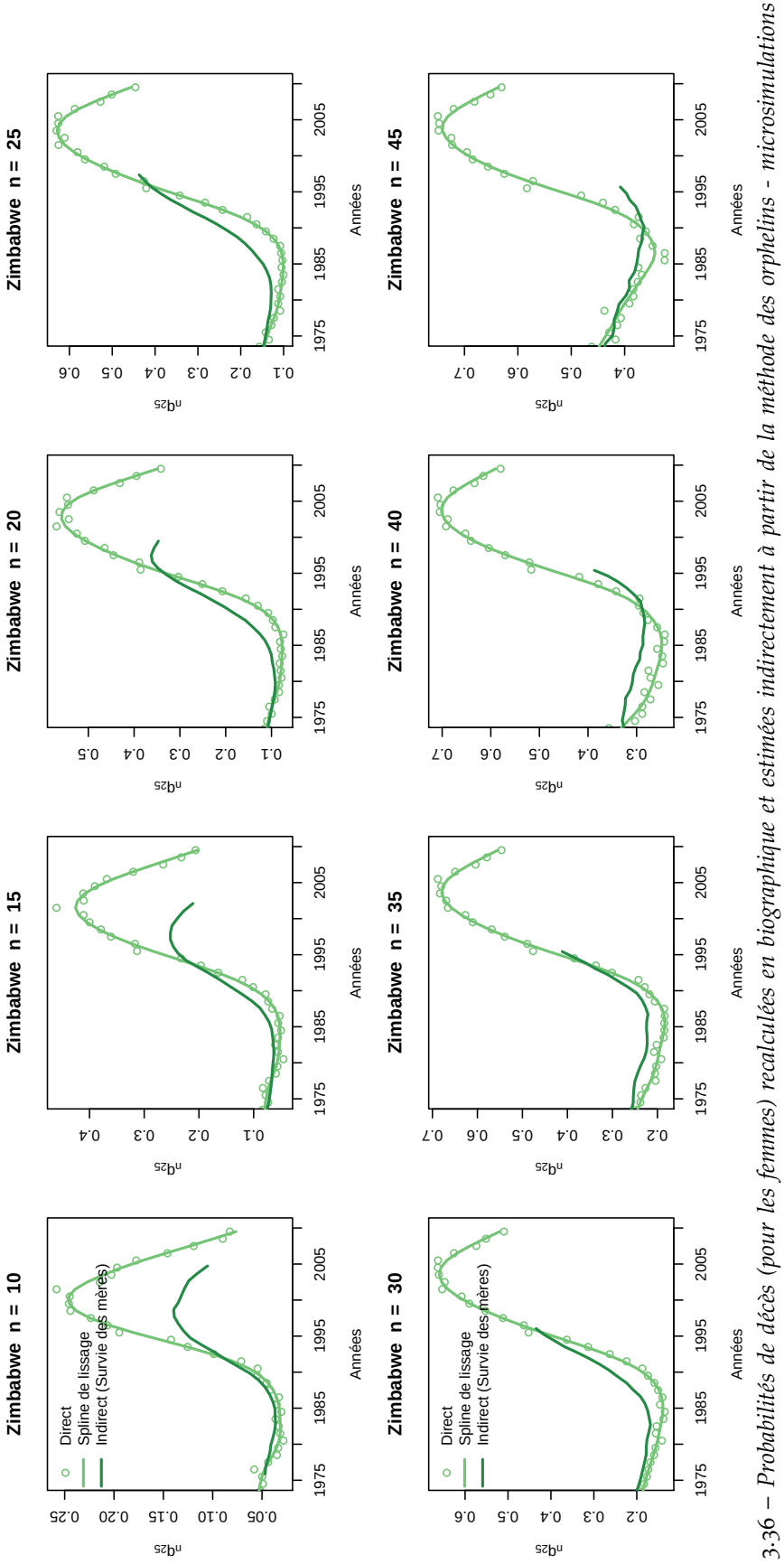

FIGURE 3.36 – *Probabilités de décès (pour les femmes) recalculées en biographique et estimées indirectement à partir de la méthode des orphelins - microsimulations menées pour le Zimbabwe.*

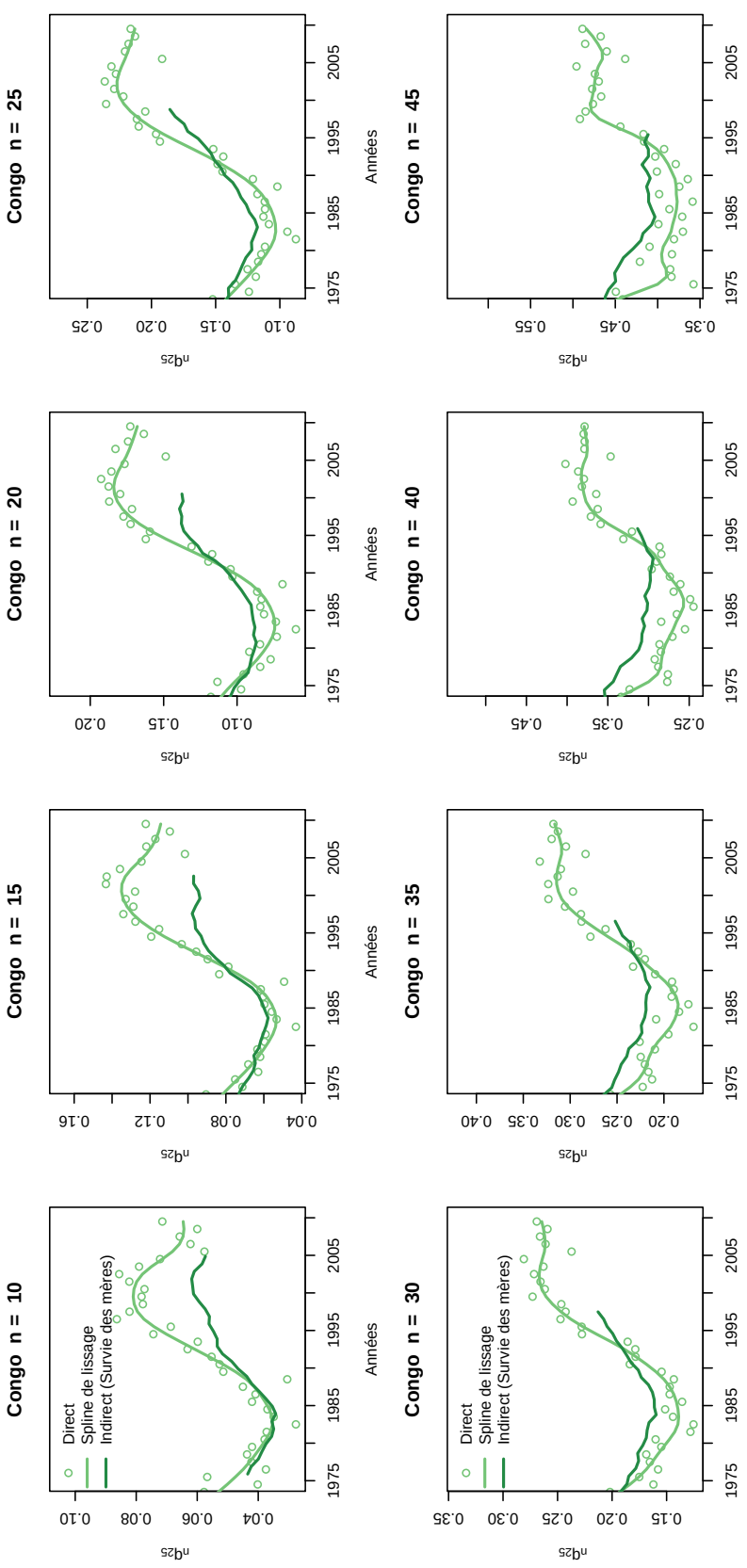

FIGURE 3.37 – *Probabilités de décès (pour les femmes) recalculées en biographique et estimées indirectement à partir de la méthode des orphelins - microsimulations menées pour le Congo (Brazza).*

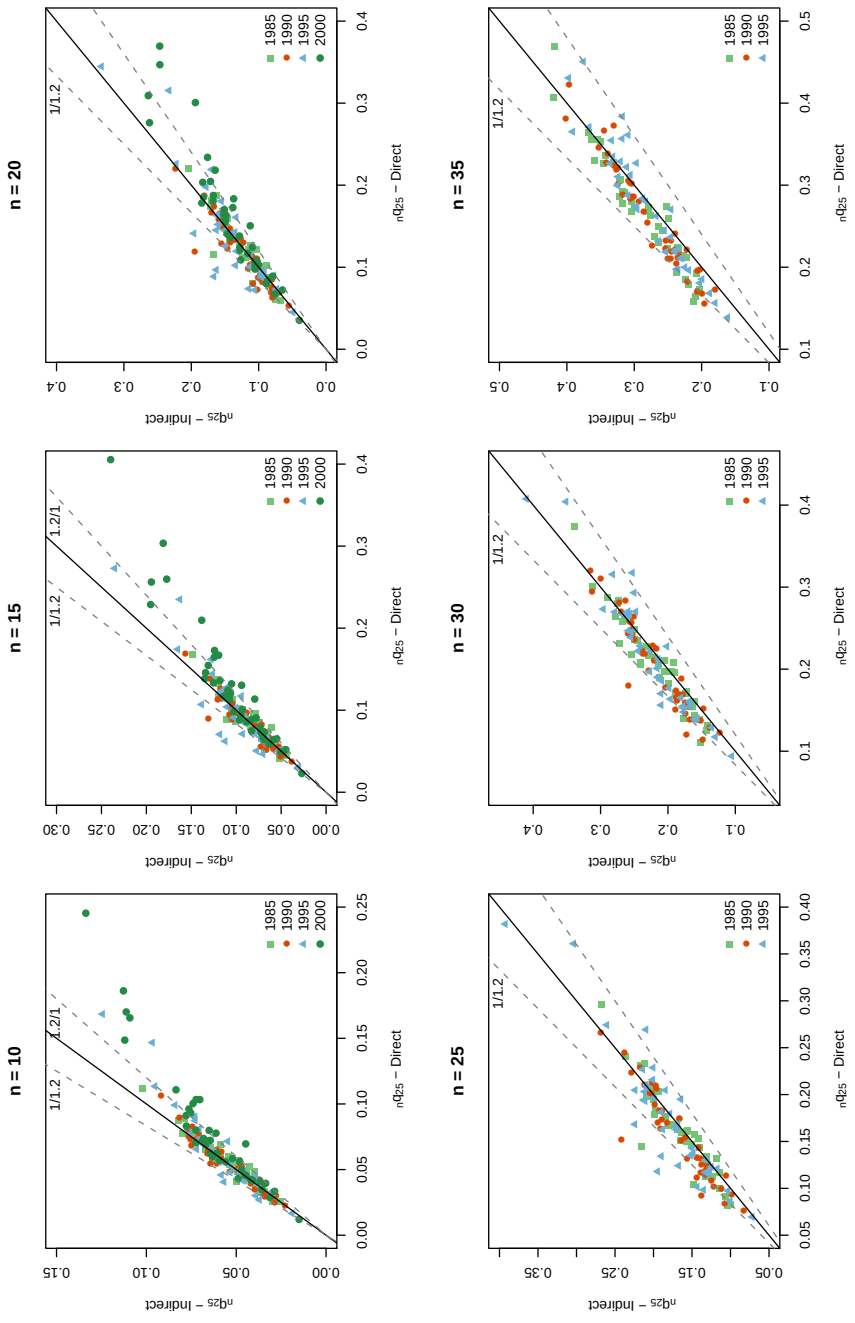

FIGURE 3.38 – *Comparaison des estimations directes et indirectes pour les 41 populations simulées - Mortalité féminine (1985, 1990, 1995, 2000)*

Le graphe 3.38 présente une comparaison plus systématique des estimations directes (en abscisses) et indirectes (en ordonnées), pour l'ensemble des 41 pays pour lesquels nous avons mené des microsimulations. Les estimations sont comparées en 1985, 1990, 1995 et 2000. On retrouve les constats faits précédemment. Pour des valeurs de n inférieures à 25, la plupart des estimations indirectes pour 1985 et 1990 se différencient de moins de 20% des estimations directes. Pour les périodes postérieures par contre, il y a davantage de points qui témoignent soit d'une *sous-estimation* de la mortalité de plus de 20% (c'est surtout le cas en 2000), soit d'une *surestimation* de la mortalité en 1995. La hausse de la mortalité, généralement due au VIH-sida, est d'abord anticipée (et donc surestimée) puis sous-estimée. Pour des valeurs de n à partir de 25 ans, les écarts les plus conséquents correspondent à une surestimation de la mortalité (il n'y a plus d'estimation indirecte possible pour 2000).

Cohortes synthétiques

Une des sources de divergences les plus importantes entre les estimations directes et indirectes semble être liée à la *datation des estimations indirectes*, et plus particulièrement à l'hypothèse de linéarité des tendances de la mortalité. Lorsque cette condition n'est pas rencontrée, il est préférable de recourir aux méthodes basées sur les *cohortes synthétiques*. La datation n'est plus nécessaire, et les estimations obtenues sont plus récentes. Nous appliquons ici la méthode de Preston, qui consiste à multiplier les proportions de non-orphelins par un facteur de correction qui va "éliminer" l'impact des tendances passées (*cf.* supra, p.152).

Le graphe 3.39 présente les estimations obtenues par ce biais, toujours à partir des microsimulations, pour Madagascar, le Congo, la Namibie et le Zimbabwe. Nous nous sommes placés dans la situation idéale où une collecte des données était faite *tous les trois ans*, en interrogeant tous les individus sur la survie de leur mère, quel que soit leur âge. A partir de là, nous estimons la mortalité pour chaque intervalle "inter-censitaire" de trois ans entre 1982 et 2009.

Pour chacun de ces intervalles, la méthode procure une série de probabilités de survie ${}_np_{25}$, dont l'étendue varie avec la valeur prise par n. Ces probabilités doivent être converties en un indice commun (généralement le quotient ${}_{45}q_{15}$) pour fournir une estimation unique. C'est là que réside une des difficultés de la méthode, car il faut opter pour une table de mortalité standard au sein de laquelle il va falloir interpoler. Or nous avons vu que le choix du standard peut avoir un impact important sur les probabilités ${}_{45}q_{15}$ obtenues[89]. Nous présentons ici les résultats pour les deux schémas de mortalité INDEPTH détaillés plus haut (p.121). Pour chaque intervalle "inter-censitaire", une moyenne est prise sur les probabilités de décès calculées à partir des répondants âgés de 20 à 39 ans. Ce sont généralement ceux qui fournissent les informations les plus fiables.

89. *Cf.* p.121. C'est le cas si l'on considère des schémas perturbés par le VIH, car les schémas plus standards procurent des estimations assez proches.

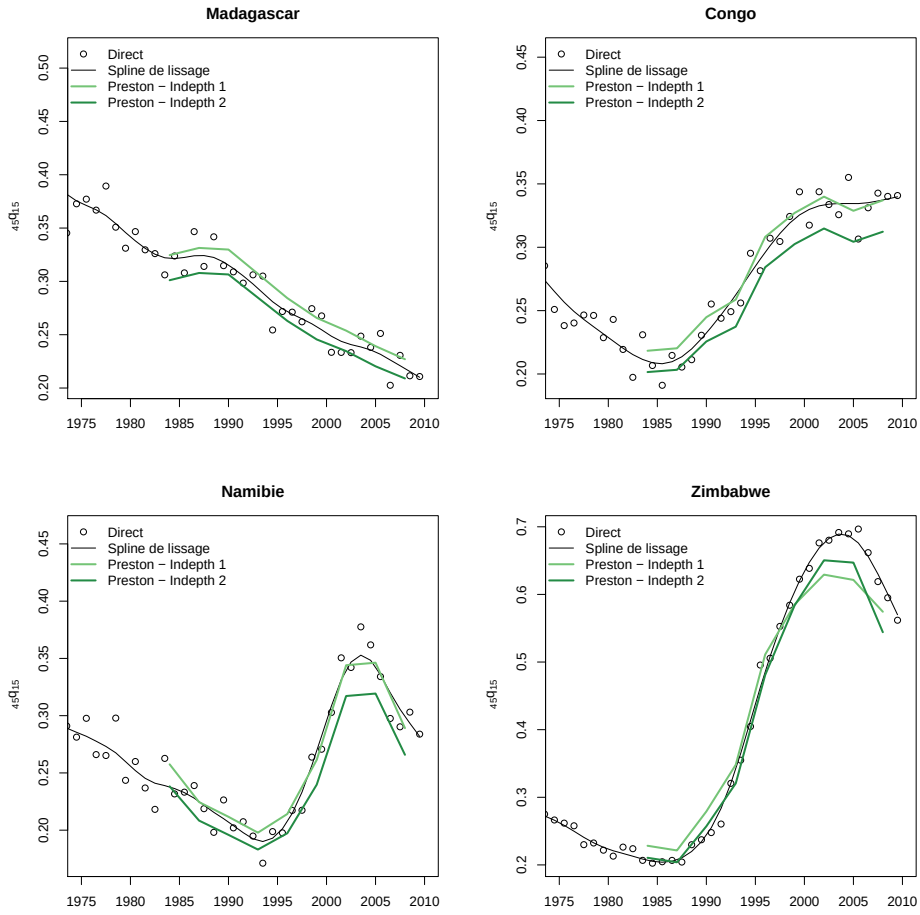

FIGURE 3.39 – *Probabilités de décès* $_{45}q_{15}$ *(féminines) calculées directement et estimées à partir de la survie des mères - Madagascar, Congo, Zambie, Zimbabwe. Moyenne prise sur les estimations calculées à partir des 20-39 ans.*

Globalement, *les estimations indirectes cadrent beaucoup mieux avec la mortalité réellement observée dans les simulations que ce n'était le cas en datant les estimations de façon classique* (section précédente). Dans les contextes perturbés par le VIH, cette approche devrait donc être privilégiée. Alors que la reprise de mortalité était très mal estimée auparavant, cette série d'estimations parvient à capturer avec précision l'ampleur de la hausse au Congo, en Namibie et au Zimbabwe.

On est toutefois ici dans un cas idéal, pour plusieurs raisons :

1. Aucune erreur de déclaration n'est introduite ici.
2. Les données sont collectées régulièrement.
3. Ces données sont disponibles pour tous les âges, et il est donc possible de faire une moyenne sur les estimations dégagées des 20-39 ans.

A ces âges, le choix d'opter pour une table de mortalité standard ou pour une table perturbée par le VIH n'a pas d'impact trop important, car la relation entre les quotients ${}_nq_{25}$ et ${}_{45}q_{15}$ est relativement stable. Ce n'est plus le cas si l'on s'en tient aux moins de 15 ans (n < 20), pour lesquels les choix faits pour convertir les probabilités ${}_nq_{25}$ en ${}_{45}q_{15}$ vont avoir un impact majeur. Par ailleurs, nous avons vu avec le cas du Zimbabwe et du Congo que lorsque le VIH-sida joue un rôle important, la mortalité est sensiblement *sous-estimée* lorsqu'elle était appréhendée à partir des plus jeunes enquêtés. On retrouve cette sous-estimation dans le graphe 3.40, basé sur les 5-9 et 10-14 ans uniquement. Pour éviter les problèmes associés au choix du standard, nous présentons ici la probabilité de décéder de 25 à 40 ans.

On retombe sur les bons niveaux de mortalité à Madagascar, et en Namibie avant 1995, mais il y a par contre une forte sous-estimation dans les autres cas, et cette sous-estimation s'accroît à mesure que l'épidémie du VIH gagne en importance. Ce que nous mettons ici en évidence est un problème sérieux, dans la mesure où une bonne part des données disponibles sur la survie des parents proviennent des enquêtes EDS, qui fournissent des informations seulement sur ces deux groupes d'âge, de 5 à 14 ans.

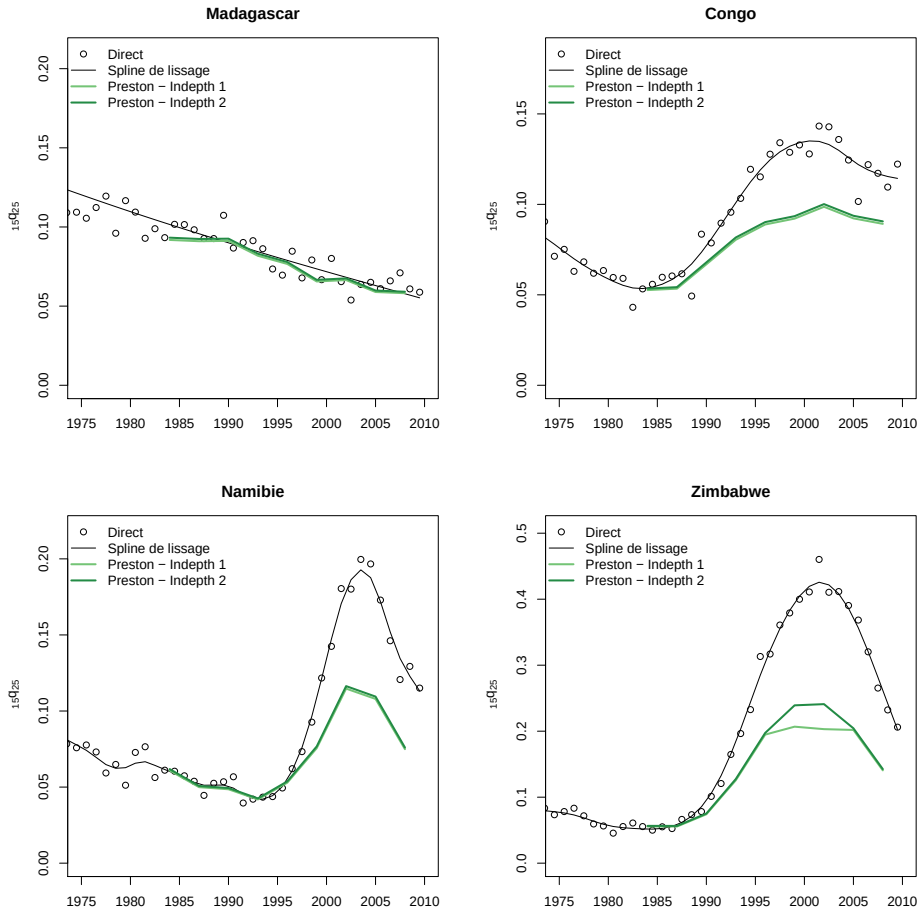

FIGURE 3.40 – *Probabilités de décès* ${}_{15}q_{25}$ *(féminines) calculées directement et estimées à partir de la survie des mères - Madagascar, Congo, Zambie, Zimbabwe. Moyenne prise sur les estimations calculées à partir des 5-14 ans.*

Conclusion

Ce chapitre a permis de détailler la méthode des orphelins, les différentes procédures de datation et les principaux biais de sélection qui la caractérisent. Dans un sens, les résultats présentés ici sont encourageants, car ils montrent que dans les pays où les tendances ont été approximativement linéaires, la méthode de base permet de procurer des niveaux de mortalité plausibles. En l'absence du VIH-sida, les biais de sélection restent limités. Les différentes extensions de la méthode fournissent par ailleurs une série d'estimations alternatives qui peuvent être utilisées pour juger de la cohérence d'ensemble des tendances dégagées.

D'un autre côté, l'application de la méthode aux microsimulations a permis de pointer une sous-estimation importante des risques de décès dans les pays fortement touchés par le VIH. Il n'existe actuellement aucune approche permettant de dater avec précision les estimations obtenues dans le cas où les tendances de la mortalité ont été perturbés. Il est peu probable qu'une telle méthode puisse être développée.

Le recours aux cohortes synthétiques permet alors de contourner ce problème de la datation, mais cette approche suppose de pouvoir disposer de données collectées régulièrement. Ce n'est pas le cas aujourd'hui dans la plupart des pays africains, mis à part à partir des enquêtes EDS, qui ne concernent que les premiers groupes d'âges (5-9 et 10-14 ans). Or précisément, à ces âges, la crise du VIH-sida a tant modifié la relation entre la proportion d'orphelins et les probabilités de survie que la méthode ne peut être utilisée comme telle.

Le chapitre suivant, consacré aux problèmes méthodologiques spécifiquement associés au VIH, propose certains amendements et de nouveaux coefficients de corrections.

Orphelins et crise du VIH-sida

4

Sommaire

4.1 Perturbations induites par le VIH-sida 198
4.2 Calcul du nombre d'orphelins de mère et de père au niveau macro 207
4.3 Retour sur les probabilités de décès 217
Conclusion 222

La crise du VIH-sida a eu un impact majeur sur le quotidien des enfants dans les pays les plus sévèrement touchés par l'épidémie. Une de ses conséquences les plus dramatiques est certainement la forte hausse du nombre d'orphelins.

Une littérature récente est consacrée à la situation des orphelins en Afrique Sub-saharienne. La plupart des travaux portent sur leur scolarisation, sur leurs conditions de vie et les ménages qui les accueillent, ainsi que sur les risques en matière de santé et de mortalité associés à la perte d'un parent (Beegle et al. 2010, Monasch et Boerma 2004, Ainsworth et Filmer 2006, Case et al. 2004, Evans et Miguel 2007, Nakiyingi et al. 2003, Ng'weshemi et al. 2003, Zaba et al. 2005).

Si ces travaux sont relativement nombreux, en revanche, la question de l'*estimation du nombre* de ces orphelins a fait l'objet de moins de développements. Les recherches les plus récentes ont été menées par Grassly et Timaeus (2005), qui ont mis au point une technique d'estimation à partir des projections de population au niveau macro. Cette méthode sert aujourd'hui de référence, et elle est notamment utilisée par l'US Census Bureau pour produire les estimations reprises par l'UNICEF et l'UNAIDS [1].

Les estimations les plus récentes sont publiées par l'UNICEF (2006). Celles-ci projettent le nombre total d'orphelins de moins de 18 ans en Afrique sub-saharienne à 53 millions en 2010, ce qui représenterait 12% de la population des 0-17 ans. 30% de ces orphelins ont perdu au moins

1. Avant cela, les modèles développés pour procéder à l'estimation du nombre d'orphelins procuraient des résultats assez variables, en raison d'hypothèses et d'approches différentes (UNAIDS 2002a). On peut citer notamment les travaux de Preble (1990) et de Gregson et al. (1994).

l'un de leurs parents en raison du VIH-sida, ce qui fait de l'Afrique subsaharienne la région qui concentre 80% des enfants orphelins en raison du sida. La forte hausse des proportions d'orphelins s'accompagne d'une hausse encore plus prononcée de la proportion d'orphelins qui perdent leurs *deux* parents, en raison des risques élevés de transmission entre partenaires. Au total, en Afrique, un peu moins d'un orphelin sur cinq en 2010 est un "double orphelin" [2].

Au niveau méthodologique, pour parvenir à ces chiffres, l'UNICEF "retourne le problème" de l'estimation indirecte de la mortalité adulte. En effet, ce sont les niveaux estimés de la mortalité qui servent alors de point de départ au calcul du nombre d'orphelins, et plus l'inverse. Dans un contexte perturbé par le VIH-sida, ce calcul doit prendre en compte plusieurs éléments : la réduction de la fécondité des mères, les risques de transmission verticale, les durées de survie des adultes et des enfants infectés, etc. Ce calcul nous intéresse ici car il relie quasiment directement les proportions d'orphelins aux probabilités de survie. Il est mené au niveau macro, et il fournit une série de proportions d'orphelins qui peut être comparée à celle de nos micro-simulations.

Ces deux approches, macro et micro, présentent toutefois leurs spécificités. Les micro-simulations sont un outil précieux, car les différentes composantes de la dynamique démographique y sont agencés de façon tout à fait cohérente : fécondité, mortalité, infection au VIH-sida, transmission verticale, ... Tout se tient, par définition, puisque tout se joue au niveau individuel au moyen de risques compétitifs. Les micro-simulations fournissent donc un cadre idéal pour appliquer les différentes méthodes indirectes et voir si l'on peut retrouver les niveaux de mortalité sous-jacents (comme nous l'avons fait dans le chapitre précédent). Par contre, à ce stade, nous pensons que les proportions d'orphelins qui sont issues de nos micro-simulations ne sont pas tout à fait satisfaisantes. Par exemple, notre modèle micro ne tient pas compte de la **surmortalité des orphelins** [3]. Par ailleurs, nous avons fixé un peu arbitrairement **la distribution par sexe de l'incidence**, ce qui a un impact sur les niveaux et les tendances de la mortalité, et donc sur les proportions d'orphelins. Enfin, la façon avec laquelle nous traitons de l'incidence revient à supposer que les probabilités d'infection au VIH de deux conjoints sont indépendantes, ce qui biaise à la hausse les **proportions d'orphelins de père** (*cf.* chapitre 2.).

Pour ces différentes raisons, dans ce chapitre consacré aux perturbations induites par le VIH sur la méthode des orphelins, nous proposons de compléter nos microsimulations par des calculs menés au niveau **macro**, afin de générer d'autres proportions d'orphelins, qui seraient plus précises dans les contextes perturbés par le sida. Comme il est impossible de reprendre directement les estimations de l'UNICEF (car celles-ci ne se réfèrent qu'à une année donnée, et il s'agit ici d'obtenir des tendances),

2. Nous reprenons ici la classification la plus commune des différents types d'orphelins. On dira qu'un enfant est "orphelin de mère" si sa mère est décédée, que son père ait survécu ou non. Cela signifie que le nombre d'orphelins de mère comprend le nombre de "doubles orphelins". Le nombre total d'orphelins correspond à la somme des orphelins de mère et des orphelins de père, somme dont on déduit les doubles orphelins.

3. Sauf pour ce qui est de la mortalité par sida chez les enfants infectés verticalement.

nous procédons à nos propres calculs en appliquant leur méthode, qui est celle développée par Grassly et Timaeus (2005). Ces proportions d'orphelins, estimées au niveau macro, vont nous permettre de mieux analyser la relation qui unit les proportions d'orphelins aux probabilités de décès là où le VIH-sida joue un rôle important.

Nous avons vu qu'en raison du VIH-sida, l'estimation *indirecte* de la mortalité à partir des parents peut fournir des niveaux assez éloignés de la mortalité sous-jacente. C'est surtout le cas pour les groupes d'âge pour lesquels on dispose de données EDS (5-9 et 10-14 ans), à partir desquels la mortalité est sensiblement sous-estimée. Il est donc important d'amender la méthode, et de chercher notamment à **développer d'autres coefficient**s permettant de convertir les proportions d'orphelins en probabilités de survie. C'est certainement là que se situent les prochaines avancées méthodologiques.

Timaeus et Nunn (1997) sont les premiers - et les seuls - à avoir cherché à rendre la méthode des orphelins applicable aux pays fortement touchés par le VIH. Ils suggèrent une série provisoire de coefficients à partir des données de l'observatoire de population de Masaka, en Ouganda. Cet amendement à la méthode "classique" se traduit par une hausse importante des quotients de mortalité estimés, comme nous le verrons plus loin. Mais en appliquant cette méthode aux micro-simulations, nous montrerons qu'elle ne permet toujours pas de récupérer les niveaux de mortalité "réellement" à l'oeuvre. Après avoir présenté brièvement leurs travaux, nous proposons donc une série de coefficients alternative, qui nous semble procurer des niveaux de mortalité plus plausibles. Pour y parvenir, nous utilisons les proportions d'orphelins calculées au niveau macro par la méthode d'estimation de l'UNICEF. Notre démarche consiste ensuite à régresser les niveaux de mortalité estimés par les Nations Unies sur ces proportions d'orphelins, en tenant compte de la prévalence du VIH. C'est par ce biais que nous proposons de nouveaux coefficients, qui devraient permettre d'appliquer la méthode des orphelins aux contextes perturbés par le sida. Avant de terminer ce chapitre, nous cherchons à valider ces coefficients à l'aide de micro-simulations.

En d'autres termes, nous nous basons sur des estimations obtenues **au niveau macro** pour produire une nouvelle série de coefficients, que nous validons ensuite **au niveau micro**. La démarche reste exploratoire mais l'utilisation de ces nouveaux coefficients a un impact conséquent sur les estimations de mortalité.

4.1 Perturbations induites par le VIH-sida

L'application de la méthode des orphelins dans les contextes où la prévalence du VIH est élevée pose problème à deux égards :

1. D'une part, *les biais de sélection sont plus prononcés*, en raison à la fois des risques de transmission verticale et de la moindre fécondité des femmes séropositives. Ceci concourt à une sous-estimation de la mortalité.
2. D'autre part, la crise du VIH-sida a généré des *profils de mortalité par âge inédits* : les risques de décès connaissent une forte hausse aux premiers âges adultes (entre 15 et 30 ans), puis, par la suite, la mortalité augmente moins vite avec l'âge que dans les schémas "classiques" [4]. Ces schémas perturbés diffèrent sensiblement de ceux qui sont à la base de la méthode des orphelins - qu'il s'agisse de tables types ou, plus souvent, du système des logits combiné au standard de Brass. Ces différences ont un effet important sur les estimations dégagées de la survie des parents. A nouveau, elles se traduisent par une sous-estimation de la mortalité lorsque celle-ci est estimée à l'aide des coefficients classiques.

Examinons plus en détail ces deux problèmes méthodologiques causés par le VIH-sida.

4.1.1 Biais de sélection associés au VIH-sida

Premièrement, la transmission verticale va renforcer le biais de sélection causé par la surmortalité des orphelins. Il en résulte une sous-estimation de la mortalité, qui affecte surtout les estimations de mortalité *féminine*. Elle affecte également les estimations *masculines*, mais dans une moindre mesure. En effet, comme les pères des enfants de mères séropositives ont une probabilité élevée d'être infectés eux aussi, la transmission verticale va se traduire par une sous-estimation de la mortalité des pères. Rappelons que sans traitement, le risque de transmission verticale du VIH-1 est estimé à 15-25% *in utero* ou lors de l'accouchement. De plus, entre 5 et 20% d'enfants non infectés intra-partum sont infectés durant l'allaitement.

Par ailleurs, les femmes séropositives ont en moyenne moins d'enfants, en tous cas à partir de 20 ans [5]. En conséquence, il y aura encore moins d'enfants pour déclarer des risques de décès plus élevés, ce qui accentue la sous-estimation de la mortalité.

4. Ce schéma "typique" de la mortalité dans les contextes perturbés par le VIH-sida a été présenté en page 121, à partir des données issues des sites d'observation démographique du réseau INDEPTH.

5. Voir la section 2.4.5, p.94. Entre 15 et 19 ans, la fécondité des séropositives est supérieure d'environ 25% à celle des séronégatives du même âge (en raison d'une activité sexuelle plus précoce). A partir des 20-24 ans, c'est l'inverse, et la fécondité des séropositives s'élève à 0.76 [0.66-0.89] de celle des femmes séro-négatives, et ce rapport chute à 0.67 [0.6-0.74] au delà de 25 ans (Lewis et al. 2004).

Estimation des biais : le cas du Lesotho

Les microsimulations peuvent être utilisées pour estimer l'ampleur du biais introduit par ces deux mécanismes de sélection. Nous travaillons sur deux scénarii différents, à partir des paramètres démographiques du Lesotho. Dans le premier scénario, nous optons pour des simulations qui comprennent des *risques de transmission verticale* ainsi que des *changements de fécondité chez les mères séropositives*. Dans le second, nous menons les mêmes microsimulations, mais en éliminant ces deux sources de biais. Le graphe 4.1 présente les séries de proportions d'orphelins par âge observées dans les deux cas.

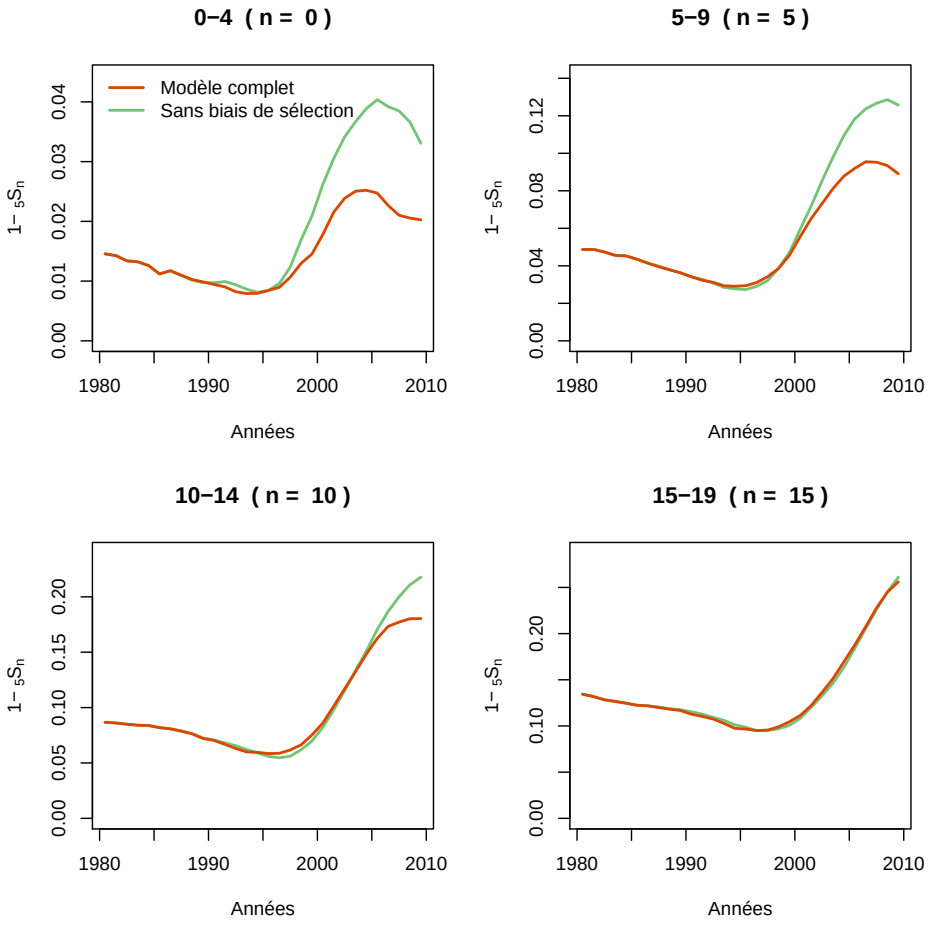

FIGURE 4.1 – *Comparaison des proportions d'orphelins en l'absence et en présence des biais de sélection associés au VIH-sida - microsimulations menées à partir du cas du Lesotho (1980-2010)*

On note effectivement que les proportions obtenues *en l'absence des deux biais de sélection* sont supérieures à celles *réellement observées*. Le biais varie dans le temps et il va en augmentant avec la hausse de la mortalité. On note que la sélection a d'abord un effet chez les 0-4 ans, puis, quelques années plus tard, chez les 5-9 ans. Elle commence à concerner les 10-14 ans un peu avant 2005, soit environ 10 ans plus tard que les 0-4 ans, puis elle disparaît chez les 15-19 ans.

On voit donc que pour un enfant *survivant*, la probabilité d'avoir une mère *survivante* est beaucoup plus élevée dans la situation où opère cette

double sélection, par rapport à la situation où cette sélection n'opère pas. Ceci s'observe également dans le tableau 4.1. Celui-ci présente, sous forme de *odds ratios*, les différences entre la probabilité d'avoir une mère survivante dans les deux configurations [6]. Par exemple, la chance d'avoir une mère survivante pour un enfant de 5 à 9 ans est 47% plus élevée en présence des biais de sélection. Les biais de sélection ont pour effet de relever de 26% la probabilité d'être non-orphelin entre 10 et 14 ans. L'ampleur du biais diminue avec l'âge et il s'estompe pratiquement chez les 15-19 ans.

	1997	2001	2005	2009
0-4	1.160	1.430	1.659	1.655
5-9	0.937	1.111	1.325	1.471
10-14	0.904	0.963	1.060	1.265
15-19	0.994	0.982	0.975	1.027

TABLE 4.1 – *Rapports de chances (odds ratios) qu'une mère soit survivante en présence de biais de sélection, par rapport à la situation sans biais de sélection (Micro-simulations, Lesotho 1997-2009)*

Ceci exprime l'effet de la sélection sur la probabilité d'avoir une mère survivante. Le biais sur la *proportion de non-orphelins*, qui est à la base des calculs, s'obtient simplement en rapportant les proportions avec sélection sur les proportions sans sélection [7]. Le biais se chiffre chez les 5-9 ans à 0.4% en 2000, puis il s'élève rapidement pour atteindre 4.2% en 2009. Chez les 10-14, ce biais est négligeable en 2000, mais il atteint 4.8% en 2009. Chez les 15-19, il ne dépasse pas 1%. Notons que ces biais sont supérieurs à ceux examinés dans le chapitre précédent, dans les cas où le VIH ne joue pas un rôle prédominant. Ils restent toutefois relativement modestes. On peut mentionner deux éléments pour l'expliquer. D'une part, une majorité des enfants des mères infectés sont nés avant cette infection (Timaeus et Nunn 1997). D'autre part, les courbes de survie sont telles que quand mères et enfants sont infectés, il est beaucoup plus fréquent que ce soit l'enfant qui décède d'abord (Ng'weshemi et al. 2003).

Timaeus et Nunn (1997) ont procédé à une estimation comparable, à partir de *macro*-simulations menées au départ des données de l'observatoire ougandais de Masaka. Les biais qu'ils obtiennent sont légèrement inférieurs à ceux calculés ici à partir des microsimulations. Cette différence

6. Timaeus et Nunn (1997) comparent également les proportions *avec* et *sans* sélection, à partir de macro-simulations. C'est de leur travail que provient notre idée d'opter pour les *odds ratios*. Ces *odds ratios* permettent de mieux comparer l'effet des biais de sélection d'un groupe d'âge à l'autre - une comparaison plus simple, qui consisterait à rapporter simplement les deux séries de proportions l'une sur l'autre rendrait les biais apparemment plus faibles chez les plus jeunes, car les proportions d'orphelins y sont nettement plus faibles. Ce n'est pas le cas des *odds ratios*. Si l'on note ${}_5S_x^*$ la proportion obtenue en présence des biais de sélection et ${}_5S_x$ la proportion obtenue sans biais de sélection, les *odds ratios* se calculent comme :

$$\frac{{}_5S_x^*}{1 - {}_5S_x^*} \times \frac{1 - {}_5S_x}{{}_5S_x}$$

7. Soit :

$$1 - {}_5S_x^* / {}_5S_x \tag{4.1}$$

tient certainement à une moindre prévalence du VIH entre l'Ouganda des années 1990 et le Lesotho des années 2000. La prévalence était plus élevée au Lesotho dans les années 2000 qu'une dizaine d'années avant en Ouganda. Or, l'ampleur du biais est directement proportionnelle à la *prévalence atteinte au moment de la naissance des enfants*, comme cela est détaillé dans la section suivante.

Facteur d'ajustement basé sur la prévalence atteinte

Le lien entre le biais de sélection et la prévalence est établi par Timaeus et Nunn (1997) à l'aide d'un modèle mathématique. Il leur permet de proposer un facteur d'ajustement à appliquer dans les pays où la prévalence est élevée.

Nous présentons en annexe A.4.1 les développements qui permettent à ces auteurs d'aboutir au facteur d'ajustement. En supposant une transmission verticale de 25% et un rapport de la fécondité des *séropositives* sur celle des *séronégatives* de 0.8, cet ajustement prend la forme suivante [8] :

$$ {}_5S_n^{adj} = \frac{1 - 0.25\ P}{1 + 0.25\ P} \times_5 S_n^* \approx (1 - 0.475\ P) \times_5 S_n^* \qquad (4.2) $$

Dans ce calcul, P renvoie à la prévalence observée parmi les mères *au moment de la naissance des enfants*. Il s'agit typiquement de la prévalence issue des services de consultations prénatales (SCPN). Lorsque la prévalence utilisée est calculée sur les toutes les femmes - c'est-à-dire au niveau de la population générale plutôt que chez les femmes qui ont accouché -, l'ajustement est réduit à $(1 - 0.4\ P)$ [9].

Pour parvenir à ces expressions, de nombreuses hypothèses sont faites. On suppose notamment que les femmes qui étaient séropositives quand elles ont donné naissance décèdent dans les cinq ans, et que les enfants infectés verticalement décèdent également à moins de cinq ans [10]. Pour tenir compte du fait qu'une fraction importante de ces enfants vont survivre au delà de cinq ans, le facteur d'ajustement peut être divisé par deux. Mais cette correction reste approximative.

Dans la pratique, seules les données provenant des 5-9 ans et des 10-14 ans devraient être ajustées. Nous avons vu à partir du cas du Lesotho que la sélection n'a pas d'effet important sur les 15-19 ans.

Le graphe 4.2 présente les résultats de cette correction sur les *proportions de non-orphelins* observées dans les microsimulations pour le Lesotho [11]. Les courbes en rouge représentent les *proportions non corrigées*, les courbes en vert (et en traits pleins) les *proportions que l'on observerait en l'absence de sélection*, et les courbes en pointillés celles obtenues à *l'aide*

8. Pour les plus de 5 ans.

9. Voir Timaeus et Nunn (1997), p.32.

10. En outre, la réduction de la fécondité des mères séropositives est supposée constante, leur mortalité ne varie pas avec la durée depuis l'infection (mais par âge), les orphelins ne connaissent pas de surmortalité en l'absence de transmission verticale, etc.

11. Nous utilisons la prévalence observée chez les femmes de 15 à 49 ans, en la décalant de 7.5 ans pour les 5-9 ans et de 12.5 ans pour les 10-14 ans (afin d'avoir la prévalence au

du facteur d'ajustement que nous venons de décrire. On constate que l'ajustement a pour effet de rapprocher les proportions de non-orphelins de celles qui seraient observées en l'absence de sélection. Les proportions corrigées apparaissent même ici légèrement trop basses.

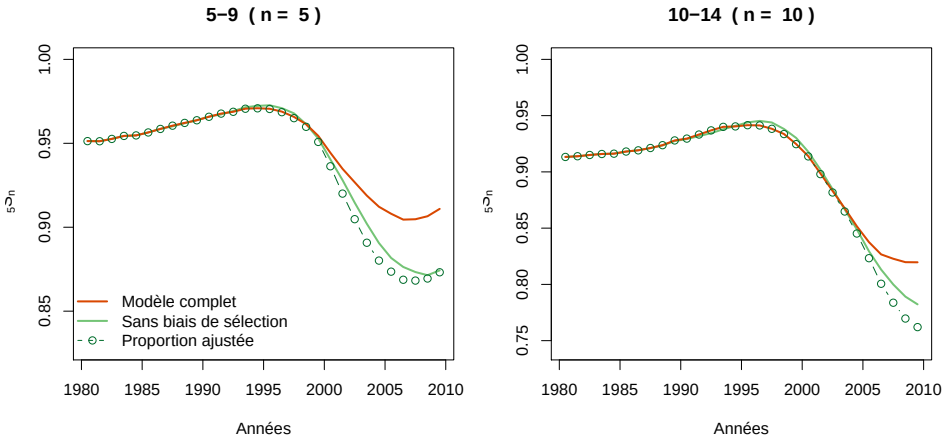

FIGURE 4.2 – *Comparaison des proportions de non-orphelins :*
- en l'absence des biais de sélection associés au VIH-sida (en vert),
- en présence de ces biais (en rouge),
- et après ajustement de Timaeus et Nunn (1997)
Microsimulations menées à partir du cas du Lesotho (1980-2010)

Le tableau 4.2 présente, toujours par le biais de *odds ratios*, le résultat de l'ajustement sous forme numérique sur les données du Lesotho. Il doit être comparé au tableau précédent. Si la sélection relevait de 47% chez les 5-9 ans (et de 26% chez les 10-14 ans) la probabilité d'avoir une mère survivante, le facteur d'ajustement de Timaeus et Nunn (1997) vient atténuer les effets de cette sélection. Les *odds ratios* sont nettement plus proches de l'unité.

		1997	2001	2005	2009
5-9	Non ajusté	0.937	1.111	1.325	1.471
10-14	Non ajusté	0.904	0.963	1.06	1.265
5-9	Ajusté	0.916	0.892	0.927	0.99
10-14	Ajusté	0.904	0.957	0.958	0.891

TABLE 4.2 – *Rapports de chances (odds ratios) qu'une mère soit survivante en présence de biais de sélection, par rapport à la situation sans biais de sélection (Micro-simulations Lesotho 1997-2009), résultat de la correction de Timaeus et Nunn (1997)*

moment de leur naissance). Nous optons pour les corrections suivantes :

$$
\begin{aligned}
{}_5S_5^{adj} &\approx (1 - 0.2\ P) \times_5 S_5^* \\
{}_5S_{10}^{adj} &\approx (1 - 0.4\ P) \times_5 S_{10}^*
\end{aligned}
$$

4.1.2 Schémas de mortalité atypiques

Nous arrivons au second problème de méthode posé par le VIH. En raison de schémas de mortalité atypiques - avec des risques de décès qui sont plus élevés dans les premiers âges adultes -, la relation entre les proportions d'orphelins et les probabilités de survie est modifiée. *Pour une même proportion de non-orphelins, la mortalité est plus élevée.* Pour cette raison, le recours aux coefficients classiques va induire une sous-estimation de la mortalité. C'est le constat que tirent Timaeus et Nunn (1997) de leur analyse des données de l'observatoire de Masaka :

> "The life table odds of surviving are overestimated substantially from data on the first two age groups. In contrast, the biases in estimates based on the reports of respondents aged more than 15 years are small" (Timaeus et Nunn (1997), p. 34).

Ces auteurs proposent donc une série de coefficients alternative. Rappelons que c'est la seule série disponible actuellement pour les contextes perturbés par le VIH, et qu'elle se limite à la mortalité féminine. A ce stade, il n'y a donc aucune option pour estimer la mortalité masculine à partir des pères lorsque le sida joue un rôle important.

n	β_0	β_1	β_2
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

TABLE 4.3 – *Coefficients fournis par Timaeus et Nunn (1997) pour convertir les proportions de mères survivantes en probabilités de survie dans les contextes perturbés par le VIH-sida*

Ces coefficients se distinguent de ceux présentés précédemment (p.115) de deux façons : les constantes (β_0) sont plus faibles, et les coefficients associés aux probabilités de survie (β_2) sont plus élevés. La "pente" de la droite de régression est plus forte. Par exemple, chez les 10-14 ans, la constante vaut -0.1718 en l'absence de sida et -0.4030 dans la nouvelle série. Le coefficient relatif à l'âge moyen à la naissance des enfants ne change pas, mais la valeur du coefficient β_2, associé à la proportion de non-orphelins s'élève maintenant à 1.3732, alors qu'il s'établissait à 1.1123 auparavant. Ces nouveaux coefficients sont donc susceptibles d'avoir un impact important sur les niveaux de mortalité obtenus.

Cet impact s'observe sur le graphe 4.3, que nous avons construit à partir des 28 enquêtes EDS de 15 pays différents. Ces enquêtes sont retenues car la prévalence nationale du VIH atteignait au moins 3% à la naissance des enfants de 5 à 9 ans (c'est-à-dire 7.5 ans avant l'enquête). Ce graphe oppose les quotients de mortalité féminine obtenus sans correction (en abscisse) et ceux obtenus en recourant à la méthode de Timaeus et Nunn (1997), repris en ordonnées.

Dans les deux graphes du dessus (A et B), les proportions de survivants qui sont à la base du calcul sont simplement diminuées au prorata

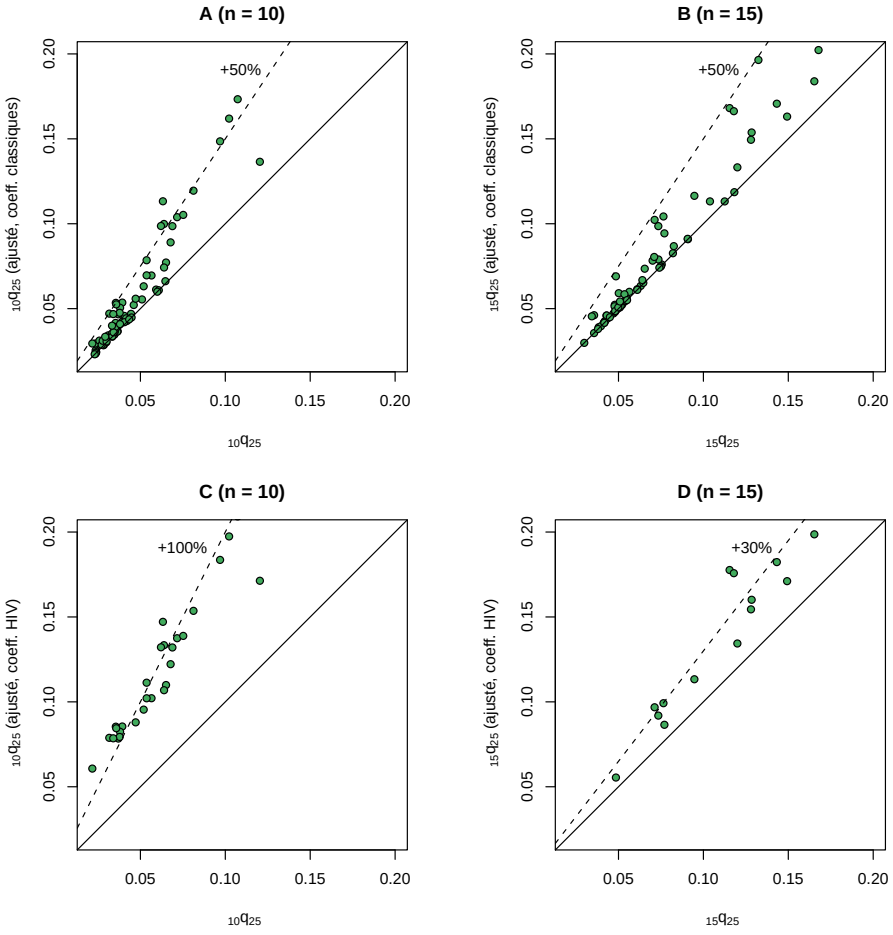

FIGURE 4.3 – *Comparaison des quotients ${}_nq_{25}$ calculés à partir de 28 (pour n = 10) et 17 (pour n = 15) enquêtes EDS (15 pays) :*
***Panel du haut** : Quotients obtenus sans correction (en abscisses) et quotients obtenus en corrigeant les proportions de survivants selon la prévalence (en ordonnées)*
***Panel du bas** : Quotients obtenus sans correction (en abscisses) et quotients obtenus en corrigeant les proportions de survivants selon la prévalence et optant pour les coefficients de Timaeus et Nunn (1997) (en ordonnées)*

de la prévalence atteinte, tout en conservant les coefficients 'classiques' [12]. A partir des 5-9 ans (schéma A), cet ajustement a pour effet de relever les quotients ${}_{10}q_{25}$ de 20% environ sur l'ensemble des enquêtes - et cette correction peut atteindre 79% (EDS Zimbabwe 1999). L'effet de cette correction est moins important pour les 10-14 ans (schéma B), notamment parce que les prévalences au moment de leur naissance étaient plus basses [13]. Dans les deux graphes du dessous (C et D), les quotients sont obtenus en recourant aux coefficients proposés par Timaeus et Nunn (1997), après avoir corrigé les proportions de non-orphelins. Par rapport aux quotients ${}_{10}q_{25}$ obtenus classiquement à partir des 5-9 ans, ceux obtenus de la sorte sont en moyenne *deux fois supérieurs* (schéma C). Il y a moins d'enquêtes où les probabilités de décès peuvent être calculées de cette façon à partir des 10-14 ans [14], mais quand c'est le cas, les quotients ${}_{15}q_{25}$ sont en moyenne relevés de 30%.

Ces deux ajustements ont donc un effet majeur sur les estimations de mortalité qui sont dégagées de la méthode des orphelins. Pour autant, Timaeus et Nunn (1997) ont bien précisé qu'ils étaient provisoires car ils n'étaient construits qu'à partir de la situation observée en Ouganda. Appliqués aux microsimulations, ils procurent en tous cas des niveaux de mortalité qui ne cadrent pas tout à fait avec ceux qui sont à la base des simulations. On le voit sur le graphe 4.4, qui, sur 4 pays, oppose (1) les probabilités de décès féminines réellement à la base des simulations, (2) celles recalculées de façon classique (en vert) à partir des orphelins et (3) celles recalculées après avoir corrigé les proportions d'orphelins et en optant pour les coefficients de Timaeus et Nunn (1997) (en orange). Les probabilités de décès ${}_{10}q_{25}$ sont ici transformées au moyen du standard INDEPTH-2 en quotients ${}_{15}q_{25}$ pour être combinées aux estimations issues des 10-14 ans. On constate que les amendements qui visent à tenir compte du VIH-sida ont bien pour effet de relever les quotients de mortalité, mais qu'il y a toujours ce décalage temporel de quelques années par rapport aux "vrais" niveaux de mortalité, ainsi qu'une sous-estimation assez importante des risques de décès.

Cela nous motive à *rechercher d'autres coefficients*, qui seraient construits sur une plus large gamme de situations, représentatives de la diversité des trajectoires démographiques africaines. Pour ce faire, une option aurait pu être de partir simplement de nos microsimulations, et de rapporter les proportions de non-orphelins aux probabilités de survie pour obtenir de nouveaux coefficients. Mais nous n'avons alors pas la possibilité de valider ces coefficients, et par ailleurs, nous ne pouvons les calculer pour les hommes. Nous avons donc opté pour une autre approche, qui consiste, rappelons-le, à partir des proportions d'orphelins estimées par la méthode utilisée par l'UNICEF.

12. Nous utilisons les corrections mentionnées dans la note 11. L'âge moyen à la naissance des enfants est celui observé au moment de l'enquête.

13. Le graphe pour les 10-14 ans fait également intervenir moins d'enquêtes (17 au lieu de 28) et moins de pays (12 au lieu de 15) car le seuil des 3% de prévalence doit avoir été passé 12.5 ans avant l'enquête (soit à la naissance des 10-14 ans).

14. Voir note précédente.

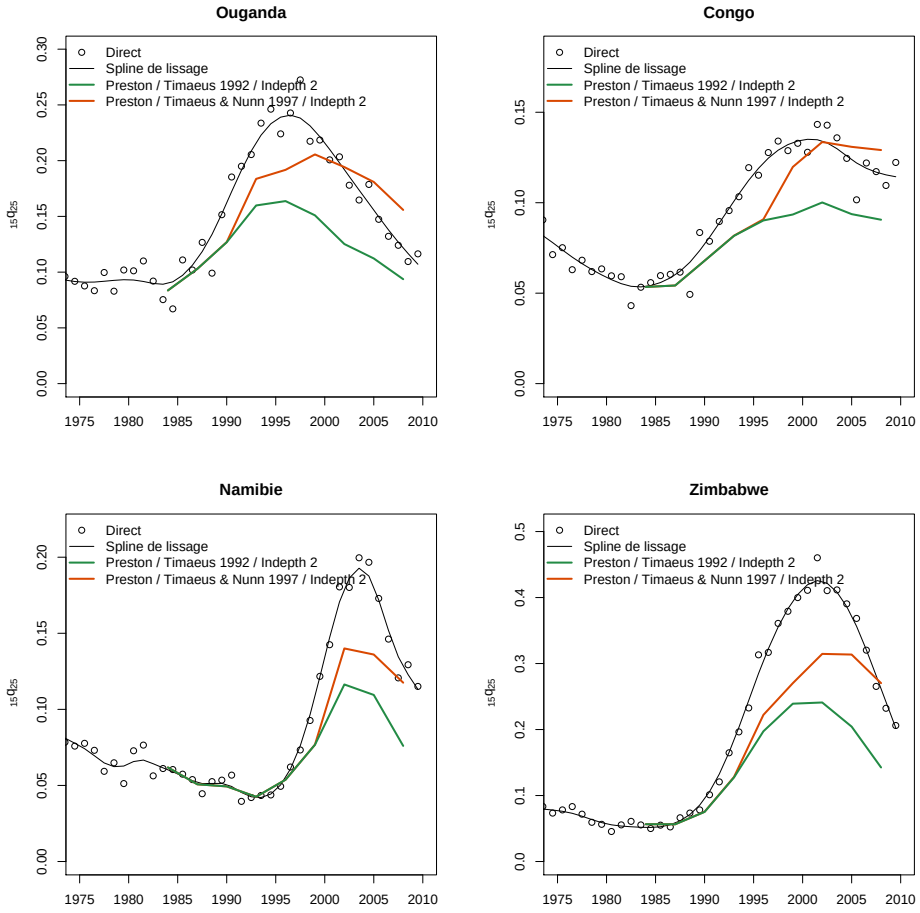

FIGURE 4.4 – *Probabilités de décès* $_{15}q_{25}$ *(féminines) calculées directement et estimées à partir de la survie des mères. Moyenne prise sur les estimations calculées à partir des 5-14 ans. Comparaison des estimations obtenues classiquement avec celles obtenues par le biais des coefficients de Timaeus et Nunn (1997).*

PRÉVALENCE DU VIH ESTIMÉE :
Ouganda : 13.9 (1990) - 8.5 (2000) - 6.3 (2005)
Congo : 5.2 (1990) - 4.7 (2000) - 4.0 (2005)
Zambie : 6.3 (1990) - 15.3 (2000) - 19.2 (2005)
Zimbabwe : 12.5 (1990) - 27.3 (2000) - 19.2 (2005)

4.2 Calcul du nombre d'orphelins de mère et de père au niveau macro

La méthode utilisée par l'UNICEF, qui a été développée par Grassly et Timaeus (2005), consiste à partir des décès adultes pour en déduire un nombre d'orphelins. Elle est en fait tout spécialement conçue pour estimer le nombre d'orphelins *dûs au sida*. Dans les quelques pages qui suivent, nous la détaillons dans les grandes lignes, surtout pour préciser les choix méthodologiques que nous avons fait en l'appliquant à 41 pays d'Afrique sub-saharienne [15]. Pour une description plus complète, nous renvoyons à l'article de Grassly et Timaeus (2005).

Tout part d'une moyenne des probabilités de survie des orphelins séropositifs et séronégatifs, pondérée par une probabilité de transmission verticale (notée ξ). Le calcul se fait en spécifiant différentes étapes au sein de la progression que les parents effectuent depuis l'infection au VIH jusqu'à leur décès. On notera n le nombre d'étapes retenues, et s l'étape atteinte par les mères. A partir de là, on peut exprimer la probabilité de survie de la naissance à l'âge a, pour un orphelin dont la mère était à l'étape s au moment de sa naissance comme :

$$\xi_s l'_a + (1 - \xi_s) l_a \tag{4.3}$$

où ξ_s correspond à la probabilité de transmission verticale à l'étape s, l'_a correspond à la probabilité de survie de la naissance à l'âge a pour un enfant séropositif et l_a à la même probabilité pour un enfant séronégatif. C'est cette somme qui constitue la base de tous les calculs. Nous présentons d'abord le calcul qui mène aux nombres d'orphelins de mère, puis celui conçu pour les orphelins de père.

Nombre d'orphelins de mère

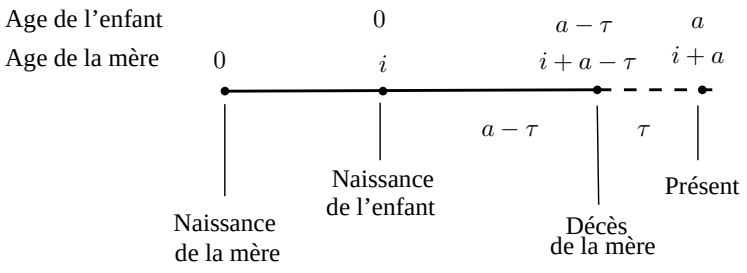

Figure 4.5 – *Schéma résumant l'âge des mères et des enfants et les différents moments clés de l'estimation du nombre d'orphelins - (Grassly et Timaeus 2005)*

La figure 4.5, reprise de Grassly et Timaeus (2005), résume les différents moments clés qui interviennent dans les calculs : la naissance de la mère, celle de l'enfant, le décès de la mère, puis le présent. On considère que le décès de la mère a eu lieu il y a τ années, et que l'enfant est né il y a a années, quand l'âge de sa mère valait i. L'enfant est donc âgé de $a - \tau$

15. Il s'agit des mêmes pays que ceux que nous avons retenus pour les microsimulations (p.56).

années au moment du décès de sa mère, qui survient quand celle-ci est âgée de $i + a - \tau$. Ces différents éléments seront utiles pour indexer, par âge et dans le temps, les taux de fécondité, le nombre des décès, etc.

La proportion de mères qui sont décédées il y a τ années alors qu'elles étaient à l'étape s de l'infection au moment de la naissance de leur enfant est notée $Y_s(a - \tau)$. On peut alors calculer la probabilité qu'un orphelin né d'une mère infectée survive jusqu'à l'âge a, en pondérant l'équation 4.3 par cette proportion $Y_s(\tau - a)$. On obtient :

$$\sum_{s=0}^{n} \left[Y_s(a-\tau)(\xi_s l'_a + (1-\xi_s) l_a) \right] \tag{4.4}$$

Pour procéder à ce calcul, il faut faire certains choix méthodologiques :

- Grassly et Timaeus (2005) supposent une probabilité de transmission ξ instantanée à la naissance [16] et constante pour toutes les étapes s. Cette probabilité est fixée à 0.3. Nous retenons la même valeur.

- Les probabilités l_a et l'_a sont basées sur les estimations de mortalité des enfants, mais elles sont ajustées à la hausse pour tenir compte d'une surmortalité en raison de la perte des parents. Nous supposons que la mortalité est multipliée par un facteur 3.9 durant l'année qui précède le décès de la mère et l'année qui suit. Cette estimation est issue des recherches menées par Zaba et al. (2005) (*cf.* p.163). Notons qu'il s'agit d'une moindre surmortalité que celle initialement prévue par Grassly et Timaeus (2005) [17]. Précisons également que l'année d'âge au cours de laquelle a lieu le décès de la mère a son importance. Pour un âge donné, plus le décès de la mère a eu lieu quand l'enfant était *jeune*, plus la surmortalité des enfants est *forte*, parce qu'elle les affecte quand ils font face à des quotients de mortalité plus élevées.

- La probabilité de survie des enfants infectés s'obtient comme $l_a \times l_a^A$, où l_a^A renvoie à la proportion d'enfants infectés qui ne sont pas encore décédés du VIH-sida à l'âge a. Pour l'évolution par âge de cette proportion, nous renvoyons à la section 2.4.4 du chapitre 2. Nous retenons ici une double distribution de Weibull et les paramètres fournis par Marston et al. (2005).

16. Ceci est justifié par le fait que le plus gros des infections post-partum surviennent dans les 6 mois qui suivent la naissance.

17. La valeur de 3.9 est inférieure à celle qu'ils retiennent pour l'année qui suit. Par ailleurs, ils prolongent cet excès de mortalité jusqu'à 15 ans. Dans la mesure où cette surmortalité est estimée chez les plus jeunes, et que les décès maternels ont surtout un effet sur la survie dans les premiers âges (sevrage précoce, etc.), nous limitons cette surmortalité aux moins de cinq ans. Zaba et al. (2005), dont nous reprenons l'estimation de la surmortalité des orphelins, précisent d'ailleurs :

> "The adverse effect of maternal death may be limited to the first 2 years of life. We have firm evidence only about the excess mortality risk in children up to the age of 3, because after that age only 30 person-years exposure to maternal death were observed" (p.279).

– Les probabilités de survie des enfants qu'il faut utiliser pour l'estimation du nombre d'orphelins sont idéalement celles des cohortes, et non les probabilités "transversales". Pour obtenir ces probabilités par cohorte, nous reprenons les tables lissées par âge et année issues des estimations des Nations Unies (*cf.* chapitre 2, p.64). Nous travaillons ensuite *en diagonale* à partir des probabilités de survie. Par exemple, la valeur de l_5 pour 1985 s'obtient comme ceci : [18]

$$l_5(1985) =_1 p_0(1980) \times_2 p_1(1981) \times_3 p_2(1982) \times_4 p_3(1983) \times_5 p_4(1984)$$

– Quant à $Y_s(x)$, la proportion de femmes adultes qui étaient à l'étape s de l'infection x années avant le décès, elle peut se calculer en résolvant une série d'équations différentielles, où interviennent les taux de progression d'une étape à l'autre (γ_s). C'est ce que montrent Grassly et Timaeus (2005). Pour chaque valeur de s, il s'agit de résoudre l'équation suivante [19] :

$$\frac{dY_s(x)}{dx} = -\gamma_s Y_s(x) + \gamma_{s+1} Y_{s+1}(x) \tag{4.5}$$

Pour modéliser la survie depuis l'infection, nous ne retenons pas les mêmes étapes et les mêmes paramètres que Grassly et Timaeus (2005), car les estimations de la durée de survie ont légèrement augmenté ces dernières années [20]. Nous optons plutôt pour la durée d'incubation utilisée actuellement par le logiciel Spectrum [21]. La survie entre l'infection et le décès (telle qu'elle est modélisée par Spectrum) est ici approximée par une succession de trois exponentielles, après avoir fixé la transition du stade clinique au décès à partir du taux repris dans Spectrum. Les deux premiers taux de transition sont estimés par approximation numérique [22].
Le graphe 4.6 présente la durée de survie de l'infection au décès prévue dans Spectrum, et la compare à notre approximation basée sur la combinaison de trois exponentielles (en traits pleins). On voit que l'approximation est tout à fait satisfaisante.
Le graphe 4.7 présente, par sexe, les proportions d'adultes décédés du VIH-sida qui étaient en phase s de l'infection selon le temps écoulé *avant* le décès. On notera les étapes s pour les femmes et r pour les hommes.

18. Cette idée est reprise de Zagheni (2010).

19. Ces équations sont résolues en posant $\gamma_0 = \gamma_{n+1} = 0$, $Y_{s=n}(0) = 1$ et $Y_{s\neq n}(0) = 0$. Nous résolvons les équations différentielles au moyen du package `deSolve` du logiciel R.

20. Grassly et Timaeus (2005) optent pour deux étapes de l'infection au décès, toutes deux modélisées par le biais d'une distribution exponentielle. Le temps médian au sein de la phase asymptotique est de 7.6 ans pour les hommes et de 8.4 ans pour les femmes, et le temps médian au sein du stade clinique de 1 ans pour les deux sexes.

21. Cette durée d'incubation prend la forme d'une distribution de Weibull avant la phase clinique puis d'une exponentielle jusqu'au décès - *cf.* graphe 2.16, p.88. Pour les hommes, la durée médiane de l'infection jusqu'au seuil des 200 cellules/mm^3 est de 8.23 ans, et la durée médiane de ce seuil au décès est de 1.5 ans. Pour les femmes, ces valeurs valent respectivement 8.55 ans et 1.95 ans. Ces valeurs diffèrent de celles retenues par les Nations Unies, mais la durée totale de l'infection au décès est similaire (10.5 ans pour les hommes et 11.5 ans pour les femmes).

22. Dans une démarche similaire à celle détaillée section 2.4.3, chap. 2, ainsi qu'en annexe A.2.8, p. 424.

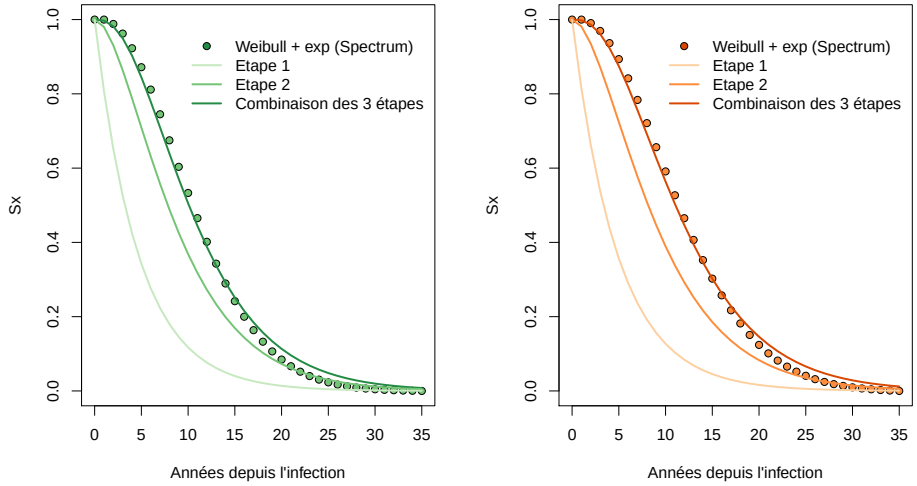

FIGURE 4.6 – *Durée de survie depuis l'infection intégrée à Spectrum, approximations obtenues en combinant trois exponentielles successives.*

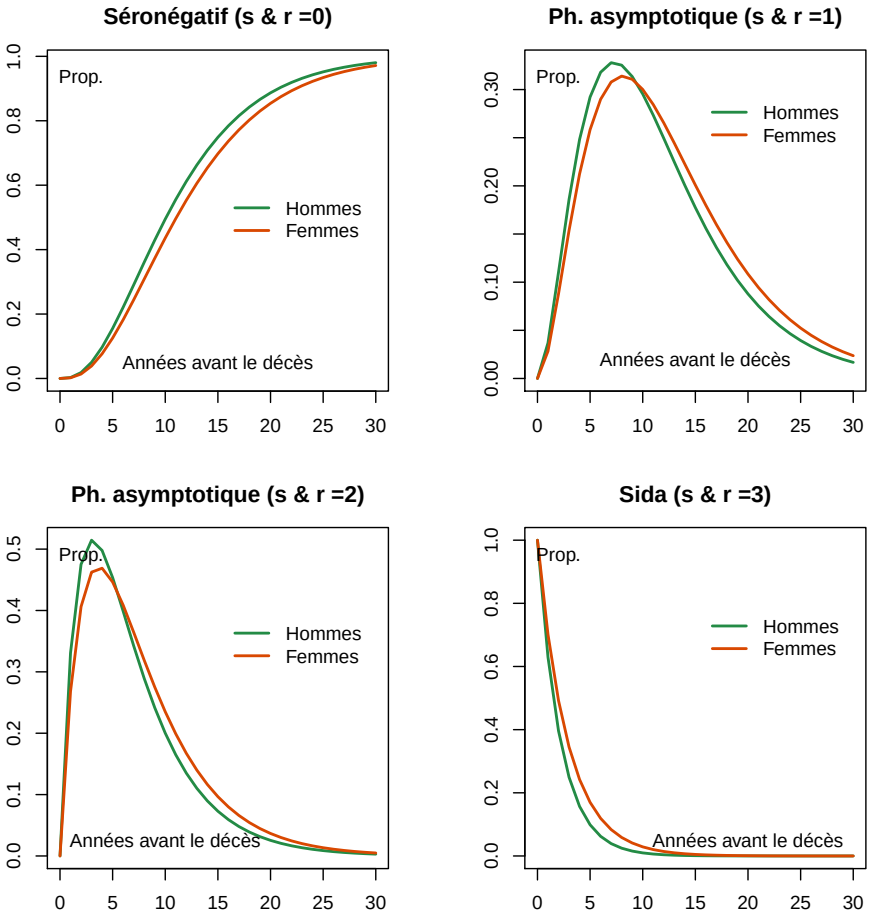

FIGURE 4.7 – *Proportion d'adultes décédés du VIH-sida qui étaient en phase s de l'infection selon le temps écoulé* ***avant*** *le décès (par sexe)*

Au départ de l'équation 4.4 (qui exprime la probabilité qu'un enfant survive juqu'à l'âge a si sa mère est décédée il y a τ années), il faut faire intervenir les taux de *fécondité* et le *nombre de décès* des femmes adultes pour obtenir le nombre d'orphelins. Ces deux éléments doivent être indexés par âge et par année, et la fécondité va également varier selon l'étape atteinte dans la progression de la maladie. En reprenant la terminologie de Grassly et Timaeus (2005), on note donc $\mu'_{t,i}$ le nombre de femmes décédées du sida à l'âge i il y a t années, et $m_{i,s,t-a}$ la fécondité des femmes d'âge i qui étaient à l'étape s de l'infection au temps $t-a$. A partir de là, le *nombre d'orphelins d'âge a au temps t dont la mère est décédée du sida il y a τ années* s'exprime comme ceci :

$$\Omega'_{t,a,\tau} = \sum_{i=15}^{49} \left[\mu'_{t-\tau,i+a-\tau} \sum_{s=0}^{n} \left[m_{i,s,t-a} \times Y_s(a-\tau)(\xi_s l'_a + (1-\xi_s) l_a) \right] \right] \quad (4.6)$$

Il s'agit de la première composante du nombre total d'orphelins. On doit y ajouter ensuite *le nombre d'orphelins dont la mère est décédée d'une cause non liée au sida*.

Pour obtenir la première composante ($\Omega'_{t,a,\tau}$), nous avons donc besoin du nombre de femmes décédées du sida par âge et par année. Grassly et Timaeus (2005) ne précisent malheureusement pas sur quelles estimations ils s'appuient. Pour notre part, nous estimons le nombre de ces décès au moyen d'une approximation détaillée dans l'encadré 9, à partir des estimations des Nations Unies. La fécondité est reprise également des estimations des Nations Unies, après avoir rehaussé celle des séropositives de 15-19 ans de 25%, et diminué celle des séropositives de 20-24 ans de 24% et au-delà, de 33% (*cf.* section 2.4.5, p.94).

Encadré 9 Estimation du nombre de décès dûs au sida

Pour estimer le nombre de décès dûs au sida, nous disposons de trois éléments : (1) les tables de mortalité des Nations Unies de la variante "moyenne", qui intègrent le sida, (2) les tables de la variante "sans sida", ainsi que (3) le nombre absolu de décès estimés par âge et par intervalle quinquennal, fourni également par les Nations Unies (UN 2009d). De ces trois types d'informations, nous devons déduire les **décès dûs au sida**. La chose n'est pas aussi simple qu'il n'y paraît, car il ne suffit pas de soustraire les décès prévus par la variante "sans sida" de ceux prévus par la variante "moyenne". Tout le monde finit un jour par décéder, et quelle que soit la variante des projections retenue, il y a toujours un décès par personne !

Notre solution consiste à partir de la **proportion de décès dûs au sida**, estimée à partir des deux variantes des tables de survie des Nations Unies (UN 2009c). Cette proportion est calculée par groupe d'âge et intervalle quinquennal comme le **rapport entre les taux de mortalité de la variante moyenne (avec sida) et les taux de la variante sans sida**. Dans la figure 4.8, les résultats de notre calcul sont présentés sous forme graphique, par âge et année, pour l'Ouganda et le Zimbabwe (sexe féminin). Dans l'ensemble, pour ces deux pays, on remarque nettement la hausse de la proportion de décès dus au sida,

surtout entre 25 et 40 ans.

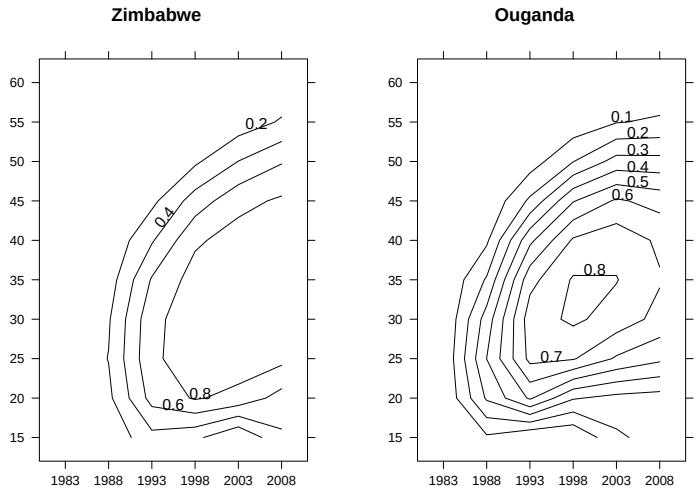

FIGURE 4.8 – *Proportions de décès féminins dûs au VIH-sida parmi l'ensemble des décès féminins - Ouganda et Zimbabwe - nos calculs menés à partir des variantes "moyennes" et "sans-sida" des Nations Unies.*
LECTURE : En 1998 en Ouganda, la proportion de décès dûs au sida à 15 ans est inférieure à 10%. A mesure que l'on monte dans les âges, cette proportion s'élève à 20%, puis 30% autour de 20 ans. Entre 35 et 35 ans, toujours pour l'année 1998, elle atteint même plus de 80%, pour baisser ensuite à moins de 20% à 50 ans.

Nous appliquons ensuite cette proportion aux **décès toutes causes** pour obtenir les décès dus au sida par groupe d'âge et intervalle quinquennal [23].

Tous les éléments permettent de procéder au calcul du nombre d'orphelins ayant perdu leur mère en raison du sida [24].

Il s'agit ensuite de calculer le nombre d'orphelins maternels dont la mère n'est *pas* décédée du sida. Comme le calcul précédent, celui-ci devrait idéalement intégrer également des risques de transmission verticale. En effet, les mères non décédées du sida peuvent avoir été infectées par le VIH, ce qui réduit le nombre d'orphelins car ceux-ci peuvent également avoir été infectés. Mais il faudrait pour cela connaître *la prévalence des mères qui ne décèdent pas du sida*, ainsi que la progression de ces mères à travers les différentes phases de la maladie (Grassly et Timaeus 2005). Ceci étant impossible, l'hypothèse est faite que les décès des mères non liés au sida

23. Ces décès sont cumulés par âge, lissés puis dé-cumulés, ce qui procure des estimations par année d'âge. Nous procédons de la même façon pour obtenir des valeurs par année calendrier. Cette procédure garantit qu'en agrégeant les résultats lissés, on récupère exactement les décès par groupe d'âge et intervalle quinquennal.

24. Ce nombre d'orphelins par âge exact se calcule en prenant la somme des valeurs de $\Omega'_{t,a,\tau}$ pour toutes les valeurs de τ. Le nombre d'orphelins par âge révolu ($\omega'_{t,a}$) se calcule ensuite en prenant la moyenne des estimations pour deux âges contigus (Grassly et Timaeus 2005) :

$$\omega'_{t,a} = \Omega'_{t,a} + \Omega'_{t,a+1} \tag{4.7}$$

surviennent uniquement chez des femmes non infectées. Dans ces conditions, le nombre d'orphelins d'âge a dont la mère n'est pas décédée du VIH-sida peut s'exprimer comme :

$$\Omega''_{t,a} = \sum_{\tau=0}^{a} \left[\sum_{i=15}^{49} \left(\mu''_{t-\tau,i+a-\tau} \times m_{i,0,t-a} \times l_a \right) \right] \tag{4.8}$$

Il s'agit en somme d'une version simplifiée de l'équation 4.6, où $\mu''_{t-\tau,i+a-\tau}$ renvoie au nombre de décès non dûs au sida survenus chez des femmes d'âge $i + a - \tau$ au temps $t - \tau$ [25].

Comme nous l'avons déjà indiqué, les dernières estimations du nombre d'orphelins sont fournies par l'UNICEF, pour l'année 2005 (UNICEF 2006). Elles reprennent notamment pour chaque pays le nombre d'enfants de moins de 18 ans dont la mère/le père sont décédés (toutes causes confondues). Elles sont obtenues au moyen de la méthode de Grassly et Timaeus (2005), mais avec peu de précisions quant aux données de base [26]. Le tableau 4.4 présente ces estimations UNICEF pour quelques pays, et les compare à nos estimations pour la même année.

	Orph. maternels	Orph. maternels	Orph. paternels	Orph. paternels
	Nos calculs	UNICEF(2006)	Nos calculs	UNICEF(2006)
Pays	(1)	(2)	(3)	(4)
Burkina Faso	329000	320000	586000	470000
Burundi	304000	310000	389000	400000
Congo RDC	2054000	2100000	3390000	2800000
Gabon	30000	32000	56000	41000
Ethiopie	2466000	2300000	3818000	3200000
Swaziland	64000	67000	91000	56000
Togo	107000	130000	203000	190000
Zambie	762000	860000	983000	800000

TABLE 4.4 – *Quelques estimations du nombre d'orphelins de mère/père en 2005 telles que nous les estimons à partir de la méthode de Grassly et Timaeus (2005), comparaison avec les estimations UNICEF.*

Ces différents pays ont été retenus ici car ils se distinguent notamment par l'importance numérique des orphelins (du Swaziland à l'Ethiopie), et par la proportion d'orphelins qui ont perdu leur mère à cause du sida. Cette proportion varie énormément, de moins de 10%, comme au Congo RDC, à plus de 70%, comme au Swaziland. Globalement, si l'on compare les effectifs absolus d'*orphelins de mère* (colonnes 1 et 2), nos calculs

25. L'hypothèse selon laquelle les décès qui ne sont pas dus au sida surviennent uniquement chez des femmes non infectées ne devrait pas avoir un impact majeur sur les estimations. Grassly et Timaeus (2005) pensent qu'elle entraîne une légère sur-estimation du nombre d'orphelins non dus au sida, mais dont l'ampleur serait limitée tout au plus à 5%.

26. Notamment, les hypothèses sur la surmortalité des orphelins, sur les courbes de survie après l'infection, ou sur les estimations des décès sont peu précisées. Comme de nouvelles estimations ou de nouveaux paramètres sont disponibles dans la littérature, il serait utile de connaître les choix faits en matière de données et d'hypothèses de base, par rapport à ceux retenus par Grassly et Timaeus (2005).

procurent des estimations comparables à celles de l'UNICEF, bien qu'il subsiste des écarts importants (au Togo, par exemple). Traduites en termes de proportions d'orphelins [27], nous obtenons des valeurs qui sont en majorité à moins de 20% de celles de l'UNICEF (graphe 4.9). Dans 7 pays, nos estimations sont inférieures de plus de 20% aux estimations officielles. C'est le cas des Comores, du Congo, de la Côte d'Ivoire, du Liberia, de la Namibie, du Togo, et du Zimbabwe. Les estimations de l'UNICEF (2006) semblent être calculées à partir de la révision 2004 des Nations Unies, et ceci pourrait expliquer les différences observées avec nos estimations. Rappelons en effet que la mortalité adulte a été sensiblement réduite entre la Révision 2004 et la Révision 2008 (Reniers, Masquelier, et Gerland forthcoming). Dans 5 pays, nos estimations surpassent les estimations officielles de plus de 20% (Guinée Equatoriale, Guinée, Gambie, Nigeria, Tanzanie).

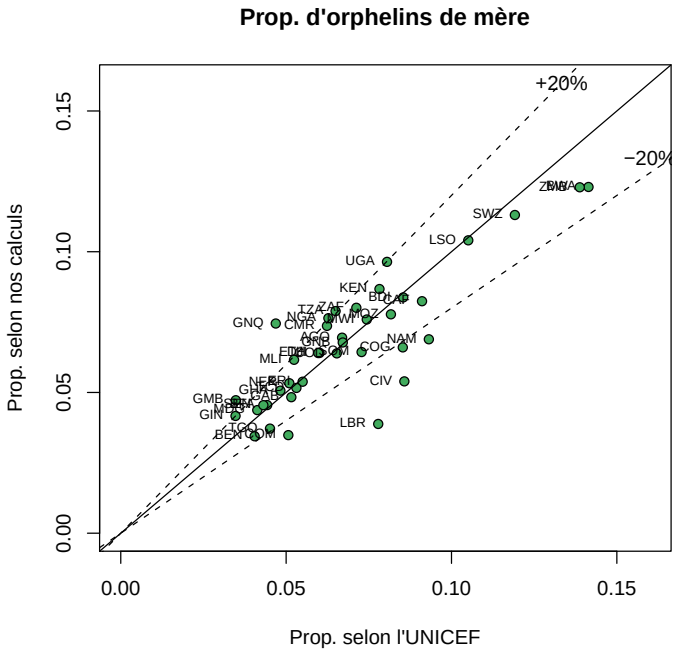

FIGURE 4.9 – *Proportions d'orphelins de mère estimées en appliquant la méthode de Grassly et Timaeus (2005) (nos calculs), et proportions estimées par l'UNICEF pour 2005.*

En dépit de ces écarts avec les estimations officielles, nous pensons que la grande majorité des proportions d'orphelins que nous obtenons sont plausibles. Elles tiennent compte de la surmortalité des orphelins, ainsi que des différentes perturbations induites par le sida. Nous allons les utiliser plus loin dans le but de proposer de nouveaux coefficients pour l'estimation de la mortalité adulte à partir de la survie des parents.

Avant cela, nous cherchons à obtenir également des estimations pour les orphelins de père. L'une des motivations principales de tous ces calculs est en effet de pouvoir proposer des coefficients relatifs à la *mortalité masculine*, ce qui n'existe pas à ce jour pour les contextes perturbés par le VIH.

27. En utilisant les effectifs d'enfants estimés par les Nations Unies pour 2005.

Nombre d'orphelins de père

Le calcul du nombre d'orphelins de père est nettement plus compliqué que le calcul des orphelins de mère, et nous ne l'abordons ici que rapidement. La difficulté réside dans le fait qu'il doit intégrer des informations sur *l'étape atteinte par les mères dans la progression de la maladie* lors de la naissance des enfants, en raison des risques de transmission verticale qui vont rehausser la probabilité de décès. C'est précisément ce qui manque dans nos microsimulations.

Notons $\Psi'_{t,a,\tau}$ le nombre d'orphelins d'âge a au temps t dont le père est décédé du sida il y a τ années. Il s'agit du pendant du même nombre pour les mères, $\Omega'_{t,a,\tau}$ (p. 211). Ce nombre dépend de plusieurs éléments, que nous ne faisons ici que mentionner :

- des décès masculins dus au sida, distribués par âge (i) et année (t), soit $\nu'_{t,i}$ (ce nombre est calculé comme l'a été $\mu'_{t,i}$ pour les femmes),
- de la probabilité, variable par année, qu'une femme enceinte soit à l'étape s de la progression de la maladie si son partenaire est à l'étape r de la progression, soit $P(s|r,t)$,
- de la proportion d'hommes à l'étape r de l'infection x années avant le décès, soit $Y_r(x)$ (calculé comme $Y_s(x)$ avec les paramètres relatifs aux hommes),
- de la fécondité des pères non infectés par âge et année, soit $\phi_{i,t}$ [28],
- enfin, comme précédemment pour les femmes, d'une moyenne des probabilités de survie des orphelins séropositifs et séronégatifs, pondérées par une probabilité de transmission verticale, soit $\xi_s l'_a + (1-\xi_s) l_a$.

Tous ces éléments se combinent dans l'équation 4.9 :

$$\Psi'_{t,a,\tau} = \sum_{i=15}^{79} \nu'_{t-\tau,i+a-\tau} \sum_{r=0}^{n} \left[Y_r(a-\tau) \times \sum_{s=0}^{n} P(s|r,t-a) k_{r,s} \phi_{i,t-a} \left(\xi_s l'_a + (1-\xi_s) l_a \right) \right] \tag{4.9}$$

Nous n'entrerons pas ici dans les détails de l'estimation par souci de concision. Nous avons suivi au plus près la méthode de Grassly et Timaeus (2005), et les principaux choix méthodologiques que nous avons fait ont déjà été mentionnés plus haut.

Précisons simplement comment est estimée la probabilité $P(s|r,t)$. Il s'agit donc la *probabilité qu'une femme enceinte soit à l'étape s de la progression de la maladie selon la stade atteint par son partenaire*. Grassly et Timaeus (2005) utilisent une régression logistique qui exprime la concordance du séro-statut des partenaires (donc le fait que les deux partenaires soient infectés ou non) comme une fonction de la prévalence observée, telle qu'elle

28. Nous utilisons, comme le suggèrent Grassly et Timaeus (2005), le standard de Paget-Timaeus, présenté en page 132. Notons qu'idéalement, il faut également tenir compte des changements dans la fécondité masculine provoqués par l'infection au VIH. On suppose ici que l'infection n'a pas d'impact sur la fécondité des hommes.

peut être estimée à partir des centres pré-nataux [29]. Leur équation prend la forme suivante [30] :

$$\P(s > 0, r > 0) = \frac{exp(\alpha + \beta \times HIVprev)}{1 + exp(\alpha + \beta \times HIVprev)} \tag{4.10}$$

Ce sont les résultats de cette régression qui permettent de tenir compte des chances qu'une femme soit infectée si son partenaire est infecté [31].

Les nombres absolus d'orphelins de père ont été comparés à ceux estimés par l'UNICEF dans le tableau précédent (repris en p. 213). Le graphe 4.10 compare les proportions d'orphelins de père que nous avons calculées et celles estimées par l'UNICEF pour 2005. A la différence du cas des mères, on observe ici d'importants écarts avec les estimations officielles. Nos estimations sont en moyenne 25% supérieures aux estimations de l'UNICEF. Les différences s'observent tant dans les pays à forte prévalence que là où le VIH-sida ne joue pas un rôle important. A ce stade, nous n'avons pas trouvé d'explication plausible à ces différences, et ce point mériterait d'avantage d'attention.

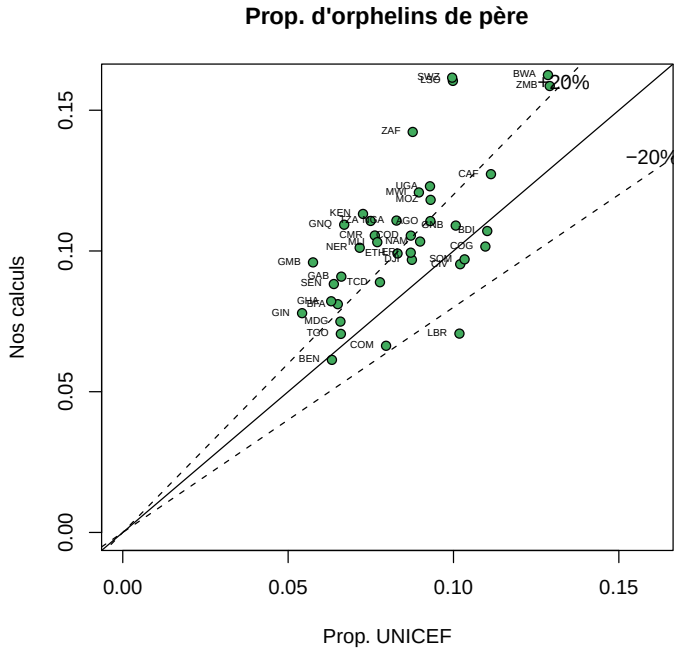

FIGURE 4.10 – *Comparaison de nos proportions d'orphelins de père estimées en appliquant la méthode de Grassly et Timaeus (2005), et des proportions estimées par l'UNICEF pour 2005.*

29. Ils s'en tiennent à deux valeurs de s, en distinguant les femmes séropositives et séronégatives uniquement.

30. Les estimateurs de maximum de vraisemblance valent -0.839 pour α et 9.134 pour β.

31. L'autre innovation principale de leur travail est qu'ils développent ensuite un modèle multi-niveau qui permet d'estimer le nombre de doubles orphelins, à partir de l'observation des enquêtes EDS. Dans notre cas, le nombre de doubles orphelins n'a pas d'intérêt en soi.

4.3 Retour sur les probabilités de décès

A partir de ces estimations, nous pouvons calculer de nouveaux coefficients, qui devraient permettre de convertir les proportions de parents survivants en indices de mortalité. *L'idée consiste ici à régresser les niveaux de mortalité estimés par les Nations Unies (${}_nq_{25}$) sur les proportions de non-orphelins calculées au niveau macro, en tenant compte de la prévalence du VIH.*

Le graphe 4.11 oppose les probabilités de survie féminines ${}_{10}q_{25}$ estimées par la Division de la Population des Nations Unies (UN 2009c), avec les proportions ${}_5S_{n-5}$ que nous venons de calculer par la méthode de l'UNICEF [32]. Dans ce graphe, on constate que la relation entre les deux séries est quasiment linéaire quand la probabilité de survie est élevée (supérieure à 0.9). C'est le cas de la plupart des points repris en vert, qui font référence aux estimations pour lesquelles la prévalence du VIH-sida étaient inférieure à 1% au moment de la naissance des enfants. Là où la prévalence est plus élevée, la tendance est quelque peu modifiée, et la pente de la droite qu'il est possible de tracer dans le nuage de points est plus "raide" que celle qui prévaut à des niveaux de mortalité plus bas. On retrouve donc ici le constat fait par Timaeus et Nunn (1997), qui a motivé le développement de leurs nouveaux coefficients. Nous pouvons faire la même observation à partir de la survie des hommes (graphe 4.12).

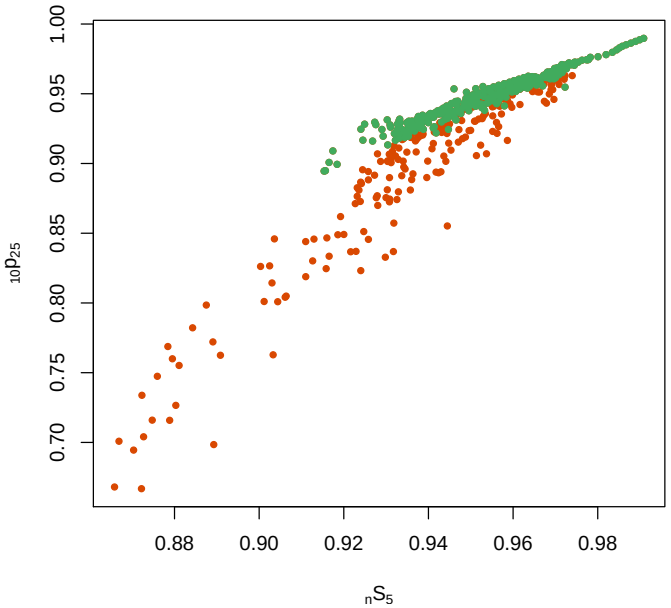

Figure 4.11 – *Rapport entre les probabilités de survie ${}_{10}q_{25}$ estimées par les Nations Unies, et les proportions ${}_5S_{n-5}$ issues des **macro**-simulations - **Mortalité féminine***

32. Pour pouvoir comparer les proportions d'orphelins aux probabilités de survie, il nous faut fixer une période de référence. Nous recourons donc aux *cohortes synthétiques*, afin d'évacuer le problème de la datation.

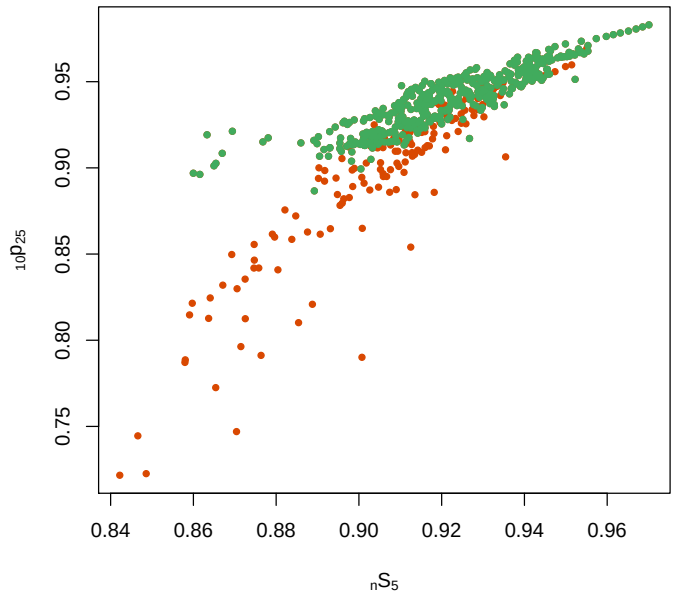

FIGURE 4.12 – *Rapport entre les probabilités de survie ${}_{10}q_{25}$ estimées par les Nations Unies, et les proportions ${}_5S_{n-5}$ issues des **macro**-simulations - **Mortalité masculine***

4.3.1 Calcul de nouveaux coefficients

A partir de là, nous optons mettre les proportions de non-orphelins en vis-à-vis des probabilités de survie pour estimer de nouveaux coefficients. Pour ce faire, nous optons pour un modèle de régression assez simple. L'équation de la régression est proche de celles utilisés auparavant (par exemple chap.3, p. 111), mais nous y ajoutons des informations sur la *prévalence atteinte au moment de la naissance des enfants*. En effet, dans les deux graphes 4.11 et 4.12, l'ampleur avec laquelle le nuage de points s'écarte de la droite esquissée par les estimations où la mortalité est faible (les points en vert) est largement déterminée par la prévalence. En d'autres termes, plus la prévalence est forte, plus la probabilité de survie sera faible pour une même proportion de parents survivants. Nous introduisons en fait deux niveaux de prévalence pour une valeur donnée de n : celle observée il y a n années et celle observée il y a $n-5$ années. Par exemple, chez les 5-9 ans, nous retenons la prévalence observée il y a 10 ans et celle observée il y a 5 ans. Ceci vise à tenir compte, même approximativement, de la progression plus ou moins rapide de l'épidémie du VIH-sida [33]. L'équation prend donc la forme suivante :

$$ {}_np_{25} = \beta_0 + \beta_1 \times M + \beta_2 \times {}_5S_{n-5} + \beta_3 \times HIV_n + \beta_3 \times HIV_{n-5} \qquad (4.11) $$

La prévalence HIV_n est celle observée chez les femmes il y a n années.

L'une des difficultés des méthodes existantes réside dans le choix des coefficients à utiliser : nous avons vu en effet que les coefficients classiques

33. Notons également que les proportions de non-orphelins qui sont reprises dans la régression ne sont pas corrigées pour tenir compte des biais de sélection (puisque cette correction dépend elle-même de la prévalence).

procurent des niveaux de mortalité très différents de ceux obtenus par le biais des coefficients de Timaeus et Nunn (1997). Mais quand faut-il opter pour l'une ou l'autre option ? Dans les situations limites, proches du seuil de prévalence qui ferait passer de l'une des options à l'autre (1, 3, 5% ?), on risque d'obtenir des "cassures" dans les tendances. Tandis qu'en introduisant directement la prévalence dans la régression, la transition se fait de façon plus souple entre les périodes à prévalence faible et les périodes où les coefficients classiques ne sont plus adaptés. Le tableau 4.5 présente les coefficients que nous proposons pour estimer la mortalité féminine à partir des informations fournies par les 5-9 et les 10-14 ans.

n	β_0	β_1	β_2	β_3	β_4
10	-0.202	0.001	1.181	-0.069	-0.47
15	-0.278	0.001	1.276	0.52	-0.952

TABLE 4.5 – *Coefficients de corrections provisoires proposés pour convertir les proportions de non-orphelins en probabilité de survie - **Mortalité féminine***

La même équation de régression est utilisée pour les hommes. Le tableau 4.6 reprend les coefficients de régression correspondants [34].

n	β_0	β_1	β_2	β_3	β_4
10	-0.076	0.004	0.954	-0.018	-0.425
15	-0.28	0.009	1.016	0.354	-0.829

TABLE 4.6 – *Coefficients de correction proposés pour convertir les proportions de non-orphelins en probabilité de survie - **Mortalité masculine***

4.3.2 Application aux microsimulations

Il est difficile de valider ces coefficients sans disposer de données fiables (du type de celles fournies par les sites d'observation démographique, par exemple). Nous pouvons les appliquer aux microsimulations, mais ceci ne constitue pas un véritable test de validation, car les estimations micro et macro ne sont pas indépendantes ; ce sont les projections des Nations Unies qui sont à la base des calculs dans les deux cas. Nous avons tout de même mené les calculs sur les populations simulées pour lesquelles la prévalence du VIH-sida a atteint des niveaux importants. Ceci est présenté dans le graphe 4.13 pour l'Ouganda, la Zambie, le Lesotho, et le Zimbabwe. En plus des probabilités de décès issues des microsimulations, ce graphe oppose les estimations obtenues à partir des coefficients classiques (en vert), celles obtenues au départ des coefficients de Timaeus et Nunn (1997) (en bleu), et enfin, celles déduites de l'application des coefficients que nous proposons (en orange). Dans les quatre pays considérés ici, nos nouveaux coefficients procurent des niveaux de mortalité bien plus proches que ceux obtenus autrement [35].

34. L'âge moyen à la naissance des enfants est estimé pour les pères à partir des taux de fécondité masculine utilisés pour générer les proportion d'orphelins, et des structures par âge des hommes estimés par les Nations Unies.

35. Ce n'est pas le cas dans les pays où la prévalence est faible, et le recours à ces coefficients devrait donc être limité aux situations extrêmes où le sida joue un rôle déterminant dans les tendances.

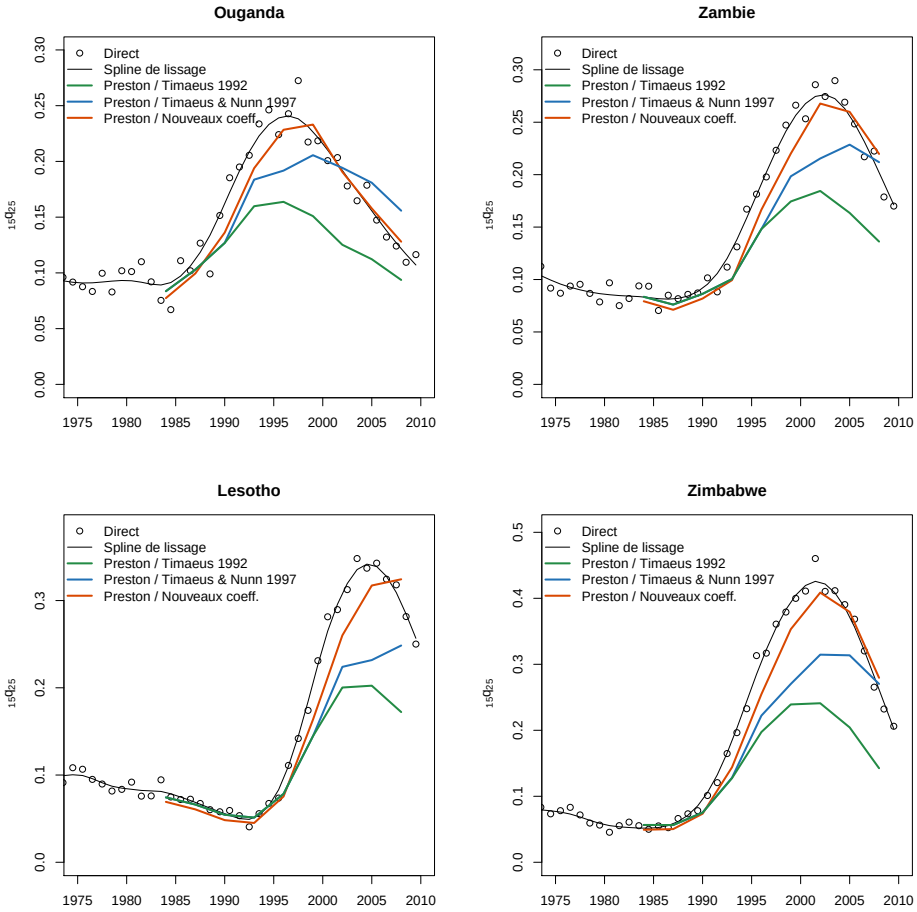

FIGURE 4.13 – *Probabilités de décès $_{15}q_{25}$ (féminines) calculées directement et estimées à partir de la survie des mères - Ouganda, Lesotho, Zambie, Zimbabwe. Moyenne prise sur les estimations calculées à partir des 5-14 ans. Comparaison des estimations obtenues classiquement avec celles obtenues par le biais des coefficients de Timaeus et Nunn (1997) et avec nos propres coefficients.*

4.3.3 Application aux enquêtes EDS

Enfin, nous appliquons ces coefficients aux enquêtes EDS, en combinant par pays les enquêtes successives pour pouvoir recourir à des cohortes synthétiques. Le graphe 4.14 oppose les valeurs obtenues avec les coefficients de Timaeus et Nunn (1997) à celles calculées à partir de nos coefficients, pour les 5-9 ans et les 10-14 ans. Seules les quelques enquêtes où la prévalence était supérieure à 3% (entre les deux enquêtes) sont retenues. On s'en tient ici à la mortalité féminine. On constate que les coefficients que nous proposons procurent en fait des niveaux de mortalité comparables à ceux obtenus par la méthode de Timaeus et Nunn (1997). Comme dans les microsimulations, la mortalité estimée à partir des 10-14 ans est peut-être légèrement supérieure (et inférieure à partir des 5-9 ans).

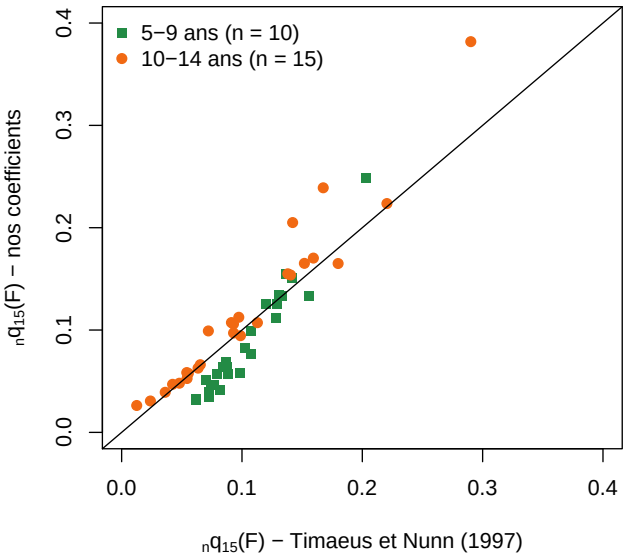

FIGURE 4.14 – *Probabilités de survie féminines ($_{10}q_{25}$ et $_{15}q_{25}$) estimées à partir de la méthode de Timaeus et Nunn (1997), et à partir de nos coefficients en faisant intervenir la prévalence du VIH*

S'ils procurent des estimations proches de celles obtenues par le biais des coefficients de Timaeus et Nunn (1997), leur intérêt est surtout qu'ils permettent aussi de générer des estimations de *mortalité* masculine dans les contextes perturbés par le sida, ce qui n'était pas possible jusqu'alors. Ces coefficients sont toutefois à prendre avec prudence, en raison du fait que les proportions d'orphelins de père que nous obtenons sont nettement supérieures à celles de l'UNICEF (*cf.* p. 216), sans que nous ayons pu à ce stade identifier les causes de ces différences.

Conclusion

Ce chapitre a permis de mettre en évidence les principaux problèmes méthodologiques posés à la méthode des orphelins par le VIH. Nous avons également procédés aux calculs mis aux points par Grassly et Timaeus (2005), afin de générer des proportions d'orphelins qui nous paraissent fiables. Celles-ci nous ont permis ensuite de proposer de nouvelles séries de coefficients à appliquer aux informations fournies par les 5-9 ans et les 10-14 ans - typiquement ce que fournissent les enquêtes EDS. Ces nouveaux coefficients sont à appliquer en recourant à la méthode des cohortes synthétiques et en faisant intervenir des informations sur la prévalence atteinte.

La démarche utilisée ici reste très exploratoire, notamment parce que nous utilisons des microsimulations pour valider nos coefficients alors que les deux séries de simulations ne sont nullement indépendantes. Dans les deux cas, ce sont les estimations des Nations Unies qui sont à la base des calculs. Il faut donc insister sur le fait que les coefficients proposés le sont à titre provisoire, surtout pour la survie des pères. On retiendra surtout de ce chapitre la conclusion suivante : *il y a moyen d'amender la méthode des orphelins pour la rendre applicable même aux situations les plus extrêmes, caractérisées par une mortalité due au sida très élevée.* L'une des voix à creuser davantage est le recours à des macrosimulations adaptés à l'estimation du nombre d'orphelins dans les contextes perturbés par le sida, comme nous le proposons ici.

En définitive, notre conviction est que les principales limites des méthodes basées sur la survie des parents se situent au niveau de la disponibilité des données - et de leur qualité -, pas au niveau des méthodes qui sont utilisées pour les exploiter.

Ceci conclut la partie consacrée à la méthode des orphelins. Le chapitre suivant est consacré à la seconde grande catégorie de données, qui portent sur les frères et sœurs.

Estimer la mortalité adulte à partir de la survie des frères et sœurs

5

Sommaire

5.1 Estimation des taux de mortalité en biographique (approche directe) 225
5.2 Biais de sélection 234
5.3 Exploitation indirecte des informations portant sur la survie des frères et sœurs 267

Par rapport aux données sur la survie des parents, les informations qui portent sur la survie des frères et sœurs présentent de nombreux avantages. Leur principal intérêt est qu'elles peuvent être *exploitées de façon directe*. En effet, elles se présentent généralement sous la forme d'une histoire génésique complète de la mère des enquêtés. En d'autres termes, on dispose donc des âges des frères et sœurs et de leurs âges au décès. L'exploitation directe que cela autorise a au moins trois avantages par rapport au calcul indirect :

1. la période d'exposition peut être délimitée en fonction des *objectifs* et de la *qualité* des données. Par exemple, les informations les plus anciennes, qui sont souvent les plus suspectes, peuvent être écartées. A l'inverse, l'exploitation indirecte va mêler les informations anciennes et récentes. Elle est donc sensible aux fréquentes omissions relatives aux événements lointains, ce qui est moins le cas de l'approche directe.
2. les estimations directes autorisent un examen des tendances de la mortalité plus aisé qu'avec les méthodes indirectes. En effet, dans le cas d'une exploitation indirecte, rappelons que les différentes périodes de référence sont estimées à partir de différents âges des enquêtés.
3. l'exploitation directe permet de procéder à des tests relatifs à la *qualité des données* souvent plus complets. C'est notamment le cas quand plusieurs enquêtes successives sont disponibles. Les données

peuvent alors être analysées dans un même modèle de régression, ce qui permet d'estimer si les décès sont davantage sous-déclarés dans le passé plus lointain (*cf*. infra).

4. Enfin, la *modélisation* des tendances ainsi que des schémas par âge de la mortalité est plus souple dans le cas de l'exploitation directe. Cette modélisation reste souvent nécessaire, parce que les estimations "brutes" sont trop erratiques (en raison du faible nombre de décès par année et par âge), mais elle est peu contraignante. Par exemple, à l'aide d'un modèle de régression, on pourra recourir à de nombreuses options pour saisir l'évolution de la mortalité dans le temps. A l'inverse, dans le cas d'une exploitation indirecte, l'hypothèse d'une évolution approximativement linéaire de la mortalité est toujours de mise.

L'exploitation *indirecte* est incontournable lorsque seules les *proportions de survivants* sont disponibles (comme c'est le cas de certaines enquêtes MICS, par exemple). Pour le reste, il nous semble préférable de privilégier l'approche directe. Car recourir à l'exploitation indirecte revient un peu à passer par une "boîte noire" qui va rendre l'interprétation des résultats et l'évaluation de leur fiabilité assez délicate. Cela dit, comme souvent, plutôt que d'opposer les deux types d'exploitations des données, il est préférable de confronter leurs résultats. En cas de mauvaise qualité des données, les estimations indirectes peuvent s'avérer plus plausibles pour au moins deux raisons :

1. elles ne sont pas sensibles aux erreurs de *datation* des décès
2. elle ne sont pas sensibles aux erreurs et aux attirances sur les *âges* des frères et sœurs et sur leurs âges au décès [1]. Elle est par contre sensible aux erreurs sur les âges des enquêtés.

Dans ce chapitre, nous présenterons donc successivement les deux types d'approches. On trouvera également dans le chapitre 7 des exemples de comparaison entre exploitation directe et indirecte.

1. A ceci près qu'il faut parfois pouvoir cibler les frères et sœurs (ou uniquement les sœurs dans le cas de l'estimation de la mortalité maternelle) qui ont survécu à quinze ans.

5.1 Estimation des taux de mortalité en biographique (approche directe)

Plusieurs options sont possibles lorsque l'on souhaite exploiter les données sur les frères et sœurs de façon directe. Par exemple, les rapports d'enquête EDS optent pour une approche comparable à celle utilisée pour les *taux de mortalité infanto-juvénile*[2] ; les taux sont calculés par période de 0-6, 7-13, et 0-13 ans avant l'enquête et par groupe d'âge quinquennal (Rutstein et Rojas 2006). Le taux de mortalité de 15 à 49 ans s'obtient ensuite en prenant la somme des taux par âge, après les avoir pondérés par la structure par âge des *enquêtés* - pas par celle des frères et sœurs.

Si c'est la structure par âge des enquêtés qui est utilisée, c'est parce que celle des frères et sœurs est tronquée ; les plus jeunes (les 15-19 ans) peuvent uniquement être mentionnés par leurs frères et sœurs au moins aussi *âgés* (de 15 à 49 ans), tandis que les plus âgés (au delà de 45 ans) sont mentionnés uniquement par leurs plus *jeunes* frères et sœurs. Par contraste, les 30-34 ans sont mentionnés par leurs plus jeunes comme par les plus vieux frères et sœurs. En conséquence, les groupes d'âge 15-19 et 45-49 sont sous-représentés parmi l'ensemble des frères et sœurs mentionnés dans l'enquête.

Ceci est illustré dans le graphe 5.1, avec l'exemple de l'enquête EDS menée à Madagascar en 2009. La répartition par âge des *sœurs des enquêtées* présente une proportion de 15-19 ans nettement moins élevée que celle observée dans la répartition par âge *des enquêtées*, ou que celle que l'on peut calculer à partir des distributions par âge de la population fournies par les Nations Unies (UN 2009d).

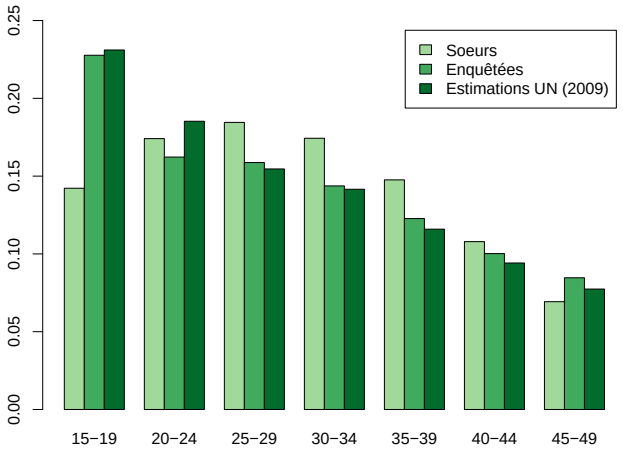

Figure 5.1 – *Répartition par âge au sein du groupe des 15-49, sœurs des enquêtées, enquêtées et estimations des Nations Unies pour 2008 - EDS Madagascar 2009*

Depuis le début des années 1990, les rapports d'enquête EDS pré-

2. Dans le cas de la mortalité des enfants, il s'agit de calculer des probabilités de survie sur de courts intervalles d'âge de 0 à 5 ans (0, 1-2, 3-5, 6-11, 12-23, 24-35, 36-47, et 48-59 mois) à partir de l'expérience réelle des cohortes puis de les combiner en des segments d'âge plus larges.

sentent des taux de mortalité adulte calculés de cette façon. Bicego (1997) a proposé une première synthèse des estimations dégagées des frères et sœurs pour l'Afrique subsaharienne, à partir de 6 enquêtes menées entre 1992 et 1996. Il considère que ces estimations sont très plausibles. Il remarque en outre que dans plusieurs enquêtes de cet échantillon, les différentiels régionaux de mortalité cadrent bien avec d'autres sources externes.

Plutôt que d'opter pour ce calcul utilisé dans les EDS, l'exploitation des données d'enquête peut se faire de façon plus souple en optant pour une approche biographique [3]. Il s'agit de reconfigurer les fichiers en *personnes périodes*. Ceci consiste à séparer chaque observation individuelle en autant de périodes de vie vécues par les frères et sœurs et définie par des âges et des périodes différentes (Schoumaker 2004). Le calcul des taux peut se faire simplement en sommant l'exposition totale et le nombre de décès dans chaque intervalle pour l'ensemble de la population, après avoir tenu compte des structures par âge. Il est également possible de recourir à des modèles de régression en temps discret, afin d'intégrer des variables qui varient dans le temps et de mener des analyses par sous-groupes. Cette seconde approche est préférable car elle permet d'"emprunter de l'information" entre les différentes enquêtes, notamment en ce qui concerne les schémas par âge de la mortalité et son évolution dans le temps.

Nous détaillons cette approche dans la section suivante, en optant pour la régression de Poisson.

5.1.1 Régression de Poisson

La régression de Poisson compte parmi les modèles les plus utilisés pour exploiter les données biographiques. Elle est appropriée lorsque les événements sont rares et surviennent au sein de larges échantillons. Elle se prête bien à l'analyse des données en personnes-périodes (Cameron et Trivedi 1998, Loomis et al. 2005, Schoumaker 2004). Régulièrement utilisée par les épidémiologues, les économistes et les démographes, elle a déjà été appliquée aux données EDS sur les frères et sœurs par Timaeus et Jasseh (2004) ainsi que Reniers, Masquelier, et Gerland (forthcoming).

Il existe bien entendu de multiples autres options, mais la régression de Poisson présente au moins deux avantages :

1. les calculs peuvent être menés au niveau individuel comme au niveau agrégé sans que les coefficients ne soient modifiés [4],
2. l'exponentielle des coefficients procure directement des rapports de taux (et l'exponentielle de la constante fournit le taux de base). Les taux pour différents groupes et périodes s'obtiennent donc facilement.

3. L'analyse biographique (*event history analysis*) connaît de nombreuses autres dénominations : analyse de survie (*survival analysis*), analyse des transitions (*transition analysis*),...

4. La somme de plusieurs variables distribuées selon une loi de Poisson est également distribuée selon une loi de Poisson, et le paramètre qui résulte de cette somme correspond à la somme des paramètres des lois qui la composent : si $Y_1 \tilde{} P(\mu_1)$ et $Y_2 \tilde{} P(\mu_2)$, alors $Y_1 + Y_2 \tilde{} P(\mu_1 + \mu_2)$

Le modèle de régression exprime le risque de décès comme une fonction de la durée d'exposition et d'un vecteur de caractéristiques (individuelles ou contextuelles). Cela est détaillé dans l'encadré 10.

Encadré 10 La régression de Poisson

La régression de Poisson compte parmi différentes variantes du modèle linéaire généralisé. Comme son nom l'indique, ce modèle étend le modèle linéaire au-delà de la linéarité et introduit des distributions non-normales. Il se caractérise par trois éléments (McCullagh et Nelder 1989) :

1. par une *composante aléatoire* : le vecteur d'observations y est la réalisation d'un vecteur de variables aléatoires Y distribuées de façon indépendante avec une moyenne notée μ. La distribution est issue de la famille exponentielle (normale, Poisson, binomiale, Gamma...)
2. par une *composante systématique* qui dépend des caractéristiques introduites ; le vecteur de μ peut s'exprimer par le biais d'un vecteur de caractéristiques x_k et un vecteur de coefficients β_k (c'est-à-dire un prédicteur linéaire, noté η) :
$$E(Y_i) = \sum_{k=1}^{K} (\beta_k x_{ki}) \tag{5.1}$$
3. par une *fonction de lien* entre la moyenne et le prédicteur linéaire de type :
$$\eta_i = g(\mu_i) \tag{5.2}$$

Au sein de cette famille de modèles de régression, la régression de Poisson se caractérise par :

1. une *composante aléatoire* qui prend la forme d'une structure des erreurs suivant une loi de Poisson. La fonction de densité de cette distribution s'exprime comme ceci :
$$P(y_i) = \frac{e^{-\lambda_i t_i} (\lambda_i t_i)^{y_i}}{y_i!} \tag{5.3}$$
La moyenne et la variance de cette distribution sont identiques et valent $\mu_i = \lambda_i t_i$, soit le produit du taux et de la durée d'exposition. La régression de Poisson s'applique sur des données de comptage (des nombres entiers non négatifs).
2. une *fonction de lien* logarithmique : $\eta_i = ln(\mu_i)$. Cette fonction de lien garantit le caractère positif de μ_i. Elle implique que les variables ont des effets multiplicatifs sur les taux - et non pas additifs[5]. L'interprétation des coefficients de régression est donc différente de celle qui caractérise la régression logistique : plutôt que d'obtenir des *odds ratios*, la régression de Poisson procure des rapports de taux.

La régression de Poisson suppose notamment que soient remplies les trois conditions suivantes :

1. la probabilité d'observer plus d'un événement dans un intervalle très court est approximativement nulle.

5. Puisque l'exponentielle d'une somme correspond au produit des exponentielles des différents termes de cette somme.

2. les événements surviennent à un taux λ invariable dans le temps (on parle de condition d'*homogénéité*). En d'autres termes, les variables explicatives conservent le même effet durant toute la durée. Cela ne pose pas de problème lorsque le fichier est restructuré en données biographiques, puisqu'il s'agit de périodes de courte durée. Par contre, pour tenir compte des durées d'exposition variables, il faut introduire dans le modèle le logarithme de la durée d'exposition de chaque observation (l'*offset*). Afin de conserver un risque proportionnel à la durée, son coefficient est fixé à l'unité. Cette variable n'est pas testée dans le modèle.
3. La probabilité d'une nouvelle occurrence est indépendante du temps écoulé depuis la dernière occurrence. Elle survient selon une distribution exponentielle constante (le temps moyen entre deux occurrences valant l/λ). En d'autres termes, il ne peut pas y avoir de "contagion" d'une occurrence à l'autre. En présence de contagion positive, le taux va augmenter avec le nombre d'événements antérieurs. Or la moyenne conditionnelle et la variance conditionnelles sont supposées égales : $E(Y) = Var(Y) = \mu$ (c'est la propriété dite d'*équidispersion*).

La principale limite de la régression de Poisson tient à l'hypothèse d'*équidispersion* ; la variance de la variable dépendante est supposée égale à sa moyenne. Cela suppose que le modèle ne laisse pas de place à l'existence de variables non observées. Or il arrive souvent que le modèle soit sous-spécifié et qu'une hétérogénéité non observée subsiste, ce qui va induire une variance plus élevée qu'attendue. Lorsque la variance réelle est supérieure à la moyenne - et donc à la variance supposée - les données présentent un phénomène appelé *surdispersion*. Cette surdispersion va compromettre les inférences au même titre que l'hétéroscédasticité présente dans un modèle linéaire[6]. Les *coefficients* sont corrects mais les *erreurs standards* sont trop faibles.

Afin de contrôler cette surdispersion, un modèle de *quasi-Poisson* peut être adopté (McCullagh et Nelder 1989). Il s'agit d'un modèle moins contraignant dans lequel un paramètre de dispersion est introduit sous une forme multiplicative : $Var(Y) = \varphi\mu$. Il sera estimé au cours de l'ajustement du modèle. Les coefficients de régression restent identiques, seule leur erreur standard va varier[7].

Dans le chapitre 7, où nous présentons les estimations basées sur la survie des frères et sœurs, nous n'encadrons pas les estimations directes

6. Lorsque les résidus ne sont pas distribués de façon homogène pour toutes les valeurs des variables explicatives.

7. Cet indice de dispersion peut être estimé en rapportant la somme des carrés des résidus de Pearson aux degrés de liberté. Notons qu'une option alternative pour capturer la surdispersion consiste à conserver une distribution de Poisson et à intégrer un paramètre de dispersion plutôt de façon additive, sous forme d'un effet aléatoire. L'analyse est menée dans un cadre multi-niveau, mais sans intégrer de niveau supérieur. Le niveau introduit reproduit en fait le premier niveau d'observations (Browne et al. 2005). Dans le cas où le rapport entre la variance et la moyenne n'est pas linéaire, le recours à une modélisation par la *loi binomiale négative* est envisageable également. D'autres alternatives sont possibles, notamment le recours aux modèles dits *zero-inflated*.

d'intervalles de confiance. Ce choix est motivé par l'absence de méthodes permettant d'estimer de tels intervalles de confiance (ou de crédibilité si l'on se place dans un cadre bayésien) autour des estimations indirectes, qu'elles soient issues des données sur la survie des parents ou des frères et sœurs. Par contre, nous avons calculé ces intervalles à l'aide d'un modèle de Quasi-Poisson pour encadrer les estimations présentées dans Reniers, Masquelier, et Gerland (forthcoming).

5.1.2 Modélisation des tendances

Après avoir présenté en quelques mots la régression de Poisson, nous détaillons ici le modèle de régression que nous avons retenu pour exploiter les données sur les frères et sœurs. Il s'agit du modèle développé par Timaeus et Jasseh (2004). Ces auteurs ont mené il y a quelques années une analyse systématique de la mortalité adulte africaine à partir des enquêtes EDS. Dans un travail plus récent, nous avons adopté leur modèle pour l'ajuster sur toutes les enquêtes EDS disponibles (Reniers, Masquelier, et Gerland forthcoming), y compris à partir des informations fournies par les *hommes* sur leurs frères et sœurs. Ce travail nous a également permis de discuter des biais de sélection, ce que nous abordons dans la section suivante. Les résultats de l'ajustement de ce modèle sont détaillés dans le chapitre 7, consacré aux niveaux de mortalité. Nous présentons ici ce modèle dans ses grandes lignes et nous renvoyons à l'article de Timaeus et Jasseh (2004) pour davantage de détails.

Modèle (adapté) de Timaeus et Jasseh (2004)

Le modèle est ajusté sur les décès calculés par année d'âge (x), sexe (g), pays (i) et année calendrier (t). Il s'agit ici d'une version légèrement différente de celle de Timaeus et Jasseh (2004)[8]. Ce modèle peut être exprimé sous la forme suivante :

$$\begin{aligned} ln(\mu(x,g,i,t)) = {} & \beta_0 + \beta_1(g,i) + \beta_2(g,i)t \\ & + \beta_3(i) ln(\mu_s(\bar{x},g)) \\ & + \left\{ (\beta_4(i) + \beta_5(\bar{x},g))\,(t - T_i) + \beta_6(i)(t - T_i)^2 I(i \in S) \right\} \\ & I(t > T_i) I(x \geq 20) \end{aligned} \tag{5.4}$$

Les différentes variables explicatives introduites sont les suivantes :

1. Le *niveau de mortalité* varie par pays, tout comme le *différentiel sexuel* de mortalité. Il s'agit du coefficient $\beta_1(g,i)$. Ces deux éléments varient également dans le temps, selon une tendance log-linéaire[9], et ce, de façon spécifique à chaque pays ($\beta_2(g,i)t$).
2. En dépit du fait que les données sur les frères et sœurs fournissent des milliers de décès et de personnes-années, la variabilité aléatoire

8. Dans leur modèle, le différentiel sexuel est constant et la structure par âge de la surmortalité liée au VIH-sida est commune à tous les pays. Ici, le différentiel sexuel varie dans le temps et la structure par âge de la surmortalité liée au VIH-sida est spécifique à chaque sous-région.

9. En raison de la fonction de lien logarithmique de la régression de Poisson.

reste importante. Il peut donc être utile de "lisser" les taux de mortalité en introduisant dans le modèle de régression un schéma de mortalité standard. Il s'agit ici du terme $\beta_3(i)ln(\mu_s(\bar{x},g))$. Les coefficients de régression traduisent l'ampleur avec laquelle les schémas 'empiriques' diffèrent de ce standard. Ces coefficients sont considérés comme fixes par pays et par sexe, et il sont censés procurer le schéma de mortalité "sans sida". Le standard introduit est donc un schéma classique, en l'occurence le modèle général développé par les Nations Unies pour les pays en développement (UN 1982). La combinaison d'un schéma fixe par pays et d'une niveau de mortalité changeant revient en quelque sorte à recourir à une table type de mortalité à deux entrées. C'est ce que notent Timaeus et Jasseh (2004) :

> "The non-AIDS component of the estimates of mortality can be considered as coming from fitted two-parameter relational model life tables in which the level of mortality changes at a different rate in each country, but the age pattern of mortality varies among countries and not over time within each country" (Timaeus et Jasseh (2004), p. 761)

Nous introduisons ce standard par groupe d'âge [10]. Timaeus et Jasseh (2004) utilisaient une version adaptée du modèle Sud de Coale et al. (1983), tandis que nous optons pour le modèle général développé par les Nations Unies pour les pays en développement (UN 1982). Les tests que nous avons menés (non présentés ici) indiquent que le choix du standard n'a qu'un impact très léger sur les estimations lorsque le niveau de mortalité retenu est intermédiaire (autour de 50 ans d'espérance de vie). Dans ce cas en effet, le logarithme des taux par âge est approximativement linéaire, comme on l'observe sur le graphe 5.2 (p.231). Les modèles Est et Sud des tables de Princeton, ainsi que le modèle général des Nations Unies (1982) sont tous les trois approximativement linéaires pour des valeurs d'espérance de vie avoisinant 50 ans. C'est nettement moins le cas pour des mortalités plus faibles, comme autour de 75 ans d'espérance de vie, par exemple.

3. Le schéma par âge de la mortalité est donc spécifique à chaque pays mais fixe dans le temps. Une exception est faite par Timaeus et Jasseh (2004) pour les pays *sévèrement affectés par le VIH-sida*, et pour les individus de plus de 20 ans. En effet, après un intervalle de quatre ans suivant l'année au cours de laquelle la prévalence du VIH a atteint le seuil de 1% (parmi les 15-49 ans), on peut s'attendre à voir se multiplier les décès imputables au VIH-sida. Le schéma par groupe d'âge de la mortalité est donc autorisé à changer selon la durée écoulée depuis cette date. Ceci renvoie au coefficient $\beta_5(\bar{x},g)(t-T_i)$, où T_i correspond à quatre ans après l'année où la prévalence a atteint le seuil de 1%.

   Timaeus et Jasseh (2004) autorisent ce schéma par âge de surmortalité à évoluer avec l'épidémie de façon *identique à tous les pays*, en

10. Dans l'équation 5.4, $\bar{x}$ renvoie au groupe d'âge, tandis que x renvoie à l'âge par années.

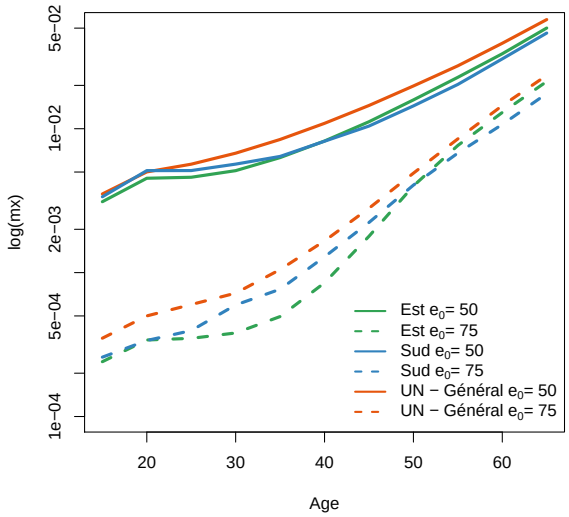

FIGURE 5.2 – *Logarithme des taux de mortalité masculins - Modèles Sud et Est de Coale et Demeny (1983), modèle général des Nations Unies (1982), pour des valeurs de e_0 de 50 et 75 ans.*

postulant donc une similarité des hausses de mortalité par âge associées au VIH-sida. Dans la mesure où davantage de données sont disponibles en 2010 et pour respecter d'éventuelles spécificités régionales, nous supposons plutôt que cette hausse est spécifique à chaque sous-région (Afrique australe, Afrique de l'Ouest, Afrique de l'Est et Afrique Centrale). Dans la pratique, nous ajustons donc le modèle sur chacune de ces quatre sous-régions prises séparément.

4. Enfin, de façon spécifique aux pays, un *terme quadratique* est ajouté en prenant le carré de "l'âge de l'épidémie" ($\beta_6(i)(t - T_i)^2$). Ceci ne se fait que dans un sous-groupe de pays (S) présentant une prévalence qui se stabilise ou qui décline sur la période d'observation. Cette variable vise à tenir compte du fait que la mortalité adulte va baisser dans les pays où les pics de prévalence ont déjà été atteints.

En plus de ces différentes variables, nous ajoutons un paramètre destiné à corriger la tendance à sous-déclarer davantage les décès qui surviennent dans le passé lointain. Cet ajustement est discuté dans le chapitre 6, consacré à la qualité des données (p.304).

Nous avons appliqué ce modèle à toutes les enquêtes EDS (standard) disposant d'un module sur la mortalité maternelle (*cf.* annexe A.1.3), à l'exception de celle menée au Nigeria en 1999. Cette enquête est exclue, car la qualité de ses données est nettement moins bonne que dans les autres enquêtes [11]. Au total, 60 enquêtes EDS, pour 31 pays sub-sahariens, ont été intégrées.

11. Ceci apparaîtra clairement lorsque nous examinerons la qualité des données. Cette enquête n'a pas été supervisée par Macro International. Selon Pullum (2008) : "It happens that the Nigeria 1999 survey went into the field without DHS approval and without the normal level of training and supervision. The survey provides a textbook example of what can go wrong with a demographic survey."

Dans neuf enquêtes, les hommes ont également été interrogés sur la survie de leurs frères et sœurs, et ces informations sont prises en compte. A ce stade, nous ne faisons pas de différence entre les enquêtes "hommes" et "femmes" ; les données sont simplement combinées.

Pour chacun des pays, la configuration en personnes-périodes est faite de façon à ce que les intervalles de temps qui définissent les personnes-périodes soient identiques d'une enquête à l'autre. La date (moyenne) de la dernière enquête sert de référence. Ce mode de calcul permet de fusionner les enquêtes pour un même pays.

Les frères et sœurs dont le sexe ou la survie sont inconnus par l'enquêté sont éliminés (ils sont rares - *cf.* chap.6). Seuls les décès et les durées d'exposition qui se réfèrent aux 9 années ayant précédé les différentes enquêtes sont retenus pour les analyses. Par exemple, si une seule enquête est disponible pour juin 2000, nous remontons jusqu'en juin 1991. Si une seconde enquête est disponible en février 1995, nous couvrons la période de juin 1986 à juin 2000. Le choix de s'en tenir à 9 ans se justifie pour deux raisons :

1. La période de 9 ans fournit déjà une certaine idée de la tendance lorsqu'une seule enquête est disponible, sans devoir remonter trop dans le passé. Comme cela apparaîtra dans le chapitre 6, la qualité des données diminue rapidement à mesure que l'on remonte dans le temps : les décès sont davantage sous-déclarés.
2. Plus on cherche à remonter dans le temps, plus le nombre de frères et sœurs de 15 à 60 ans diminue, en raison du fait que seuls les enquêtés de 15 à 49 fournissent des informations sur leurs frères et sœurs. Par exemple, plus de 10 ans avant l'enquête, il y aura peu de frères et sœurs de 15-49 déclarés par des enquêtés de 15-19 ans. Ceci se traduit par des estimations qui sont de plus en plus erratiques (*cf.* infra).

Il y a toutefois une exception à cette période de 9 ans, et nous suivons en cela Timaeus et Jasseh (2004). C'est le cas du Rwanda, où la période de référence est limitée à 6 ans dans le cas de l'enquête de 2000 (06/2000 - 08/2000), afin d'éviter les distorsions importantes induites par le génocide (printemps 1994).

Notons également que les données sont pondérées, et que les poids qui sont alloués aux enquêtés sont ceux que nous appliquons aux frères et sœurs. Nous discuterons plus loin des biais de sélection - précisons simplement ici que les informations sur les enquêtés eux-mêmes (qui font par définition partie de leur fratrie) ne sont pas comptabilisés dans les périodes d'exposition.

Au total, les 60 enquêtes fournissent plus de 17.6 millions de personnes-années (après pondération) et plus de 106 000 décès. L'ensemble des données couvre une période allant de 1983 à 2009. Pour ajuster le modèle, les estimations de prévalence que nous utilisons sont

celles de l'UNAIDS (2008) [12].

Les résultats de l'ajustement de ce modèle sont détaillés dans le chapitre 7 pour la plupart des pays considérés. A titre illustratif, nous présentons ici les taux de mortalité de 15 à 60 ans calculés pour le Zimbabwe et la Zambie (graphe 5.3). Les taux par groupe d'âge sont transformés en taux de 15 à 60 ans par le biais de la structure par âge estimée par les Nations Unies. Les taux bruts sont présentés par années sous forme de carrés, tandis que les courbes correspondent au résultat de l'ajustement. Elles esquissent une tendance qui cadre bien avec les estimations brutes. Les niveaux "ajustés" apparaissent légèrement supérieurs aux taux "bruts", parce qu'ils sont corrigés pour tenir compte de la sous-déclaration des décès dans les périodes éloignées de l'enquête. Ceci est discuté plus loin (p. 304).

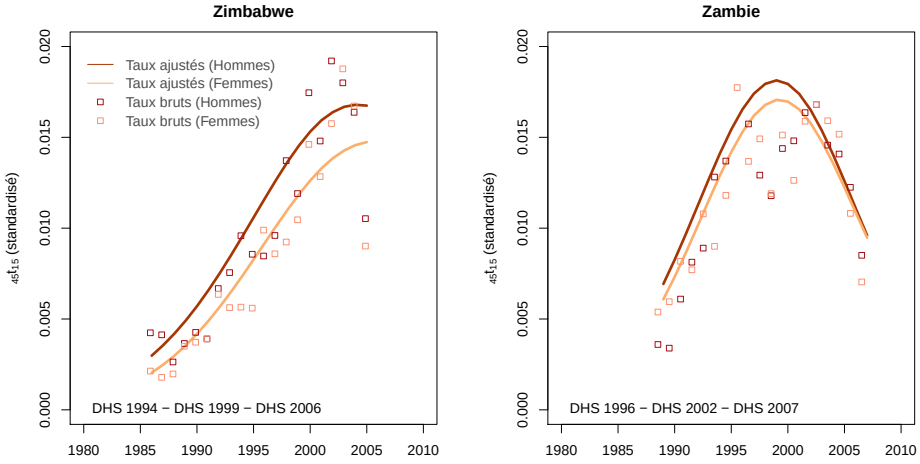

FIGURE 5.3 – *Taux de mortalité de 15 à 60 ans calculés à partir de la survie des frères et sœurs et du modèle de Timaeus et Jasseh (2004)- Zimbabwe et Zambie*

Modèle de Obermeyer et al. (2010)

Obermeyer et al. (2010) ont récemment proposé un modèle de régression alternatif pour exploiter les données EDS. Ce modèle est basé sur la régression logistique. Les tendances sont modélisées de façon non paramétriques, par intervalles de 5 ans et par groupes d'âges [13]. Leur modèle complet regroupe la plupart des enquêtes EDS disponibles, en ce compris celles menées dans d'autres pays que les pays sub-sahariens. Le schéma par âge de la mortalité est modélisé en croisant les groupes d'âge avec différentes variables dichotomiques, qui rassemblent les personnes-années en quatre groupes :

1. le premier regroupe les personnes-années où les pays ont enregistré de nombreux décès causés par *la guerre* [14].

12. http ://data.unaids.org/pub/GlobalReport/2008/.

13. L'ensemble des do.file STATA nécessaires pour ajuster leur modèle aux données EDS peut être téléchargé sur le site http ://www.healthmetricsandevaluation.org/

14. Les auteurs utilisent la base de données PRIO de l'université d'Uppsala. Le seuil retenu est de 0.75 décès liés à la guerre (survenus lors de com-

2. le deuxième regroupe les personnes-années où la prévalence du *VIH-sida* est au moins égale à 7% [15],
3. les deux groupes suivants rassemblent les autres observations, classées en deux catégories en fonction du niveau de *mortalité des enfants* estimée par Murray et al. (2007) [16].

Leur modèle incorpore également un facteur d'ajustement pour tenir compte d'une sous-déclaration des décès plus importante dans le passé lointain, ainsi qu'un système de pondération qui vise à corriger les biais de sélection. Ces différents aspects sont discutés dans les sections suivantes.

A notre connaissance, ces deux modèles sont à ce jour les seuls à avoir été utilisés en combinant plusieurs enquêtes EDS comprenant des informations sur les frères et sœurs. Nous retenons le modèle de Timaeus et Jasseh (2004) car il nous semble mieux modéliser les tendances et les schémas par âge de la mortalité. D'autres options seraient bien entendu possible, par exemple en étendant la régression dans un cadre multi-niveaux (avec des effets contextuels au niveau des régions ou de certains groupes de pays).

5.2 Biais de sélection

Dans la mesure où les données sur les frères et sœurs sont collectées de façon rétrospective, diverses formes de biais de sélection sont introduites.

- D'une part, il n'y a aucune information sur les familles pour lesquels il n'y a *plus aucun survivant âgé de 15 à 49 ans* [17].
- D'autre part, les fratries à mortalité élevée sont *sous-représentées*, puisque cette forte mortalité élimine des répondants potentiels.

Les diverses formes de dépendance entre (1) la survie d'un individu, (2) la survie de ses frères et sœurs et (3) la taille de sa fratrie vont moduler l'ampleur de ces deux biais.

5.2.1 Démonstration mathématique de Trussell et Rodriguez (1990)

Trussell et Rodriguez (1990) ont démontré mathématiquement qu'à trois conditions, on pouvait obtenir un indicateur non biaisé de la mortalité à partir des déclarations faites sur les frères et sœurs :

1. tous les frères et sœurs présents dans le plan d'échantillonnage sont interrogés,
2. les enquêtés eux-mêmes sont exclus des calculs,
3. les risques de mortalité ne varient pas avec la taille des fratries.

bats ou non) pour 1000 personnes. La base de données est téléchargeable sur http ://www.prio.no/CSCW/Datasets/Armed-Conflict/.

15. Après avoir introduit un délai de 5 ans pour tenir compte du décalage entre les changements de la prévalence et la hausse des niveaux de mortalité

16. Les personnes-années sont codées selon qu'elles appartiennent aux deux premiers terciles de la distribution des niveaux de mortalité des enfants ou au troisième tercile.

17. Ou plus aucune sœur survivante de 15 à 49 ans si seules les femmes sont interrogées.

Pour faire le détail de leur démonstration, imaginons d'abord (1) une situation où tous les individus (hommes et femmes) sont interrogés sur leurs fratries, et (2) une probabilité de décès pour les sexes réunis notée p, indépendante du nombre de frères et sœurs, noté n. Le nombre de décès survenus dans chaque fratrie suit donc une loi binomiale avec une moyenne $\mu = np$ et une variance $\sigma^2 = np(1-p)$.

Le calcul des taux, qu'il se fasse de façon directe ou indirecte, se fait généralement en déduisant l'information relative aux enquêtés eux-mêmes [18]. Dans ce cas, si tous les adultes survivants ont été interrogés, le nombre de répondants de chaque fratrie vaut $(n-x)$, le nombre de décédées vaut x, et la taille de la fratrie considérée pour le calcul de l'exposition vaut $(n-1)$. L'expérience d'un individu est donc répétée dans les données autant de fois qu'il y a d'autres enquêtés.

La proportion de décès observée *dans chaque fratrie* peut donc s'exprimer comme ceci, si $f(x)$ correspond à la fonction de probabilité de la loi binomiale [19] :

$$PD = \frac{\sum_{x=0}^{n-1} x(n-x)f(x)}{\sum_{x=0}^{n-1}(n-1)(n-x)f(x)} \tag{5.5}$$

En procédant à une simplification, on peut établir que cette équation équivaut à :

$$PD = \frac{npq(n-1)}{nq(n-1)} = p \tag{5.6}$$

Les étapes permettant de passer de la première à la seconde équation sont détaillées en l'annexe A.5.1 et dans l'article de Trussell et Rodriguez (1990).

Jusque là, nous n'avons qu'une information par groupes de fratries regroupées selon leur taille (n). Il faut donc simplement faire intervenir la distribution des différentes tailles de fratries pour compléter la démonstration. Notons le numérateur $N(n)$ et le dénominateur $D(n)$. Considérons ensuite une distribution des tailles des fratries $g(n)$. La proportion des décès observés sera alors obtenue en prenant la somme des produits entre les numérateurs $N(n)$ et les distributions $g(n)$, rapportée à la somme des produits entre les dénominateurs $D(n)$ et les distributions $g(n)$. En somme, nous prenons simplement la somme des décès observés sur la somme des frères et sœurs observés (répondants exclus), toutes tailles de fratrie confondues :

$$PD = \frac{\sum g(n)N(n)}{\sum g(n)D(n)} = \frac{\sum g(n)\sum_{x=0}^{n-1} x(n-x)f(x)}{\sum g(n)\sum_{x=0}^{n-1}(n-1)(n-x)f(x)} \tag{5.7}$$

Comme $N(n) = pD(n)$ et que p est identique pour toutes les fratries, il est possible de le factoriser. Cette équation se simplifie donc également en

18. Voir les rapports EDS, Timaeus et al. (2001), Timaeus et Jasseh (2004), Feeney (2001), etc.

19. Cette fonction de probabilité s'exprime comme ceci :

$$f(x) = \frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}$$

$PD = p$. Cela démontre que la proportion observée est bien un estimateur non biaisé de la probabilité de décès lorsque (1) tous les frères et sœurs sont interrogés, que (2) la même probabilité de survie est appliquée quelle que soit la taille des fratries, et que (3) la durée d'exposition du répondant est exclue des calculs.

On peut illustrer cette démonstration mathématique par le biais d'un exemple simple, toujours basé sur la distribution binomiale. Imaginons des fratries de tailles variables (entre 1 et 10), une probabilité de décès de 0.05, et tirons 100 fratries au hasard. Nous supposons que la distribution des tailles de fratries (c'est-à-dire des descendances finales des mères) suit une loi de Poisson avec un paramètre λ fixé à 5 (soit des fratries de 5 personnes en moyenne environ). On pourrait alors observer les valeurs reprises dans le tableau 5.1, où la proportion recalculée à partir des différentes déclarations est bien équivalente à la probabilité de décès sous-jacente. Nous parlerons de "calcul EDS" pour qualifier ce type d'exploitation des données, car c'est l'option retenue dans les rapports d'enquête.

n	1	2	3	4	5	6	7	8	9	10
p	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
N(n)	0.000	0.095	0.285	0.570	0.950	1.425	1.995	2.660	3.420	4.275
D(n)	0.0	1.9	5.7	11.4	19.0	28.5	39.9	53.2	68.4	85.5
nb frat.	3	8	16	25	13	13	12	4	3	3
PD(réelle)							24.15/483		=	**0.05**
PD(observée)							108.11/2162.2		=	**0.05**

TABLE 5.1 – *Exemple avec probabilité indépendante de la taille de la fratrie, calcul EDS*

Les taux calculés de façon classique à partir des enquêtes EDS sont donc issus d'une combinaison de trois biais de sélection, qui s'annulent lorsque la mortalité ne varie pas avec la taille des fratries, à savoir : l'absence d'information sur les fratries sans survivants, l'absence de prise en compte de la survie des enquêtées, et la sous-représentation des fratries à mortalité élevée. Notons par ailleurs que ce calcul ne fait pas intervenir d'information sur les enfants uniques ; car dans le cas d'une exploitation de type EDS, $D(1)$ est nul.

Si, dans la même configuration, nous ne retenons qu'une seule déclaration par fratrie, la probabilité observée s'exprime alors comme ceci :

$$PD = \frac{\sum g(n) \sum_{x=0}^{n-1} x f(x)}{\sum g(n) \sum_{x=0}^{n-1} (n-1) f(x)} \tag{5.8}$$

Cette situation, qui est celle visée lorsque l'on cherche à ne retenir qu'une sœur par ménage (c'est le cas de l'EDS menée en Bolivie en 1989), procure une probabilité observée trop élevée par rapport à la probabilité réelle, comme on le constate dans le tableau 5.2.

n	1	2	3	4	5	6	7	8	9	10
p	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
N(n)	0.000	0.095	0.150	0.200	0.250	0.300	0.350	0.400	0.450	0.500
D(n)	0.000	0.997	2.000	3.000	4.000	5.000	6.000	7.000	8.000	9.000
nb frat.	3	8	16	25	13	13	12	4	3	3
PD(réelle)							24.15/483		=	**0.05**
PD(observée)							23.953/382.976		=	**0.063**

TABLE 5.2 – *Exemple avec probabilité indépendante de la taille de la fratrie, un seul répondant*

Nous avons repris jusqu'ici les conclusions auxquelles parviennent Trussell et Rodriguez (1990). Leur message est donc qu'*il ne faut absolument pas se contenter d'une seule déclaration par ménage mais bien interroger tous les frères et sœurs éligibles*. Ce constat peut être étendu en faisant varier l'intensité de la mortalité p et la taille des fratries λ. Nous faisons varier les tailles moyennes de fratries entre 2 et 7, et les probabilités de décès entre 0.01 et 0.9 [20]. Pour chacune de ces combinaisons, une distribution des tailles de fratries est générée aléatoirement à partir de la loi de Poisson. Nous calculons ensuite l'erreur relative faite sur la proportion de décédés [21]. Comme on l'observe sur le graphe 5.4, cette erreur relative se chiffre à moins de 20% lorsque la taille moyenne des fratries est supérieure à 6 et que la probabilité de décès est supérieure à 0.65. Elle se situe entre 20 et 40% si les fratries sont d'une taille de 3 à 6 personnes et que la probabilité de décès varie entre 0.3 et 0.65.

Ce petit exemple illustre l'importance d'*interroger toutes les individus éligibles* présents dans les ménages sur la survie de leurs frères et sœurs : n'en interroger qu'un seul tout en exploitant les données de façon classique (en éliminant le répondant, etc.) amène à surestimer la mortalité, surtout dans les contextes où la mortalité est faible et où les fratries sont de petite taille.

Notons toutefois qu'aucune situation réelle ne correspond à ces simulations. Il faudrait en effet, pour ne retenir qu'une seule déclaration par fratrie, pouvoir *identifier complètement les fratries dupliquées dans le fichier*, ce qui est impossible. Même lorsqu'une seule enquêtée est retenue par *ménage*, une même fratrie peut être échantillonnée à partir de différents ménages. En effet, les individus ne cohabitent qu'avec une petite fraction de leur fratrie. A partir des enquêtes EDS africaines, nous avons aussi estimé qu'entre 2 et 21% seulement des enquêtés cohabitent avec au moins un frère ou une sœur éligible (*cf.* chapitre 1). Ce biais varie donc avec

20. Pour les tailles de fratries, 100 valeurs de λ sont prises entre 1.6 et 7. Comme le cas où $x = 0$ n'est pas observé, la moyenne des tailles observés vaut $\lambda/(1-e^{-\lambda})$, et la variance vaut $e^{\lambda}(e^{\lambda}-1-\lambda)\lambda/(1-e^{\lambda})^2$. 100 valeurs de la probabilité p sont également prises entre 0.01 et 0.9.

21. Calculée comme $\frac{valeur_{approchee}-valeur_{reelle}}{valeur_{reelle}}$

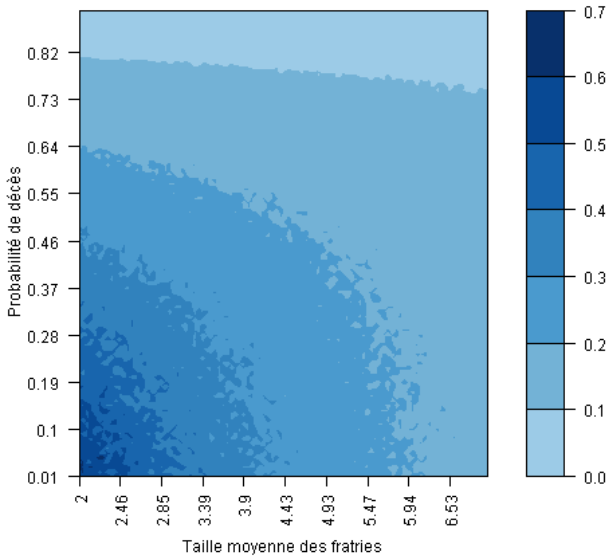

FIGURE 5.4 – *Erreur relative de la probabilité de décès observée lorsqu'une seule déclaration est retenue par fratrie.*
LECTURE : Pour une fratrie de 3 membres et une probabilité de décès de 0.5, l'erreur relative se situe entre 20 et 30%. Plus la taille des fratries diminue, plus l'erreur augmente. Plus la probabilité de décès diminue, plus l'erreur augmente.

la concentration géographique des frères et sœurs (Trussell et Rodriguez 1990), et il est impossible de le quantifier avec précision [22].

Notons également que certains auteurs continuent à ne retenir qu'une déclaration par fratrie. Par exemple, dans leur analyse de la mortalité adulte au cours du génocide rwandais, de Walque et Verwimp (2010) disent avoir éliminé les fratries dupliquées du fichier [23]. Leurs résultats doivent donc être interprétés avec prudence.

5.2.2 Applications aux micro-simulations

Nous avons jusqu'ici raisonné dans le cadre simple d'une mortalité constante dans le temps et identique quels que soient l'âge et le sexe. Nous avons également considéré que tous les individus fournissaient des informations sur leurs frères et sœurs, ce qui n'est pas le cas dans les enquêtes EDS. Dans la plupart de ces enquêtes, seules *les femmes de 15 à 49 ans* sont concernées par le module sur la mortalité maternelle. Dans une sur cinq environ, les *hommes* le sont également (sur un sous-échantillon). Nous pouvons nous rapprocher des situations réelles en recourant aux résultats de la microsimulation. Dans la série de simulations que nous avons construite (chapitre 2), la taille des fratries n'est pas associée à la

22. Les fratries sont également réparties sur différents clusters, et même si l'identification des fratries pouvait se faire en dehors des ménages, elle ne pourrait l'être qu'au sein des clusters échantillonnés.

23. Ils ne précisent pas comment s'est fait le repérage des fratries dupliquées. Selon nos estimations, environ 14% des enquêtées co-habitent avec une sœur éligible dans l'enquête rwandaise de 2000.

mortalité [24], ce qui nous place dans les conditions requises par la démonstration mathématique de Trussell et Rodriguez (1990).

A partir des populations simulées pour le Mozambique, nous comparons les valeurs "réelles" du ${}_{45}q_{15}$, c'est-à-dire celles qui renvoient à la mortalité réellement à l'oeuvre dans les microsimulations, avec celles qui sont recalculées à partir de la survie des frères et sœurs. Quatre cas de figure sont présentés dans le graphe 5.5, selon le *sexe* des enquêtés (femmes uniquement ou hommes et femmes), et selon les *plages d'âge retenues* (tous âges ou 15-49 ans uniquement) :

A - tout les individus survivants au terme des simulations fournissent des informations sur leur fratrie, quels que soit leur âge ou leur sexe,

B - seuls les survivants âgés de 15 à 49 ans sont "enquêtés" (les hommes et les femmes),

C - seuls les femmes survivantes sont "enquêtées" (quel que soit leur âge),

D - seules les femmes de 15 à 49 ans sont "enquêtées".

La dernière configuration est donc celle de la plupart des EDS, et la configuration B est celle de quelques enquêtes où les hommes sont interrogés également sur leur fratrie. Pour produire ces graphes, les probabilités de décès sont calculées de façon "brute", sans modéliser les tendances ni les schémas par âge. A partir des populations simulées, nous construisons un fichier "frères et sœurs", puis les décès sont simplement rapportés aux unités d'exposition.

Pour la première décennie qui précède l'enquête, ces différentes approches procurent toutes des niveaux de mortalité très proches des niveaux réellement à l'oeuvre dans les microsimulations (en traits pleins). Ceci indique que *même en tenant compte des âges et du sexe des répondants, les trois biais de sélection s'annulent toujours* si, comme nous l'avons postulé à ce stade, les risques de décès sont tous issus de la même distribution. Ce qui est intéressant de noter également à partir des schémas C et D, c'est qu'interroger les sœurs sur leurs frères procure un résultat sans biais, et ce, *même si la mortalité féminine diffère de la mortalité masculine*. Nous pouvons expliquer cela par trois raisons :

1. par définition, le fait d'intégrer ou non le répondant dans les calculs n'a pas d'impact, puisque ces calculs portent sur les individus du sexe opposé.
2. les données masculines relatives aux fratries à faible mortalité féminine sont sur-représentées (et *vice versa*), mais cela ne constitue pas un biais, compte tenu de l'indépendance des risques de décès au sein des fratries.
3. pour la même raison, le fait que certaines fratries n'ont plus aucune femme survivante pour déclarer des décès n'a pas d'impact sur l'estimation de la mortalité masculine (et *vice versa*).

24. Dans certains pays affectés par le VIH, la mortalité des femmes séropositives est réduite et la mortalité de leurs enfants est plus élevée, mais peu d'enfants infectés intra- ou post- partum survivent jusqu'aux âges adultes (les ART n'étant pas étendus aux enfants dans nos simulations).

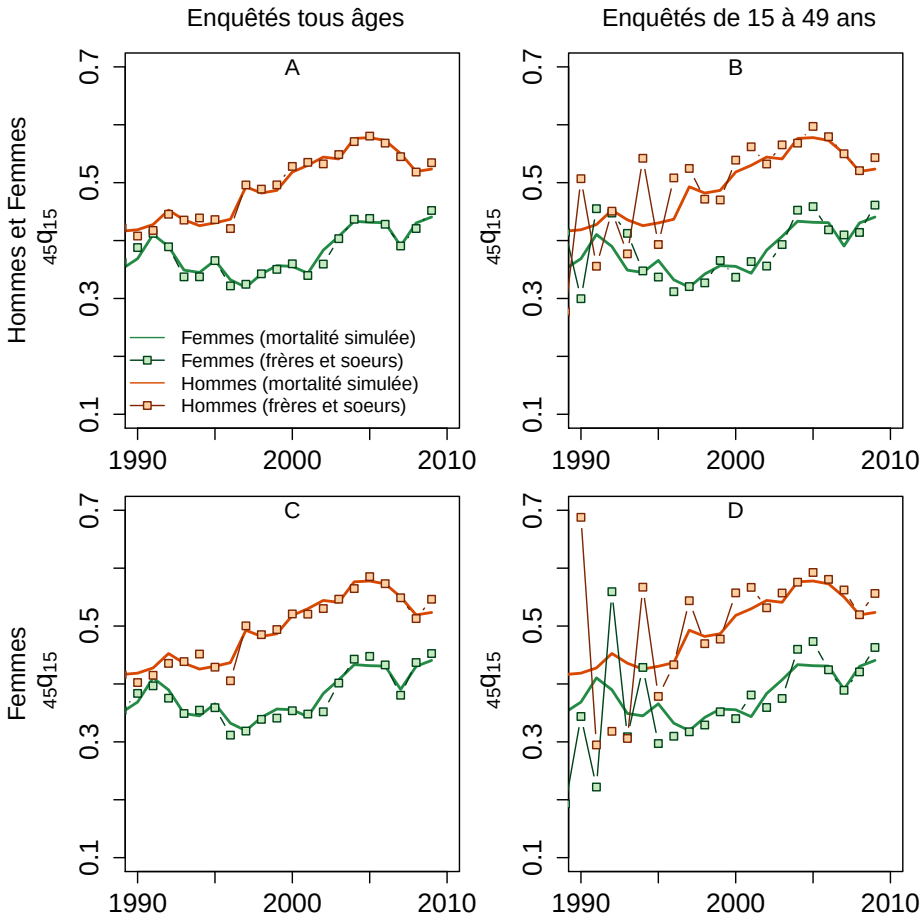

FIGURE 5.5 – *Probabilités de décès entre 15 et 60 ans simulées (les traits) et recalculées (les points) à partir des frères et sœurs - microsimulations - scénario Mozambique*

Remarquons toutefois que dans la configuration où seuls les adultes âgés de 15 à 49 ans répondent (graphes B et D), les valeurs ré-estimées de ${}_{45}q_{15}$ s'écartent davantage des valeurs sous-jacentes. La variation aléatoire va en grandissant à mesure que l'on remonte dans le temps. Cela s'explique par le fait que de moins en moins d'enquêtés ont des frères et sœurs qui étaient âgés de 15 à 49 ans sur les périodes considérées.

Il faut relever à cet égard le fait que les valeurs présentées dans le graphe 5.5 sont établies à partir de *tous* les répondants potentiels observés dans les simulations. Typiquement, cela représente plus de 300 000 individus lorsque tous les survivants en 2009 sont interrogés (ce qui correspond au graphe A, en haut à gauche). Dans une configuration EDS où les questions sont uniquement posées aux femmes en âge de reproduction (graphe D, en bas à droite), entre 20 et 30% seulement du nombre total de survivants sont interrogés. C'est toutefois encore beaucoup plus d'enquêtés que ne le prévoient les échantillons EDS, qui varient entre 4 800 et 33 400 femmes adultes (avec une médiane qui s'établit à 8 000 personnes).

Afin de nous rapprocher de cette configuration, nous pouvons tirer au hasard dans les populations simulées un certain nombre d'enquêté(e)s, et mesurer ensuite la variation autour des probabilités de décès recalculées à partir des frères et sœurs. Prenons par exemple le cas de l'Afrique du Sud. A partir de la population obtenue au terme des microsimulations, nous tirons 100 échantillons différents de 8000 femmes âgées de 15 à 49 ans. A partir de ces femmes, nous reconstruisons des fichiers "frères et sœurs" comparables à ceux que l'on obtient au départ des enquêtes EDS. Les valeurs de ${}_{45}q_{15}$ sont ensuite recalculées pour chacun de ces échantillons, à partir des informations fournies sur les fratries. Les graphes 5.6 confrontent par sexe les valeurs "réelles" des ${}_{45}q_{15}$ calculées sur toute la population avec les valeurs réestimées à partir des frères et sœurs sur base de ces différents échantillons. Ces valeurs réestimées sont présentées sous forme de boxplots. Les deux graphes de droite présentent ensuite le *coefficient de variation*, calculé comme la racine carrée de l'erreur quadratique moyenne divisée par la valeur "réelle" du quotient ${}_{45}q_{15}$. Trois cas de figure sont présentés, pour des échantillons de 6 000, 8 000 et 10 000 enquêtées.

On constate que les tailles d'échantillon influencent relativement peu le coefficient de variation (ou l'erreur moyenne) faite sur les valeurs de ${}_{45}q_{15}$. Celle-ci augmente rapidement à mesure que l'on s'éloigne de la date de l'enquête. De 5 à 7% environ immédiatement avant l'enquête, l'erreur moyenne faite sur les valeurs de ${}_{45}q_{15}$ grimpe à plus de 25% lorsque l'on remonte plus de dix ans avant l'enquête.

De tout cela, on retiendra plusieurs choses :

- La démonstration mathématique de Trussell et Rodriguez (1990) est corroborée par les microsimulations. Elle peut en outre être étendue aux situations des enquêtes, où seules les femmes adultes de 15 à 49 ans sont interrogées.
- Il est important d'interroger tous les individus éligibles et de ne pas

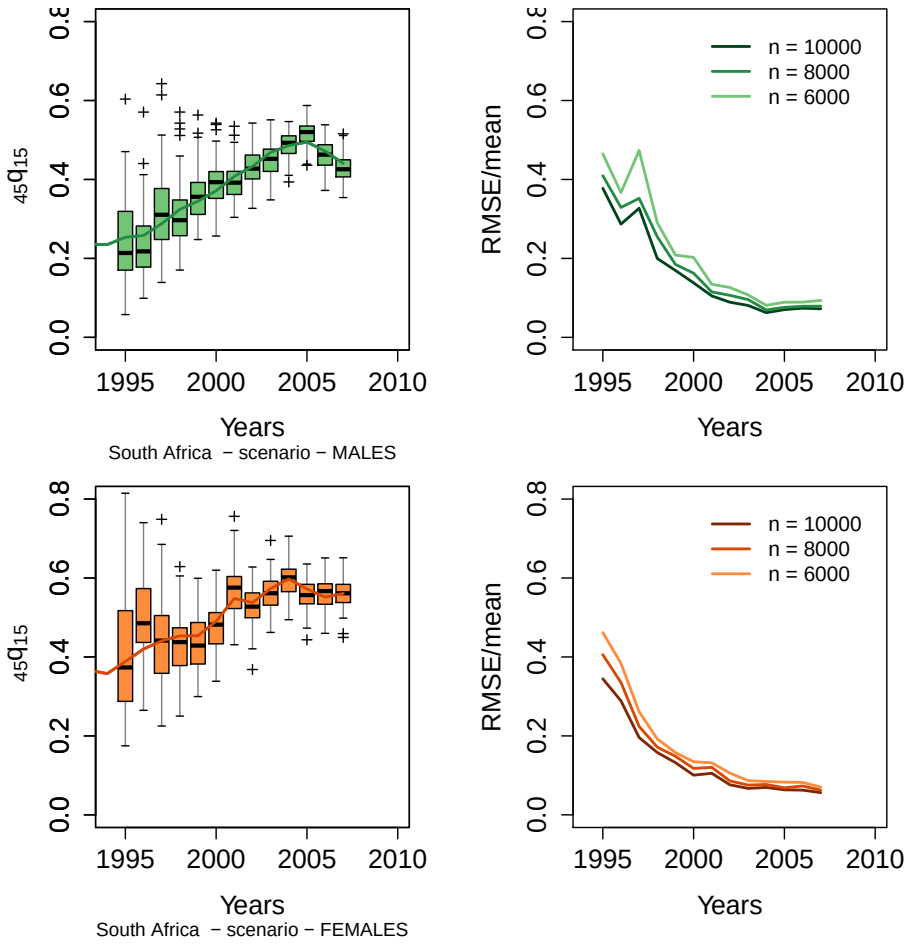

FIGURE 5.6 –
Panel de gauche *: Probabilites de décès entre 15 et 60 ans simulées et recalculées à partir des frères et sœurs sur 8000 femmes survivantes de 15 à 49 ans (Exemple de l'Afrique du Sud)*
Panel de droite *: RMSE/moyenne des* ${}_{45}q_{15}$

s'en tenir à une déclaration par fratrie ou par ménage, sans quoi la mortalité risque d'être surestimée, surtout là où la mortalité reste faible et où les fratries sont de petite taille.
– Il est préférable de ne pas remonter trop loin dans le passé, car les erreurs aléatoires augmentent rapidement (sans parler encore de la sous-déclaration des décès).

5.2.3 Association entre mortalité et taille des fratries

Dans le cas où la mortalité varie avec la taille des fratries, les différents biais de sélection ne s'annulent plus. En effet, les probabilités $p(n)$ ne peuvent plus être factorisées (*cf.* équation 5.9). Comme $N(n) = D(n)p(n)$, le calcul classique procure une moyenne des probabilités $p(n)$ pondérée par le produit entre *la distribution des tailles de fratries et le nombre de frères et sœurs observés* ($g(n)D(n)$). Or la mortalité qui est réellement à l'oeuvre correspond à la moyenne pondérée des probabilités $p(n)$ par le produit entre *la distribution des tailles de fratries et leur taille* ($g(n)\ n$) :

$$\begin{aligned} PD &= \frac{\sum g(n)N(n)}{\sum g(n)D(n)} = \frac{\sum g(n)\left[D(n)p(n)\right]}{\sum g(n)D(n)} \\ &= \frac{\sum g(n)\left[n(1-p(n))(n-1)\right]p(n)}{\sum g(n)n(1-p(n))(n-1)} \\ &\neq \frac{\sum g(n)\left[p(n)n\right]}{\sum g(n)n} \end{aligned} \tag{5.9}$$

Par rapport à la pondération exacte ($g(n)n$), la pondération observée donne plus d'importance aux fratries de grande taille (via le facteur $n-1$) ainsi qu'aux fratries où la mortalité est faible (via le facteur $1-p$).

En fonction des différentes combinaisons entre la taille des fratries et la mortalité, le calcul classique "de type EDS" va donc induire *soit* une surestimation, *soit* une sous-estimation. Une forte corrélation positive entre la taille des fratries et la mortalité induit une *surestimation*, puisque les fratries de grandes tailles à forte mortalité sont alors nettement surreprésentées. Ceci peut s'illustrer en reprenant un exemple simple, basé sur des probabilités de décès fixées arbitrairement de 0.03 à 0.075 selon la taille des fratries. On observe dans le tableau 5.3, que l'exploitation de type EDS induit dans ce cas une surestimation. La proportion de décédés observée ($PD(o)$) se chiffre à 0.058, tandis que la proportion réelle ($PD(r)$) ne s'élève qu'à 0.054. La mortalité est donc surestimée de plus de 7%.

C'est l'inverse si la mortalité *diminue* fortement avec la taille des fratries ; dans ce cas, les petites fratries à forte mortalité sont alors doublement pénalisées, et sont nettement sous-représentées. Dans le tableau 5.4, la proportion observée ($PD(o)$) se chiffre à 0.047, tandis que la proportion réelle ($PD(r)$) s'élève à 0.051.

Notons que les biais diminuent quand diminue la corrélation entre la taille des fratries et la mortalité, mais *l'absence de corrélation n'est pas une condition suffisante pour obtenir un estimateur véritablement sans biais*. On

n	1	2	3	4	5	6	7	8	9	10
p	0.030	0.035	0.040	0.045	0.050	0.055	0.060	0.065	0.070	0.075
N(n)	0.000	0.068	0.230	0.516	0.950	1.559	2.369	3.403	4.687	6.244
D(n)	0.00	1.93	5.76	11.46	19.00	28.35	39.48	52.36	66.96	83.25
nb frat.	3	8	16	25	13	13	12	4	3	3
W(r)	0.006	0.033	0.099	0.207	0.135	0.161	0.174	0.066	0.056	0.062
W(o)	0.000	0.007	0.043	0.134	0.115	0.172	0.221	0.098	0.094	0.117
PD(r)							25.87/483		=	**0.054**
PD(o)							124.574/2143.48		=	**0.058**

TABLE 5.3 – *Exemple du calcul de la probabilité de décès lorsque celle-ci **augmente** avec la taille de la fratrie, calcul EDS*

n	1	2	3	4	5	6	7	8	9	10
p	0.075	0.070	0.065	0.060	0.055	0.050	0.045	0.040	0.035	0.030
N(n)	0.000	0.130	0.365	0.677	1.039	1.425	1.805	2.150	2.432	2.619
D(n)	0.00	1.86	5.61	11.28	18.90	28.50	40.11	53.76	69.48	87.30
nb	3	8	16	25	13	13	12	4	3	3
W(r)	0.006	0.033	0.099	0.207	0.135	0.161	0.174	0.066	0.056	0.062
W(o)	0.000	0.007	0.041	0.130	0.113	0.171	0.222	0.099	0.096	0.121
PD(r)							24.845/483		=	**0.051**
PD(o)							101.25/2169.54		=	**0.047**

TABLE 5.4 – *Exemple du calcul de la probabilité de décès lorsque celle-ci **diminue** avec la taille de la fratrie, calcul EDS*

l'observe dans le tableau 5.5, que nous avons conçu de façon à ce que la corrélation entre n et p soit nulle, avec des probabilités p variables. L'erreur faite sur la proportion de décédés est faible, mais non nulle. Le problème se situe bien au niveau de la pondération des différentes valeurs de p.

n	1	2	3	4	5	6	7	8	9	10
p	0.055	0.060	0.030	0.050	0.065	0.040	0.070	0.075	0.035	0.045
N(n)	0.000	0.113	0.175	0.570	1.216	1.152	2.734	3.885	2.432	3.868
D(n)	0.00	1.88	5.82	11.40	18.70	28.80	39.06	51.80	69.48	85.95
nb frat.	3	8	16	25	13	13	12	4	3	3
W(r)	0.006	0.033	0.099	0.207	0.135	0.161	0.174	0.066	0.056	0.062
W(o)	0.000	0.007	0.043	0.132	0.113	0.174	0.218	0.096	0.097	0.120
PD(r)							25.485/483		=	**0.053**
PD(o)							115.986/2152.87		=	**0.054**

TABLE 5.5 – *Exemple du calcul de la probabilité de décès lorsque celle-ci **varie** avec la taille de la fratrie, calcul EDS*

5.2.4 Correction de Gakidou et King (2006)

Une méthode a récemment été proposée par Gakidou et King (2006), afin de tenir compte d'une éventuelle association entre la taille des fratries et l'intensité de la mortalité. Cette méthode a deux faces : la première consiste en un mécanisme de pondération, et la seconde repose sur une extrapolation. La pondération procure les taux pour les fratries où il subsiste *au moins un répondant*, tandis que l'extrapolation permet d'estimer le nombre de décès survenus dans les fratries *sans aucun survivant* [25].

Cette nouvelle approche a reçu une attention considérable dans la littérature portant sur la mortalité adulte. Elle été appliquée, sous diverses formes, sur les données de l'*Iraq Family Health Survey* (Iraq Family Health Survey Study Group 2008), sur les enquêtes WHS (Obermeyer et al. 2008), ainsi que sur les données EDS (Obermeyer et al. 2010, de Walque et Verwimp 2010). Obermeyer et al. (2010) disent même avoir développé une nouvelle méthode d'estimation, appelée 'Corrected Sibling Survival', à partir des travaux de Gakidou et King (2006). Nous avons déjà présenté en quelques mots leur modèle de régression (p.233). La différence avec les approches antérieures tient surtout dans cette pondération, qui procure des niveaux de mortalité bien supérieurs à ceux obtenus classiquement :

> "Our findings suggest that levels of adult mortality prevailing in many developing countries are substantially higher than previously suggested by analyses of sibling history data. In sub-Saharan African populations largely unaffected by HIV, we estimate the risk of death between ages 15 and 60 years to be 20%-35% for females and 25%-45% for males, though considerable heterogeneity exists among countries. In Southern African countries where the HIV epidemic has been most pronounced, rates are uniformly and strikingly high : at current rates, as many as eight out of ten men alive at age 15 will be dead by age 60, as will as many as six out of ten women". (Obermeyer et al. (2010), p.7)

Gakidou et King (2006), comme l'équipe qui est derrière les travaux de Obermeyer et al. (2010), ont rendu publiques leurs syntaxes, ce qui nous permet de discuter de leurs travaux de façon précise [26]. Dans les sections suivantes, nous les présentons dans le détail. Si les propositions de Gakidou et King (2006) nous paraissent intéressantes, nous avons par contre de nombreuses réserves à formuler sur l'application de cette approche aux données EDS. Nous pensons notamment que l'application qui en est faite par Obermeyer et al. (2010) est erronée.

25. Les résultats publiés dans Gakidou et King (2006) peuvent être répliqués au moyen des syntaxes STATA rendues publiques sur http ://dvn.iq.harvard.edu/dvn/dv/king/

26. Les syntaxes utilisées par Gakidou et King (2006) sont disponibles sur le site http ://dvn.iq.harvard.edu/dvn/dv/king/. Celles de Obermeyer et al. (2010) sont disponibles sur le site de l'Institute for Health Metrics and Evaluation (IHME), de l'Université de Washington : www.healthmetricsandevaluation.org

Pondération des données

L'idée à la base de la pondération suggérée par Gakidou et King (2006) est la suivante : n'interroger que les survivants, cela revient à échantillonner dans cette population spécifique de survivants (notés $i = 1, \ldots, n$) alors qu'il aurait fallu échantillonner dans la population d'ensemble, c'est-à-dire en intégrant tous les individus présents au début de la période d'observation (notés $j = 1, \ldots, N$), y compris les individus décédés entretemps. Lorsque tous les individus sont intégrés, chaque tirage au hasard d'un individu revient à tirer une fratrie avec une proportion équivalente à la taille de la fratrie. Mais quand seuls les survivants sont interrogés, cette proportion est établie à partir du nombre de survivants. Dès lors, la probabilité d'observer des individus issus de fratries où il y a eu moins de survivants est réduite, et celle d'observer des individus issus de fratries où il reste beaucoup de survivants est augmentée, par rapport à la situation en début de période.

Gakidou et King (2006) proposent de corriger cela, pour les fratries où il reste au moins un survivant. En effet, dans ce cas, le nombre de frères et sœurs au départ de la période d'observation est connu. Notons le B_i. Le nombre de frères et sœurs survivants au moment de l'enquête l'est également : il est noté S_i. Il est donc possible de rééquilibrer les choses par le biais de facteurs de correction qui correspondent au rapport entre le *nombre de frères et sœurs présents en début de période* et le *nombre de frères et sœurs survivants* :

$$W_i = B_i / S_i \tag{5.10}$$

Mais il faut alors que toutes les informations soient retenues, c'est-à-dire à la fois celles qui portent sur les individus eux-mêmes (parfois appelées *own-report* en anglais) et celles qui concernent leurs frères et sœurs (*proxy-report*). A la différence de l'approche classique, il s'agit de ne pas éliminer les enquêtés du calcul des durées d'exposition. Cette précision est essentielle.

Dans les exemples précédents, nous avons chaque fois rapporté le nombre de décès déclarés par les survivants (D_i) au nombre de frères/sœurs qu'ils avaient au départ (B_i), en excluant les enquêtés et en autorisant les répétitions. Il s'agit du calcul que nous avons appelé de "type EDS" [27]. Gakidou et King (2006) expriment cet estimateur comme :

$$\hat{q} = \frac{\sum_{i=1}^{n} D_i}{\sum_{i=1}^{n} B_i - 1} \tag{5.11}$$

Pour faire le lien entre la terminologie de Gakidou et King (2006) et celle utilisée précédemment, précisons que cette équation est identique à l'équation 5.7, avec une formulation différente :

$$\hat{q} = \frac{\sum g(n) \sum_{x=0}^{n-1} x(n-x) f(x)}{\sum g(n) \sum_{x=0}^{n-1} (n-1)(n-x) f(x)}$$

27. Ce qu'ils appellent l'estimateur *standard*, noté $\breve{q}$.

Nous avons vu que cet estimateur est sans biais à trois conditions [28].

En ajoutant le répondant dans le calcul, ces deux équations sont transformées de cette façon [29] :

$$\tilde{q} = \frac{\sum g(n) \sum_{x=0}^{n-1} x(n-x) f(x)}{\sum g(n) \sum_{x=0}^{n-1} n(n-x) f(x)} = \frac{\sum_{i=1}^{n} D_i}{\sum_{i=1}^{n} B_i} \tag{5.12}$$

Il est évident que dans les mêmes conditions, cet estimateur va sous-estimer la mortalité.

Nous nous situons jusque là au niveau individuel. En d'autres termes, on fait simplement la somme des décès observés divisée par la somme des individus présents au départ. Alternativement, il est possible de calculer des *taux par fratrie*, c'est-à-dire, plus précisément, de prendre la moyenne, pondérée par le nombre de répondants, des proportions de décès déclarées par chacun des répondants (M_j). C'est à partir de ce niveau, *celui des fratries*, que Gakidou et King (2006) présentent leur pondération. Ceci induit une certaine confusion, car les données sont généralement exploitées au niveau individuel (et pas au niveau des fratries). Le calcul par fratrie prend la forme suivante :

$$\hat{q} = \frac{\sum g(n) \sum_{x=0}^{n} \frac{x}{n}(n) f(x)}{\sum g(n) \sum_{x=0}^{n} (n) f(x)} = \frac{\sum_{j=1}^{N} D_j / B_j}{N} = \frac{\sum_{j=1}^{N} M_j}{N} \tag{5.13}$$

Si tout le monde pouvait répondre (décédés y compris), et que les répondants étaient intégrés dans les calculs, cet estimateur serait sans biais. On l'observe dans le tableau 5.6 où la probabilité de décès augmente avec la taille des fratries.

n	1	2	3	4	5	6	7	8	9	10
p	0.030	0.035	0.040	0.045	0.050	0.055	0.060	0.065	0.070	0.075
N(n)	0.03	0.07	0.12	0.18	0.25	0.33	0.42	0.52	0.63	0.75
D(n)	1	2	3	4	5	6	7	8	9	10
nb frat.	3	8	16	25	13	13	12	4	3	3
PD(réelle)							25.87/483		=	**0.054**
PD(observée)							25.87/483		=	**0.054**

TABLE 5.6 – *Exemple de calcul par fratrie - tous répondent*

Mais lorsque seuls les survivants sont interrogés, cela conduit plutôt à ceci :

$$\hat{q} = \frac{\sum g(n) \sum_{x=0}^{n-1} \frac{x}{n}(n-x) f(x)}{\sum g(n) \sum_{x=0}^{n-1} (n-x) f(x)} = \frac{\sum_{i=1}^{n} D_i / B_i}{n} \tag{5.14}$$

A nouveau, nous présentons deux formulations, mais elles sont équivalentes.

28. (1) Tous les frères et sœurs sont interrogés, (2) la même probabilité de survie s'applique quelle que soit la taille des fratries, et (3) la durée d'exposition du répondant n'est pas comptée.

29. Ce que Gakidou et King (2006) appellent l'estimateur *naïf*, noté $\tilde{q}$.

n	1	2	3	4	5	6	7	8	9	10
p	0.030	0.035	0.040	0.045	0.050	0.055	0.060	0.065	0.070	0.075
N(n)	0.000	0.034	0.077	0.129	0.190	0.260	0.338	0.425	0.521	0.624
D(n)	0.97	1.93	2.88	3.82	4.75	5.67	6.58	7.48	8.37	9.25
nb	3	8	16	25	13	13	12	4	3	3
PD(réelle)							25.87/483		=	**0.054**
PD(observée)							19.77/457.13		=	**0.043**

TABLE 5.7 – *Exemple de calcul par fratrie - seuls les survivants répondent*

On remarquera dans le tableau 5.7 que dans ce cas, l'association positive entre la mortalité et la taille de la fratrie ne se traduit pas par une surestimation de la mortalité comme précédemment (tableau 5.3, p. 244), car le mode de calcul est ici différent. Les taux observés pour les différentes tailles de fratries sont tous sous-estimés, et ce, d'autant plus que les fratries sont petites [30].

C'est ici qu'intervient la pondération proposée par Gakidou et King (2006). Pour obtenir un taux de mortalité non biaisé relatif aux fratries pour lesquelles il subsiste au moins un survivant, ils proposent simplement de pondérer les taux observés par fratrie par $W_i = B_i/S_i$ [31] :

$$\hat{q} = \frac{\sum_{i=1}^{n} D_i/B_i \times B_i/S_i}{\sum_{i=1}^{n} B_i/S_i} = \frac{\sum_{i=1}^{n} D_i/S_i}{\sum_{i=1}^{n} B_i/S_i} \tag{5.15}$$

La même pondération s'exprime comme $n/(n-x)$ si nous traduisons cette équation dans la formulation que nous avons retenue - qui est celle utilisée par Trussell et Rodriguez (1990) :

$$\hat{q} = \frac{\sum g(n) \sum_{x=0}^{n-1} x\, f(x)}{\sum g(n) \sum_{x=0}^{n-1} n\, f(x)} \tag{5.16}$$

Les résultats de cette pondération sont présentés dans le tableau 5.8. La probabilité observée est très proche de la probabilité réelle... La différence tient au fait que les fratries sans aucun survivant sont omises du calcul.

n	1	2	3	4	5	6	7	8	9	10
p	0.030	0.035	0.040	0.045	0.050	0.055	0.060	0.065	0.070	0.075
N(n)	0.000	0.068	0.120	0.180	0.250	0.330	0.420	0.520	0.630	0.750
D(n)	0.970	1.998	3.000	4.000	5.000	6.000	7.000	8.000	9.000	10.000
nb	3	8	16	25	13	13	12	4	3	3
PD(réelle)							25.87/483		=	**0.054**
PD(observée)							25.764/482.894		=	**0.053**

TABLE 5.8 – *Exemple de calcul par fratrie - seuls les survivants répondent - avec pondération*

Nous nommerons cette pondération la "pondération GK", pour faire simple. Celle-ci élimine les biais associés à la sous-représentation des

30. Car dans ce cas, la probabilité de n'avoir aucun survivant est plus élevée.

31. "We replace the simple average of M_j in $q = \sum_{j=1}^{N} M_j$ with the weighted average of M_i, using weight $W_i = B_i/S_i$", p.572

fratries à mortalité élevée. Il subsiste uniquement un biais structurel lié à l'*absence d'information pour les fratries sans survivants*. En d'autres termes, ce mode de calcul permet de procurer les taux relatifs aux fratries pour lesquels il reste au moins un individu qui peut répondre à l'enquête (quelle que soit la taille de ces fratries). Ceci constitue donc un estimateur légèrement biaisé à la baisse, que la seconde étape - l'extrapolation - viendra relever [32].

C'est donc au niveau des calculs menés *par fratrie* que s'applique comme telle la pondération de Gakidou et King (2006). Cela dit, généralement l'estimation de la mortalité se base sur le calcul des décès et de l'exposition déclarés par *chaque survivant* (plutôt qu'à partir des taux observés par chacun dans la fratrie). Nous pensons qu'il suffit alors de pondérer les décès ainsi que les frères et sœurs observés en début de période par l'inverse du nombre de survivant ($1/S_i$ ou $1/(n-x)$). L'équation 5.12 est revue de la façon suivante :

$$\hat{q} = \frac{\sum g(n) \sum_{x=0}^{n-1} x \left[(n-x)/(n-x)\right] f(x)}{\sum g(n) \sum_{x=0}^{n-1} n \left[(n-x)/(n-x)\right] f(x)} = \frac{\sum_{i=1}^{n} D_i/S_i}{\sum_{i=1}^{n} B_i/S_i} \tag{5.17}$$

Ce qui équivaut bien à l'équation 5.15. La distinction est importante, car nous avons trouvé différentes applications dans la littérature (*cf.* infra) : en cas de calcul par fratrie, il faut pondérer par B_i/S_i, en cas de calcul individuel, il faut pondérer par $1/S_i$.

En somme, la démonstration de Gakidou et King (2006) consiste jusqu'ici à montrer qu'il est possible de neutraliser les perturbations induites par l'association entre la taille des fratries et la mortalité. Avant d'appliquer cette approche aux enquêtes EDS, détaillons ce qu'ils proposent pour estimer les décès survenus dans les fratries sans survivant.

Correction par extrapolation pour les fratries sans survivants

Pour tenir compte du biais résiduel lié à l'absence d'information sur les fratries sans survivants, Gakidou et King (2006) suggèrent de recourir à l'extrapolation. Leur approche consiste à estimer un nombre de décès dans les fratries sans survivants (un nombre noté ζ) à partir des fratries pour lesquels il subsiste *encore au moins un survivant*. L'équation 5.15 est donc revue en conséquence :

$$\hat{q} = \frac{\sum g(n) \left[\sum_{x=0}^{n-1} x\, f(x)\right] + n f(n)}{\sum g(n) \left[\sum_{x=0}^{n-1} n\, f(x)\right] + n f(n)} = \frac{\sum_{i=1}^{n} D_i/S_i + \hat{\zeta}}{\sum_{i=1}^{n} B_i/S_i + \hat{\zeta}} \tag{5.18}$$

Il s'agit simplement de rajouter, au dénominateur comme au numérateur, le nombre de décès survenus dans les fratries sans survivants (ζ) [33].

32. Notons que ces équations valent pour les calculs menés par *cohorte*. Lorsque nous travaillons en personnes-années, le dénominateur change ; il correspond à la somme des totaux d'années d'exposition observées (notés B_i^*), rapportée au nombre de répondants (toujours S_i).

33. Un nombre qui est bien sûr relatif au nombre total de survivants.

La démarche pour estimer ζ consiste à ajuster une courbe sur le nombre de décès observés selon le nombre de survivants observés ($s = 1, 2, 3\ldots$). A partir de cette courbe, on prédira ensuite le nombre de décès selon s, pour pouvoir extrapoler au cas où s vaut zéro. Il s'agit donc de calculer le nombre de décès déclarés dans les fratries avec 1 survivant, puis 2 survivants, puis 3 survivants, etc [34], afin de prolonger la tendance. Sur base de l'examen des données EDS, Gakidou et King (2006) proposent deux approches pour ajuster le nombre de décès :

1. D'abord une courbe du second degré ajustant le logarithme du nombre de décédés *par survivant* dans les fratries de 1, 2, 3,... survivants :
$$ln(\sum_{i:S=s} D_i/S_i) = \beta_0 + \beta_1 s + \beta_2 s^2 \ (s = 1, \ldots, 7)$$
2. Alternativement, une courbe du second degré ajustant simplement le logarithme du nombre *absolu* de décédés. Selon les auteurs, cette approche semble mieux ajuster les données :
$$ln(\sum_{i:S=s} D_i) = \beta_0 + \beta_1 s + \beta_2 s^2 \ (s = 1, \ldots, 7)$$

Dans les deux cas, une transformation de β_0 procure $\hat{\zeta}$ [35]. Nous avons procédé à ce calcul en conservant les enquêtes africaines de l'échantillon de Gakidou et King (2006) (16 enquêtes pour 14 pays) [36]. Le graphe 5.7 présente les différentes valeurs du logarithme du nombre de décédés [37] et les résultats de l'ajustement avec une courbe de second degré. On note effectivement un très bon ajustement dans quasiment toutes les enquêtes.

Ceci fait dire à Gakidou et King (2006) qu'il y a là une relation stable qui permet de prédire avec une certaine confiance le nombre de décès quand s vaut 0 [38]. Ils précisent toutefois qu'il est tout à fait possible que la courbe des décédés selon le nombre de survivant ne se prolonge pas pour $s = 0$. Par exemple, on peut s'attendre à un plus grand nombre de fratries sans survivants en cas de *forte corrélation des risques de décès entre les fratries*. Ce sera le cas dans les contextes de guerres et de génocides, si certaines familles sont plus particulièrement la cible d'attaques, par exemple (un problème discuté par de Walque et Verwimp (2010) pour le Rwanda).

Obermeyer et al. (2010) utilisent *une approche alternative* pour récupérer cette information sur les fratries sans survivants. Ils parlent de cette information manquante comme du "*zero-female-survivor bias*". Ils montrent

34. Les auteurs suggèrent de s'en tenir à 7 survivants ; au delà, les fluctuations aléatoires sont trop importantes.

35. Plutôt que de prendre l'exponentielle de β_0, les auteurs suggèrent de calculer $exp(\beta_0 + \sigma^2/2)$, avec l'erreur standard de la régression noté σ : "One might think that we could merely exponentiate the constant term, e^{β_0}, to remove the log scale, but this procedure is biased because the expected value of the log (which the regression estimates) is not equal to the log of the expected value (which this calculation would produce)." Gakidou et King (2006), p.574.

36. Nous ne retenons pas les données de l'Érythrée car elles ne sont pas distribuées.

37. Exprimé sous forme de la proportion des décès totaux.

38. "...although one should not make too much of a close fit of a model with three parameters to seven data points, the fact that the estimated relationship between sibship size and mortality is highly similar across this long list of diverse countries estimated independently is additional evidence that we have found a persistent, stable pattern that may be useful in extrapolating to deaths in families with zero survivors.", Gakidou et King (2006), p. 580

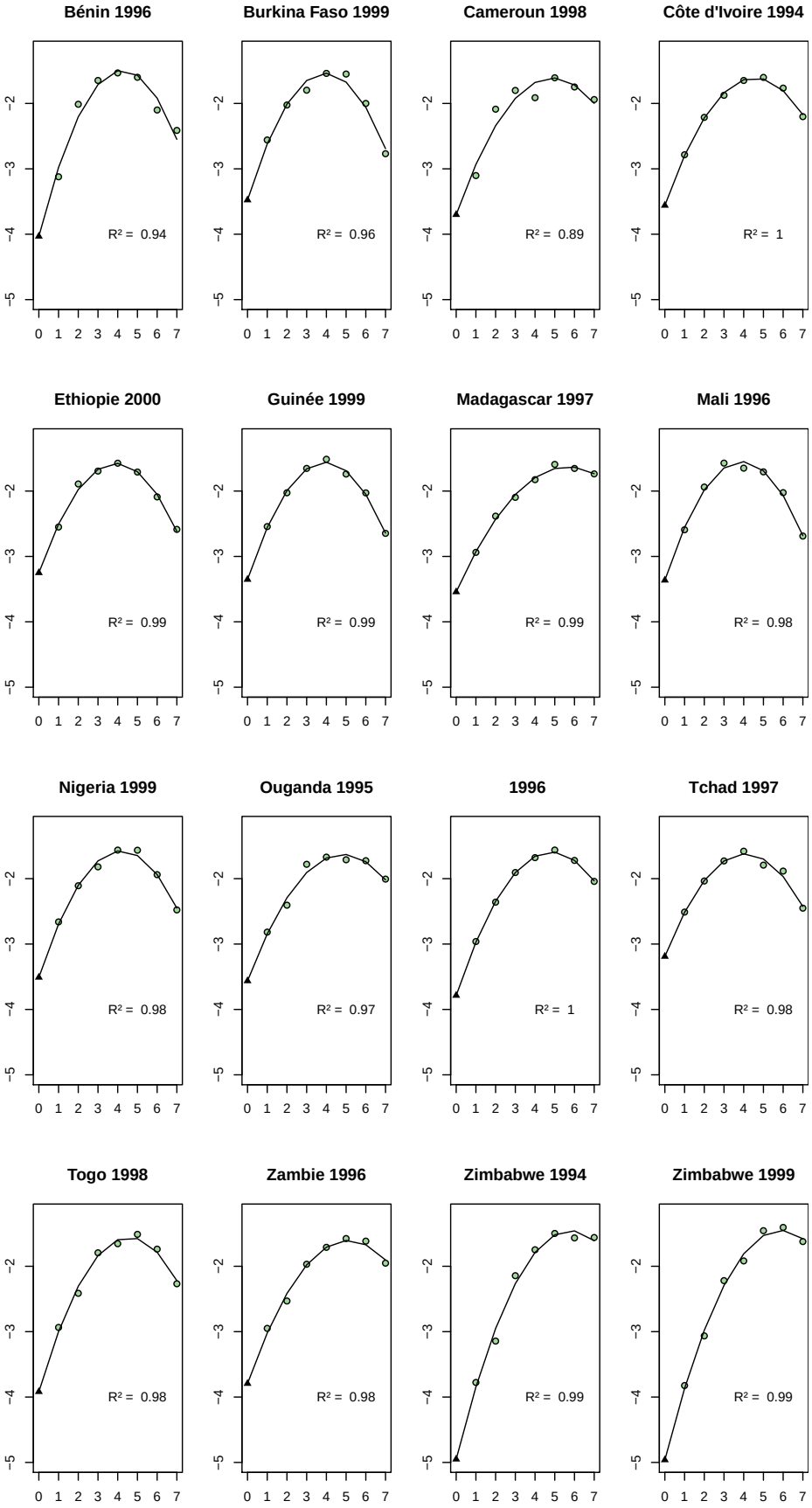

FIGURE 5.7 – *Logarithme du nombre de décédés selon la taille des fratries, différentes enquêtes EDS - nos propres calculs à partir des données EDS sur l'échantillon retenu par Gakidou et King (2006).*

que dans un cadre binomial, pour une taille de fratrie donnée (n), la probabilité de décès *observée* (p_n^{obs}) peut s'exprimer comme une fonction de la probabilité de décès *réelle* (notée p plutôt que p_n pour simplifier). Leur démonstration est relativement simple :

$$p_n^{obs} = \frac{\sum_{x=0}^{n-1} x\, f(x)}{n - \sum_{x=0}^{n-1} n\, f(x)} = \frac{np - np^n}{n - np^n} = \frac{p - p^n}{1 - p^n} \tag{5.19}$$

L'erreur relative qui caractérise la probabilité observée peut donc être facilement calculée. C'est bien sûr le cas pour des valeurs de n supérieures à 1, car si n vaut 1, p_{obs} sera toujours nulle. Nous avons illustré cette erreur relative [39] dans le graphe 5.8 pour différentes valeurs de n et de p, en supposant toujours une distribution binomiale. La sous-estimation augmente de façon quasiment linéaire avec la mortalité et diminue - logiquement - à mesure qu'augmente la taille des fratries : plus les fratries sont de grande taille, plus la probabilité que tous ses membres décèdent est faible. Lorsque n s'élève au dessus de 4, l'erreur relative est limitée - sauf pour des valeurs de p élevées (supérieures à 0.5).

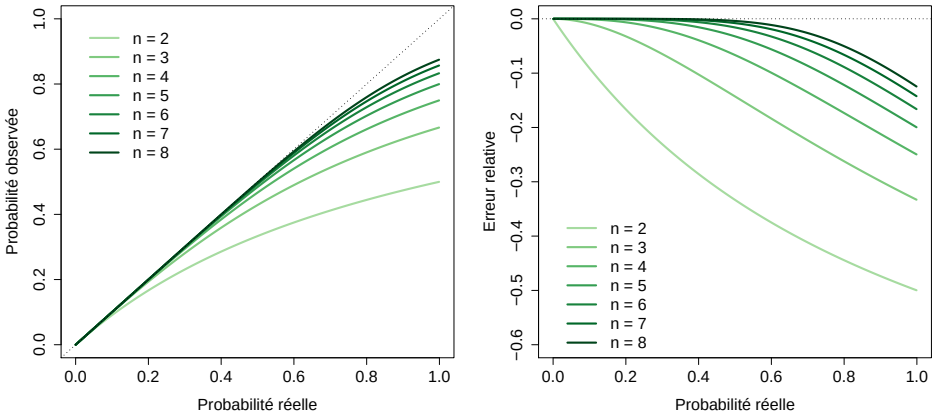

FIGURE 5.8 – *Probabilités réelles et observées - Erreur relative due aux décès non observés selon la taille des fratries et la mortalité - exemple de calcul dans un cadre binomial.*

A partir de là, Obermeyer et al. (2010) proposent de corriger les valeurs des ${}_{45}q_{15}$ qu'ils obtiennent des EDS au moyen de leur modèle de régression (présenté p.233). La démarche pour procéder à cette correction est la suivante :

1. Ils utilisent l'équation 5.19 pour calculer de façon algébrique les valeurs des proportions de décès réelles à partir des proportions de décès observés. Ce calcul est fait après avoir pondéré les données par les "poids GK", et uniquement pour les fratries dont la taille est inférieure à 3 ($n \leq 2$) [40].
2. Ils supposent ensuite une relation linéaire entre la taille des fratries et la proportion de décédés et extrapolent pour obtenir la proportion de décès dans les fratries où n vaut 1.

39. Calculée comme $\frac{p_{obs} - p_{reelle}}{p_{reelle}}$.

40. L'équation 5.19 peut facilement se résoudre algébriquement dans ces deux cas (Obermeyer et al. 2010). Si la probabilité corrigée est notée p_c et la probabilité observée p_{obs}, on

3. A partir de ces pourcentages corrigés et du nombre de décès observés, ils estiment le nombre total de décès (observés et non observés), quelle que soit la taille des fratries [41].
4. En rapportant ce nombre total de décès au nombre de décès observés, ils obtiennent un coefficient qui est ensuite utilisé pour revoir à la hausse les estimations de ${}_{45}q_{15}$ issues du modèle de régression.

Obermeyer et al. (2010) estiment que sur l'ensemble des données EDS qu'ils exploitent, cette correction a pour effet de relever les valeurs de ${}_{45}q_{15}$ de 0.2 à 4.0%. Cet ajustement n'est donc pas celui qui a l'impact le plus important sur leurs estimations, c'est plutôt la pondération qu'ils utilisent qui va relever sensiblement les estimations.

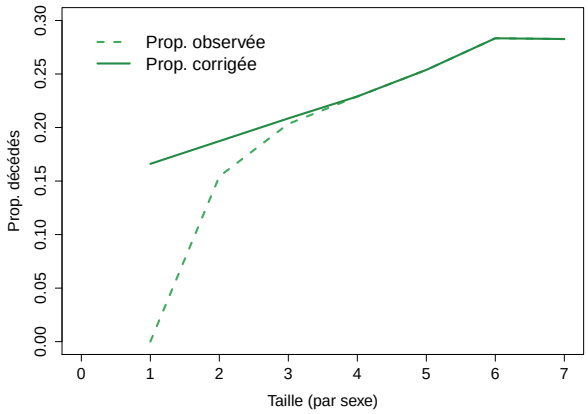

FIGURE 5.9 – *Proportion de sœurs décédées (Di) (pondérées) selon le nombre de sœurs à l'origine (Bi) - application de la méthode de Obermeyer et al. (2010) - nos calculs à partir de l'enquête EDS Mali 2006*

Ils illustrent le résultat de cet ajustement à partir de l'enquête malienne de 2006, et nous avons répliqué leurs calculs pour produire le même

peut établir :

$$
\begin{aligned}
n = 2 \Rightarrow & p_c = p_{obs}/(1 - p_{obs}) \\
n = 3 \Rightarrow & p_c = \frac{(1 - p_{obs}) - \sqrt{-3p_{obs}^2 + 2p_{obs} + 1}}{2p_{obs} - 2}
\end{aligned}
$$

41. A nouveau, cette valeur peut se calculer algébriquement. Notons $N(k)$ le nombre de fratries de taille k. Notons $N(k)_{obs}$ le nombre de fratries observés parmi ces fratries de taille k (c'est-à-dire celles où il subsiste au moins un survivant). Notons $D(k)$ le nombre de décès totaux et $D(k)_{obs}$ le nombre de décès observés. A partir de ces différents éléments, on peut établir, toujours en supposant une distribution binomiale des risques p, les relations suivantes (Obermeyer et al. 2010) :

$$
\begin{aligned}
N(k)_{obs} &= N(k) - p^k N(k) \\
N(k) &= N(k)_{obs}/(1 - p^k) \\
D(k) &= D(k)_{obs} + p^k \times N(k) \times k \\
D(k) &= D(k)_{obs} + p^k \times \frac{N(k)_{obs}}{(1 - p^k)} \times k
\end{aligned}
\tag{5.20}
$$

graphe (5.9)[42]. Nous obtenons une proportion de frères et sœurs décédés qui s'élève à 0.17 lorsque $n = 1$ et à 0.19 et 0.21 lorsque n vaut respectivement 2 et 3. Ces proportions sont logiquement supérieures à celles observées : 0 quand n vaut 1, 0.15 quand n vaut 2 et 0.20 quand n vaut 3. En recalculant le nombre de décès total estimé pour les fratries de taille $n < 10$[43], nous estimons qu'un peu plus de 2% des décès n'ont pas été observés dans l'enquête malienne de 2006.

Y a-t-il réellement une association entre la taille des fratries et la mortalité aux âges adultes ?

Pour justifier leur pondération, Gakidou et King (2006) indiquent qu'à partir d'une trentaine d'enquêtes EDS, ils observent une *forte association positive entre la taille des fratries et la mortalité*. Pour le démontrer, ils calculent des proportions de frères et sœurs décédés par taille de fratrie en appliquant leur pondération et estiment ensuite le coefficient de corrélation entre les tailles des fratries et ces proportions. En se limitant aux EDS africaines de leur échantillon, nous avons procédé aux mêmes calculs, ce qui procure les valeurs reprises dans la quatrième colonne du tableau 5.9 (ρ_1). Elles sont identiques à celles publiées par Gakidou et King (2006) à partir des mêmes enquêtes.

	Pays	Année	ρ_1	ρ_2	Sign.(2)
1	Bénin	1996	0.95	0.97	***
2	Burkina Faso	1999	0.95	0.97	***
3	Cameroun	1998	0.75	0.81	**
4	Tchad	1997	0.93	0.96	***
5	Côte d'Ivoire	1994	0.75	0.89	**
6	Ethiopie	2000	0.71	0.90	***
7	Guinée	1999	0.80	0.97	***
8	Madagascar	1997	-0.19	-0.16	
9	Mali	1996	0.86	0.97	***
10	Nigeria	1999	0.93	0.96	***
11	Togo	1998	0.74	0.92	***
12	Ouganda	1995	-0.06	0.33	
13	Tanzanie	1996	0.82	0.93	***
14	Zambie	1996	0.47	0.73	*
15	Zimbabwe	1994	0.76	0.85	**
16	Zimbabwe	1999	0.69	0.67	*

TABLE 5.9 – *Corrélations entre la taille des fratries et la mortalité - enquêtes EDS (nos calculs à partir de Gakidou et King (2006) : mortalité corrigée (ρ_1) et non corrigée (ρ_2)*

Pour être précis, il nous faut noter que ces corrélations sont en fait obtenues à partir des taux de mortalité *corrigés* pour tenir compte de l'ab-

42. Les pondérations utilisées par Obermeyer et al. (2010) pour construire ce graphe sont de type B/S, avec B et S calculés sur tous les frères et sœurs.

43. Ce facteur de correction varie logiquement avec les valeurs de n retenues pour calculer les décès totaux.

sence d'information dans les fratries sans survivants[44] (au moyen de la démarche présentée p.250). Nous pouvons recalculer les corrélations sans cette correction : elles sont reprises dans la cinquième colonne (ρ_2) du tableau 5.9. Ces nouvelles corrélations "non ajustées" s'avèrent dans l'ensemble plus élevées, ce qui est logique, puisque l'ajustement relève davantage la mortalité dans les petites fratries (la probabilité que les fratries disparaissent complètement est plus élevée quand les fratries sont plus petites). Quoi qu'il en soit, les proportions de frères et sœurs décédés augmentent sensiblement avec la taille des fratries. C'est bien la conclusion qu'en tirent Gakidou et King (2006) :

> "[These results] demonstrate unambiguously that mortality is not empirically independent of sibship size, as the standard estimator assumes. In the vast majority of surveys, the correlation is very high, often above 0.90. In two surveys, the correlation is negative. Any deviation from a zero correlation invalidates the standard estimator, but this table does not even suggest a tendency toward a zero correlation". - Gakidou et King (2006), p.579.

Ce constat nous amènerait donc à penser qu'une exploitation classique de type EDS des données risque de *surestimer* la mortalité. Il y aurait davantage de répondants pour témoigner de risques de mortalité plus élevés.

C'est ici que nous prenons nos distances par rapport à la démarche de Gakidou et King (2006) et de Obermeyer et al. (2010). Leurs calculs sont en effet menés sur *tous les frères et sœurs*. En d'autres termes, les proportions de décédés sont calculées sans tenir compte ni de l'âge, ni du sexe des frères et sœurs. C'est le cas des proportions ainsi que des poids B_i/S_i. Or, à partir des données EDS, les problèmes de biais de sélection vont se poser s'il y a une association entre la mortalité aux âges *adultes* et le nombre de *répondants potentiels* (c'est-à-dire généralement le nombre de femmes âgées de 15 à 49 ans). Le calcul mené par Gakidou et King (2006) mettant en évidence de fortes corrélations positives entre la taille des fratries et la mortalité nous paraît donc peu convainquant :

1. Les *enfants décédés avant 15 ans* représentent une fraction importante des décès, et ces décès sont susceptibles d'influencer les corrélations obtenues, alors qu'il s'agit ici de travailler uniquement sur les adultes.
2. De la même façon, les *hommes* sont pris en compte dans le calcul des poids B_i/S_i ainsi que dans celui des proportions de décédés, alors que seules les femmes de 15 à 49 ans répondent à l'enquête.
3. Comme ce calcul ne tient compte ni de l'âge du répondant ni de celui de ses frères et sœurs, l'ampleur de la corrélation observée va être surtout déterminée par des effets plus "structurels". Par exemple, si la fécondité a baissé de façon continue sur les dernières décennies,

44. Même si les auteurs disent l'inverse dans le texte. Notons que les enquêtés de moins de 25 ans sont exclus des calculs car toutes leurs mères n'ont pas terminé leur période de reproduction. Les fratries de plus de 10 membres sont exclues également. Ces choix sont ceux de Gakidou et King (2006), et nous les adoptons ici.

les tailles de fratrie les plus élevées seront observées *chez les enquêtées les plus âgés*. Or les frères et sœurs de ces enquêtées plus âgés ont été soumis *plus longtemps* au risque de décéder. Dans ce cas, on observera une corrélation entre la taille des fratries et les proportions de décès, mais cela tient à la baisse de la fécondité, et non à des risques de mortalité plus élevés dans les plus grandes fratries. Et nous pourrions imaginer d'autres mécanismes qui donneraient lieu à une telle corrélation, à partir par exemple des changements de mortalité. Ainsi, pour peu que les fratries plus nombreuses soient plus anciennes et que la mortalité ait baissé, on pourrait également observer une corrélation positive.

Selon nous, les fortes corrélations mises en évidence dans le tableau 5.9 indiquent uniquement que le *nombre total* de frères et sœurs déclaré *par les femmes de 15 à 49 ans* est associé à la *proportion de frères et sœurs décédés à tous âges*. Mais ce faisant, elles ne disent pas grand chose sur l'association qu'il pourrait y avoir entre des fratries plus nombreuses aux âges adultes et des risques de mortalité adulte plus élevés. Pour montrer qu'elles saisissent en fait tout autre chose, nous avons appliqué les mêmes calculs *sur les résultats des microsimulations africaines* présentées dans le chapitre 2. Dans ces dernières, la taille des fratries n'est pas associée à la mortalité, et nous ne devrions donc pas observer de corrélation significative. Pour mener ces calculs, les résultats des microsimulations sont restructurés de façon à obtenir une configuration EDS, c'est-à-dire que nous n'avons conservé que les fratries déclarées par les femmes survivantes de 25 à 49 ans [45].

	Pays	ρ_1	ρ_2	Sign.(2)
1	Bénin	0.29	0.56	
2	Burkina Faso	0.77	0.86	**
3	Cameroun	0.75	0.85	**
4	Côte d'Ivoire	0.66	0.91	***
5	Ethiopie	0.24	0.71	*
6	Guinée	0.69	0.84	**
7	Madagascar	0.58	0.70	*
8	Mali	0.15	0.82	**
9	Nigeria	0.33	0.75	*
10	Ouganda	0.27	0.67	*
11	Tanzanie	0.76	0.80	**
12	Tchad	0.17	0.44	
13	Togo	0.28	0.52	
14	Zambie	0.56	0.84	**
15	Zimbabwe	0.91	0.95	***

TABLE 5.10 – *Corrélations entre la taille des fratries et la mortalité - calculs menés sur les populations microsimulées*

Le tableau 5.10 reprend pour différentes simulations les deux séries des coefficients de corrélation (ρ_1 et ρ_2), calculées comme précédemment [46]. On note à nouveau une forte association entre la taille des fratries

45. Les répondants de moins de 25 ans sont éliminés car leur fratrie n'est pas complète. *Cf.* note 44.

46. Les données sont assez erratiques car elles sont basées sur une seule série de mi-

et la proportion de décédés, alors que les risques de décès ne varient pas avec la taille des fratries. Nous en concluons que les fortes corrélations obtenues par Gakidou et King (2006) sont en fait induites par la façon avec laquelle les données sont exploitées (sans tenir compte des âges, etc.). *Il n'y a pas, jusqu'alors, de résultats empiriques qui iraient dans le sens d'une association entre la taille des fratries (aux âges adultes) et la mortalité de ces adultes.* Ceci n'invalide pas le recours aux pondérations qu'ils suggèrent, mais cela invite à chercher une telle association en contrôlant mieux les différents éléments en jeu.

Il nous paraît plus rigoureux de chercher à savoir si les risques de décès varient effectivement avec le nombre de frères et sœurs adultes en utilisant un modèle de régression. Cela permet de tenir compte des âges des frères et sœurs, des âges des enquêtés ainsi que des tendances de la mortalité. Il faut également mieux circonscrire les plages d'âges retenues pour mener les calculs, en fonction des caractéristiques des enquêtés. En d'autres termes, il faut tenir compte du fait que bien souvent, seules les femmes de 15 à 49 ans sont enquêtées.

Par exemple, afin de disposer d'une cohorte homogène, nous ne retiendrons que les femmes survivantes 10 ans avant l'enquête et âgées *à l'époque* de 20 à 39 ans. Les frères sont éliminés du calcul, ainsi que les enquêtées âgées de moins de 30 ans au moment l'enquête (car la fratrie des plus jeunes évolue encore, puisque leur mère est toujours en âge de reproduction).

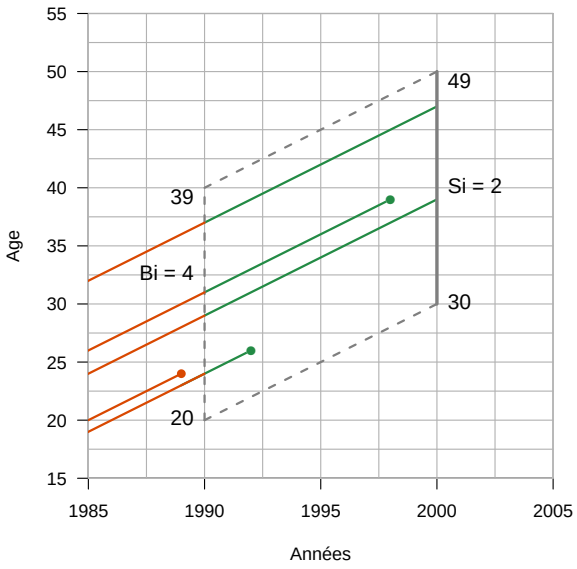

FIGURE 5.10 – *Sélection des enquêtées et des sœurs pour le calcul de l'association entre la taille des fratries aux âges adultes et la mortalité féminine.*

La période sous observation et les âges qui y correspondent sont présentés dans le graphe 5.10, sous la forme d'un diagramme de Lexis. Dans

crosimulations, soit, par pays, sur 30 000 survivants en 2010. Mais notre objectif n'est pas ici d'obtenir des estimations précises mais de montrer que les simulations procurent elles aussi des corrélations élevées.

cette fratrie, par exemple, quatre sœurs étaient survivantes en 1990, et deux sœurs sont décédées durant la période d'observation. Toutes les autres sœurs sont supprimées du calcul (y compris, dans notre exemple, celle décédée à 24 ans un peu avant 1990).

On cherche ensuite à savoir si les risques de décès durant la période d'observation varient de façon *significative* avec le nombre de sœurs présentes au début de cette période. Si c'est le cas, les conditions ne seront pas réunies pour l'estimation classique et il faudrait alors recourir à un mécanisme de pondération comparable à celui proposé par Gakidou et King (2006). Pour mener les calculs, nous optons pour une version simplifiée du modèle de Timaeus et Jasseh (2004) présenté plus haut, en recourant toujours à la régression de Quasi-Poisson, mais en ajustant le modèle sur les enquêtes prises individuellement[47]. Nous y ajoutons comme variable explicative l'âge des enquêtées, car nous soupçonnons que les enquêtées plus âgées (dont les tailles de fratries devraient généralement être plus élevées) vont avoir tendance à davantage sous-déclarer les décès (*cf.* infra).

Pour chacune des enquêtes EDS, nous ajustons deux modèles ; l'un en introduisant, sous forme continue, le nombre de sœurs de 20 à 39 ans au début de la période d'observation, l'autre sans cette variable. Nous comparons ensuite ces deux modèles emboîtés au moyen d'un test du rapport des vraisemblances[48]. Le tableau 5.11 reprend les *p-valeurs* issues de cette analyse pour les 57 enquêtes analysées. Ces valeurs indiquent si l'information apportée par la variable relative aux tailles des fratries représente un apport significatif ou non. Si c'est le cas, ces p-valeurs sont inférieures à 0.05. On constate que dans une très large majorité d'enquêtes, *la mortalité féminine entre 20 et 49 ans ne semble pas significativement associée au nombre de sœurs adultes présentes au début de l'observation*. Ce n'est le cas que dans 7 enquêtes seulement (au seuil 0.05), dont 3 au Kenya et 2 au Mozambique.

Ces résultats sont provisoires et ils méritent certainement une attention plus particulière, notamment pour établir si les cas du Kenya et du Mozambique ne sont pas dus à des erreurs de données (par exemple, une sous-déclaration des décès plus élevée dans certaines fratries). Mais dans l'ensemble, ils nous amènent à penser que l'exploitation "classique" des données reste valable et que les biais de sélection sont vraisemblablement assez faibles.

47. Nous introduisons toujours un schéma de mortalité standard, mais les coefficients qui y correspondent sont donc spécifiques à chaque enquête. La hausse de la mortalité due au sida est également spécifique à chaque enquête.

48. Dans `R`, il s'agit de la fonction `anova` après avoir ajusté les modèles au moyen de la commande `glm`.

Pays	Année	p-valeur	Sign	Pays	Année	p-valeur	Sign	Pays	Année	p-valeur	Sign
Afrique du Sud	1998	0.503		Lesotho	2004	0.751		Ouganda	2001	0.221	
Bénin	1996	0.857		Madagascar	1992	0.619		Ouganda	2006	0.970	
Bénin	2006	0.958		Madagascar	1997	0.059	.	Rwanda	2000	0.370	
Burkina Faso	1999	0.832		Madagascar	2004	0.791		Rwanda	2005	0.111	
Burkina Faso	2003	0.120		Madagascar	2009	0.202		Sénégal	1993	0.704	
Cameroun	1998	0.369		Malawi	1992	0.499		Sénégal	2005	0.541	
Cameroun	2004	0.147		Malawi	2000	0.003	**	Sierra Leone	2008	0.128	
Centrafique	1995	0.358		Malawi	2004	0.695		Swaziland	2006	0.077	.
Congo	2005	0.642		Mali	1996	0.415		Tanzanie	1996	0.186	
Congo RDC	2007	0.419		Mali	2001	0.768		Tanzanie	2004	0.084	.
Côte d'Ivoire	1994	0.683		Mali	2006	0.177		Tchad	1997	0.000	***
Ethiopie	2000	0.068	.	Mozambique	1997	0.011	*	Tchad	2004	0.580	
Ethiopie	2005	0.898		Mozambique	2003	0.002	**	Togo	1998	0.447	
Gabon	2000	0.741		Namibie	1992	0.326		Zambie	1996	0.062	.
Guinée	1999	0.748		Namibie	2000	0.099		Zambie	2002	0.764	
Guinée	2005	0.946		Niger	1992	0.602		Zambie	2007	0.118	
Kenya	1998	0.007	**	Niger	2006	0.282		Zimbabwe	1994	0.545	
Kenya	2003	0.013	*	Nigeria	2008	0.706		Zimbabwe	1999	0.412	
Kenya	2009	0.026	*	Ouganda	1995	0.541		Zimbabwe	2006	0.463	

TABLE 5.11 – *P-valeurs des tests du rapport des vraisemblances - comparaison entre le modèle de régression avec la 'taille des fratries' et le modèle sans cette variable - Enquêtes EDS (n=57)*

NOTE : Ces p-valeurs ne nous donnent pas d'indication sur l'**intensité** de l'association entre la mortalité et la taille des fratries. Il s'agit uniquement de la probabilité de se tromper quand on affirme que le modèle intégrant la taille des fratries apporte une information significative. Dans la majorité des enquêtes, ces valeurs sont si basses qu'on peut considérer que la mortalité ne varie pas avec la taille des fratries aux âges adultes.

La pondération proposée par Gakidou et King (2006) reste intéressante, mais à ce stade, il ne nous paraît pas certain qu'elle soit nécessaire : les risques de décès ne semblent pas varier significativement avec la taille des fratries aux âges adultes. De plus, *elle ne peut pas être appliquée comme telle aux EDS*. Il faut en effet calculer des valeurs de B_i et S_i qui tiennent compte des spécificités des données, notamment du fait que seuls les adultes de 15 à 49 ans procurent des informations sur leur fratrie. A ce jour, les différentes applications de cette pondération aux données d'enquête nous paraissent approximatives, et même parfois erronées [49]. Nous en discutons dans la section suivante.

Le problème de l'application aux enquêtes EDS

Quand il s'agit de traiter des biais de sélection qui caractérisent les données sur les frères et sœurs, la littérature récente présente une grande diversité d'approches et . . . une certaine confusion.

1. D'une part, les rapports d'enquêtes EDS *se basent toujours sur une exploitation classique*, c'est-à-dire sans pondération [50], et en excluant les enquêtés des calculs. Il en va de même pour les travaux menés avant que Gakidou et King (2006) ne publient leurs travaux. C'est le cas notamment des recherches menées par Timaeus et Jasseh (2004), ainsi que par Feeney (2001) pour le Zimbabwe, ou encore de Walque (2005) pour le Cambodge. Pour produire les estimations reprises dans le chapitre 7 ainsi que dans Reniers, Masquelier, et Gerland (forthcoming), nous avons opté pour la même approche.

2. Une deuxième série de travaux pondèrent les données au niveau individuel *par l'inverse du nombre de répondants potentiels* (c'est-à-dire les sœurs survivantes âgées de 15 à 49 ans). C'est le cas notamment d'une enquête irakienne menée par l'OMS, qui a adopté un module sur la fratrie comparable à celui des enquêtes EDS (Iraq Family Health Survey Study Group 2008). Pour analyser les risques de décès durant le génocide rwandais, de Walque et Verwimp (2010) font de même, mais en excluant les enquêtés des calculs.

3. Une troisième série de travaux *s'inspirent plus directement de la pondération de Gakidou et King (2006)*. Il s'agit surtout de chercheurs de l'*Institute for Health Metrics and Evaluation* (IHME), de l'Université de Washington. Dans un article s'inscrivant dans le champ très controversé de l'estimation des décès liés aux guerre, Obermeyer et al. (2008) exploitent le module portant sur la fratrie des enquêtes santé de l'OMS (WHS) [51]. Dans un autre article récent, Obermeyer et al. (2010) appliquent également la pondération de Gakidou et

49. Sauf peut-être les travaux qui portent sur l'enquête de mortalité irakienne (Iraq Family Health Survey Study Group 2008)

50. Mis à part celle qui vise à tenir compte de l'échantillonnage.

51. Dans lesquelles un seul adulte est tiré au hasard par ménage pour procurer des informations sur ses frères et sœurs (*cf.* chapitre 1). Sans doute la question des biais de sélection se pose-t-elle en des termes différents dans ce cas. . .

King (2006) aux enquêtes EDS. Nous avons vu qu'ils indiquent que cette pondération a un impact déterminant sur les estimations, avec des valeurs de ${}_{45}q_{15}$ augmentées de 6 à 66% selon les pays, avec une hausse de 28% en moyenne par rapport à l'estimation classique. Ceci leur fait dire que la méthode proposée permet de corriger les biais de sélection et de procurer des niveaux de mortalité plus plausibles. Ces estimations ont été ensuite reprises dans d'autres travaux, notamment dans un article récent qui fait le point sur la mortalité adulte au niveau mondial, publié dans le Lancet par Rajaratnam et al. (2010). La pondération de Gakidou et King (2006) joue également un rôle dans de nouvelles estimations de mortalité maternelle, basées partiellement sur les enquêtes EDS, et publiées elles aussi dans le Lancet par Hogan et al. (2010).

Si nous revenons quelques pas en arrière, il paraît *a priori* étrange que Obermeyer et al. (2010) obtiennent des niveaux de mortalité *plus élevés* avec la pondération GK. En effet, ceci voudrait dire que la taille des fratries est *négativement* associée à la mortalité (la mortalité serait plus faible dans les grandes fratries, et l'approche classique sous-estimerait donc la mortalité). Or c'est précisément l'inverse que Gakidou et King (2006) avaient mis en évidence : ils pointaient une forte corrélation *positive* entre tailles des fratries et mortalité ! Il y a donc apparemment une contradiction entre les travaux de Obermeyer et al. (2010) et ceux de Gakidou et King (2006), alors qu'il s'agit de la même équipe et que les premiers disent utiliser la pondération des seconds.

En fait, nous pensons que cette contradiction n'est qu'apparente. C'est plus largement l'application de la méthode de Gakidou et King (2006) aux EDS qui pose problème. Nous pensons que les niveaux de mortalité publiés par Obermeyer et al. (2008; 2010), et par Rajaratnam et al. (2010) sont *surestimés, particulièrement chez les hommes*. Cette surestimation de la mortalité masculine pourrait atteindre 15 à 20%, peut-être plus. Nous justifions cette position en trois points :

1. Premièrement, comme le font Gakidou et King (2006), Obermeyer et al. (2010) construisent leur pondération à partir de *tous les frères et sœurs*, quel que soit leur âge ou leur sexe. Ils font donc intervenir des frères et sœurs décédés dans l'enfance, alors que ces derniers ne contribuent ni aux durées d'exposition ni aux décès. De plus, les données utilisées par Obermeyer et al. (2010) sont uniquement celles collectées auprès des femmes, mais les pondérations retenues sont également *basées sur les frères*, alors que ceux-ci n'interviennent pas au niveau des biais de sélection [52].
2. Deuxièmement, sur un plan plus technique, nous avons distingué plus haut deux types de pondérations :

52. Cela affecte non seulement les valeurs des probabilités de survie estimées à partir du modèle mais également le coefficient de correction que les auteurs utilisent pour tenir compte du biais lié à l'absence d'information sur les fratries sans répondants. Notons que Gakidou et King (2006) avaient précisé que leur pondération devait être adaptée dans les cas où la survie masculine étaient estimée à partir des sœurs, et vice-versa. Mais ceci n'a pas été appliqué aux enquêtes.

  - Gakidou et King (2006) suggèrent de pondérer les données par B_i/S_i. Mais cette pondération vaut lorsque l'on travaille au niveau des *taux par fratrie observés par chaque enquêté*.
  - Si l'on travaille dans un fichier individuel, en faisant la somme des décès et des expositions observés par les enquêtés, il faut alors pondérer les données par *l'inverse du nombre de survivants*, soit $1/S_i$, ou $1/(n-x)$ selon la formulation choisie.

  Or Obermeyer et al. (2010) exploitent les données au niveau individuel dans un fichier personnes-périodes, en reprenant également les répondants [53], mais en appliquant ensuite des poids de type B_i/S_i [54]. Une erreur, selon nous, qui revient en fait à sur-représenter les fratries où il y a eu beaucoup de frères et sœurs dans l'enfance. Notons que si la mortalité *adulte* ne semble pas varier avec le nombre de frères et sœurs *adultes*, elle pourrait varier avec le nombre de frères et sœurs *dans l'enfance*.

3. Troisièmement, il y a un petit problème technique pour certaines enquêtes, qui introduit également un biais à la hausse. Dans 6 enquêtes EDS sur 10 environ pour l'Afrique, la taille des fratries de frères et sœurs des *enquêtés fils ou fille unique* n'a pas été codée comme "zéro" mais comme "manquante". Ces filles uniques ont été éliminées par Obermeyer et al. (2010), ce qui induit une sous-estimation des durées d'exposition [55]. Dans les enquêtes concernées, ceci représente seulement 1.7% des enquêtés en moyenne, et cela devrait donc le plus souvent avoir un faible impact sur les probabilités de décès. Mais cette proportion s'élève néanmoins à 4.2% dans l'enquête Guinée 2005 et à 5% dans l'enquête Congo RDC 2007.

Ces trois éléments concourent ensemble à la surestimation de la mortalité.

Nous pouvons illustrer en particulier les effets de la pondération en recourant aux microsimulations. Reprenons le cas du Mozambique, qui a déjà été utilisé pour comparer quatre configurations, selon le sexe des enquêtés (femmes uniquement ou hommes et femmes), et selon les plages d'âge retenues (tous âges ou 15-49 ans uniquement) (*cf.* graphe 5.5, p.240). A partir de là, nous avions montré que la mortalité estimée à partir des frères et sœurs de façon classique, sans pondération, permettait de retomber sur les niveaux de mortalité "réellement" à l'oeuvre dans les simulations [56].

53. "We first reshape each dataset so that there is one observation per sibling, including one observation for the respondent. We then reshape the data again, so that one observation refers to one person-year of one sibling. In this data structure, every member of a sibship receives an observation for every year that they were alive and an observation for the year in which they died, if they died.", p.7

54. "This method ... incorporates a family-level weight, $W_f = B_f/S_f$, where B_f is the original sibship size and S_f is the number of siblings in family f surviving to the time of the survey. ... To generate the final weight for each observation, we multiply W_f by the DHS sampling weight.", p. 11

55. Ceci peut s'observer à partir du nombre de frères et sœurs reportés par les auteurs dans leur tableau en annexe, ainsi que dans les do.file STATA distribués sur le site www.healthmetricsandevaluation.org.

56. Pour peu que la mortalité ne varie pas avec la taille des fratries.

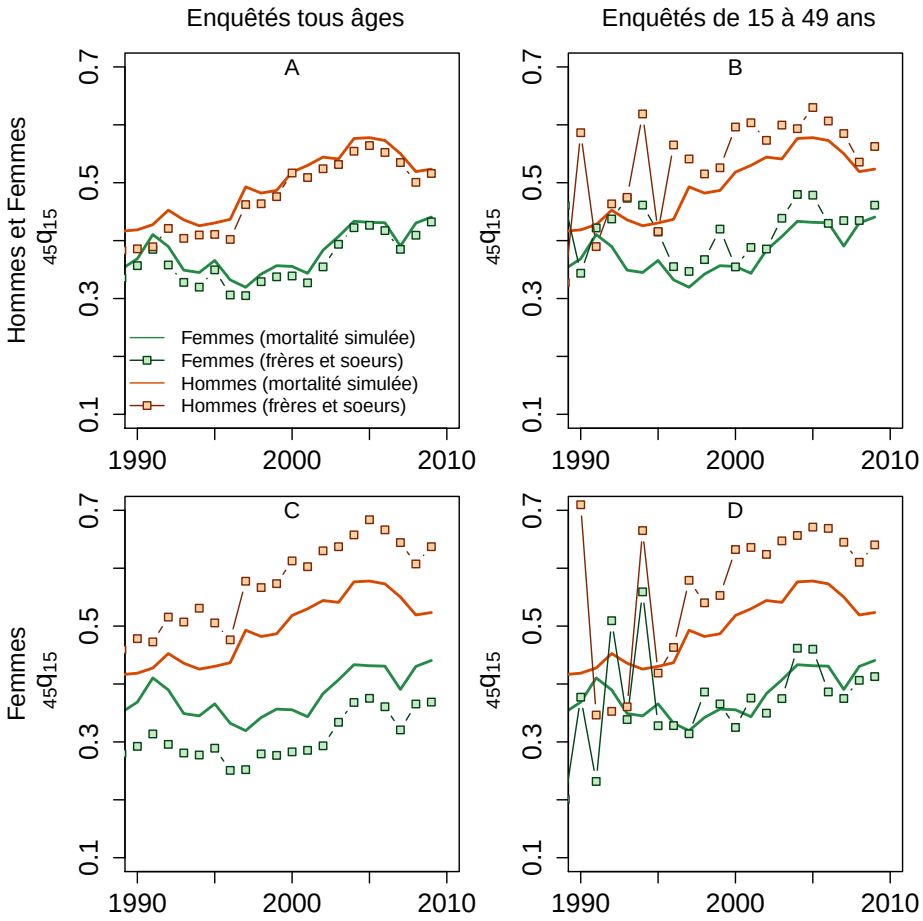

FIGURE 5.11 – *Probabilités de décès entre 15 et 60 ans simulées et recalculées à partir des frères et sœurs en appliquant la pondération de Obermeyer et al. (2010) - microsimulations - scénario Mozambique*

Ce n'est pas le cas si nous retenons les poids utilisés par Obermeyer et al. (2010)[57]. Comme précédemment, les deux figures du haut (A et B) du graphe 5.11 correspondent à la situation où les hommes et les femmes sont interrogés :

- Dans le graphe A, en haut à gauche, tous les survivants, quel que soit leur âge, procurent des informations sur leur fratrie. La pondération de Gakidou et King (2006) fonctionne bien : elle nous fournit des niveaux de mortalité très proches des niveaux "réels" des microsimulations. S'ils sont cependant légèrement inférieurs à la mortalité "réelle", c'est en raison de l'absence d'informations sur les fratries sans survivants. Une correction pour en tenir compte aurait pour effet de relever de quelques pourcents les quotients de mortalité, et de les ramener au niveau des quotients "réels".
- Dans le graphe B, seuls les adultes de 15 à 49 ans sont enquêtés. Dans ce cas, la mortalité tant féminine que masculine est surestimée. Les raisons de cette surestimation ne nous paraissent pas toutes évidentes, mais un des problèmes (sans doute le plus important) se situe du côté de la pondération, qui ne tient pas compte de la structure par âge des enquêtés ni de celle des frères et sœurs.

Dans les deux figures du bas (C et D), seules les femmes sont éligibles aux questions sur les frères et sœurs. Et dans ces deux cas, la mortalité masculine est sensiblement surestimée en procédant à la méthode utilisée par Obermeyer et al. (2010).

- Dans le graphe C, cette surestimation se double d'une sous-estimation de la mortalité féminine. Ceci est vraisemblablement lié aux différentiels sexuels de mortalité, et au fait que les poids sont calculés sans distinction de sexe.
- Dans le graphe D, qui correspond à la situation des EDS, la mortalité féminine est plus proche de la mortalité "réellement" à l'oeuvre dans les simulations. Par contre, la mortalité masculine est toujours surestimée. Dans cet exemple, entre 2000 et 2010, les valeurs recalculées de ${}_{45}q_{15}$ sont en moyenne supérieures de 18% aux valeurs tirées des simulations.

Les microsimulations ne nous permettent pas cependant de faire le tour de la question, notamment parce que les risques de décès y sont constants d'une fratrie à l'autre. *Avant de conclure cette section, nous avons donc appliqué l'approche de Obermeyer et al. (2010) aux enquêtes EDS.* Pour dégager des estimations de mortalité, nous retenons le modèle de régression présenté précédemment, repris à Timaeus et Jasseh (2004). Cela permet de comparer les estimations "corrigées" aux estimations qui sont obtenues classiquement (ces dernières sont détaillées dans le chapitre 7).

Les deux figures du haut du graphe 5.12 opposent, pour les hommes et les femmes, les valeurs obtenues avec la pondération de Obermeyer et al.

57. Notons que nous obtenons un graphe identique à celui-ci si nous pondérons les données par $1/S_i$ plutôt que par B_i/S_i. La sur-représentation des fratries de grande taille n'a pas d'effet dans les simulations, puisque la mortalité est identique quelle que soit cette taille. Ce sont donc plutôt les problèmes liés à l'utilisation de poids qui ne sont pas adaptés aux caractéristiques des répondants qui sont illustrés ici.

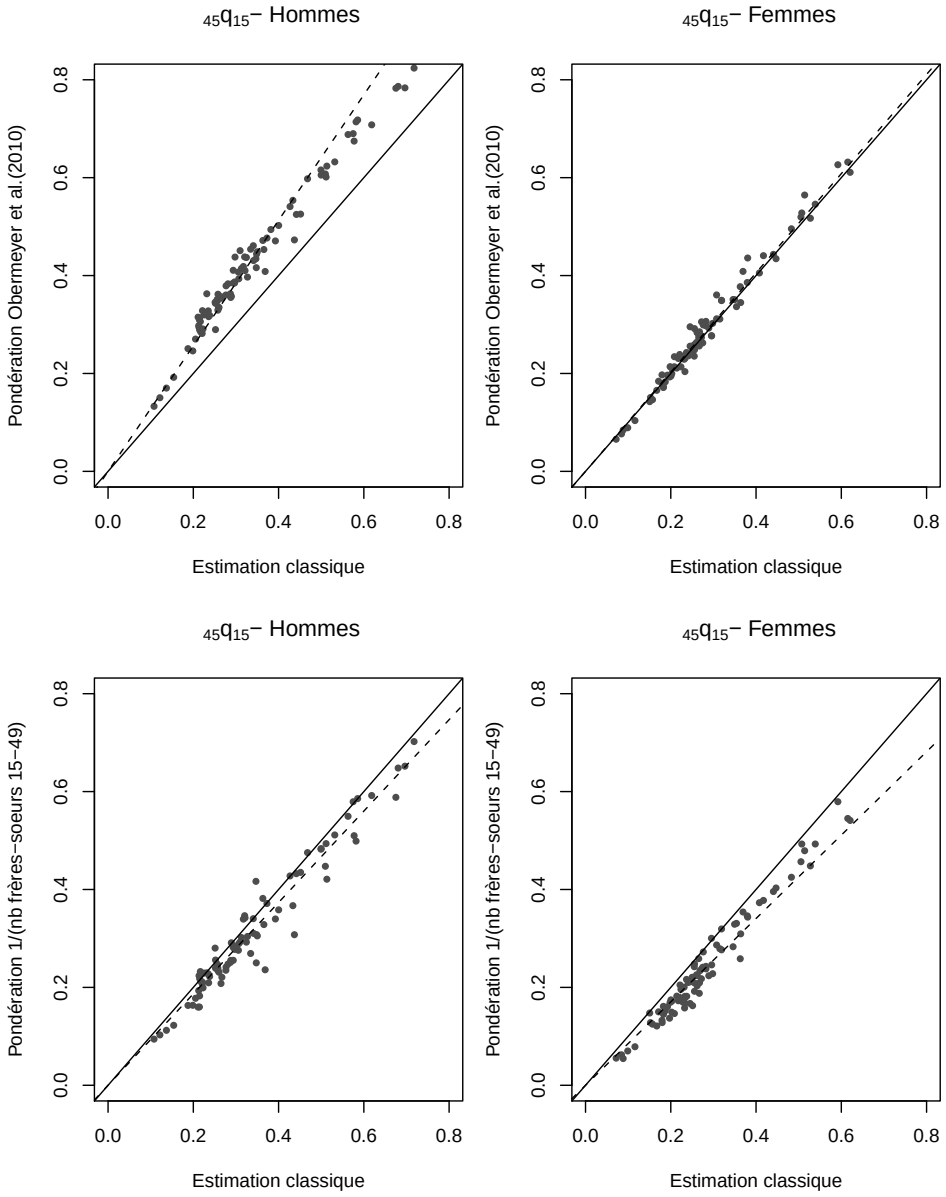

FIGURE 5.12 –
Panel du haut *: Comparaison des probabilités de décès entre 15 et 60 ans issues des données sur les frères et sœurs (enquêtes EDS), obtenues de façon "classique" (calcul EDS) et en appliquant la pondération de Obermeyer et al. (2010)*
Panel du bas *: Comparaison des probabilités de décès entre 15 et 60 ans issues des données sur les frères et sœurs (enquêtes EDS), obtenues de façon "classique" (calcul EDS) et en pondérant les données par l'inverse du nombre d'enquêtés potentiels.*

(2010) (en ordonnées), avec celles issues du calcul EDS (en abscisses)[58]. Les droites en pointillés correspondent à la médiane du rapport entre les premières et les secondes. Si l'on pense que les estimations obtenues classiquement sont fiables, alors la mortalité adulte masculine est sensiblement surestimée par Obermeyer et al. (2010). La médiane du rapport entre les deux séries de valeurs se chiffre à 1.28, ce qui correspond tout à fait à l'estimation que donnent ces auteurs de la différence entre les deux types d'exploitation des données. Par contre, on note un bon accord entre les deux séries de quotients ${}_{45}q_{15}$ chez les femmes[59]. Ces deux graphes viennent donc confirmer ce que la microsimulation avait déjà montré : chez les femmes, l'application que font Obermeyer et al. (2010) de la pondération semble procurer les mêmes résultats que l'exploitation classique, mais elle induit par contre une sensible surestimation chez les hommes. Il est d'ailleurs anormal que la correction affecte disproportionnellement un des deux sexes[60].

Quoiqu'il en soit, sur le plan théorique, puis à partir des exemples issus des simulations et des enquêtes, nous pensons avoir montré ici que *les niveaux plus élevés de mortalité obtenus par Obermeyer et al. (2010) résultent de l'application d'une pondération qui est à la fois erronée* (puisqu'il eut fallu pondérer les données par $1/S_i$) *et non adaptée aux enquêtes EDS*. Dans la mesure où ils se basent sur les mêmes calculs, cela affecte sans doute également les travaux récents de Rajaratnam et al. (2010), et de Hogan et al. (2010).

Selon nous, si l'on souhaite appliquer la pondération de Gakidou et King (2006) aux données d'enquêtes, il convient plutôt de pondérer les observations par l'inverse du nombre de répondants potentiels. Par exemple, on multipliera par 0.5 les observations dans les fratries où il subsiste deux femmes survivantes et âgées de 15 à 49 ans[61]. C'est cette pondération qui a été utilisée dans l'enquête de l'OMS menée en Irak en 2006 et 2007 (Iraq Family Health Survey Study Group 2008). On obtient alors les valeurs reprises dans les deux graphes du bas de la figure 5.12 (p.265). Les estimations "corrigées" s'avèrent là plus basses que les estimations classiques, ce qui paraît logique. En effet, ces estimations "corrigées" devraient correspondre aux quotients de mortalité pour les fratries où il subsiste au moins un enquêté. Elles sont inférieures aux estimations classiques de 7% chez les hommes[62] et de 15% chez les femmes.

58. Il s'agit des estimations prédites par le modèle de régression, pour les années 1985, 1990, 1995, 2000 et 2005 (selon la disponibilité des données EDS par pays). Un même pays peut donc se retrouver plusieurs fois sur ce graphe.

59. Même si les valeurs pondérées devraient ensuite être révisées à la hausse de quelques pourcents pour tenir compte du biais causé par l'absence de survivants

60. Lorsqu'ils présentent les résultats de leur pondération, Obermeyer et al. (2010) ne disent pas avoir rehaussé davantage la mortalité masculine. Or c'est ce que nous observons dans les microsimulations, ainsi qu'en appliquant cette pondération aux enquêtes. Cette question mérite un examen plus attentif.

61. Pour construire le graphe 5.12, nous avons tenu compte du fait que les hommes sont également répondants dans certaines enquêtes. Dans ce cas, les poids sont établis à partir du nombre de *frères* adultes survivants.

62. Cela correspond à la médiane des rapports entre les deux séries de valeurs.

Ceci nous conduit au second problème posé par la pondération de Gakidou et King (2006) (le premier étant celui de son application aux enquêtes). Cette approche nécessite en effet que l'on traite de l'*estimation du nombre de décès survenus dans les fratries sans survivants*, un problème qui ne se pose pas dans l'exploitation "classique", puisque les différents biais sont supposés s'annuler. Or ce calcul est entouré de nombreuses incertitudes, notamment si l'on souhaite pouvoir corriger les estimations dans le temps et par âge. Il n'existe pas à ce jour de solution pleinement satisfaisante pour "récupérer" cette information sur les fratries disparues. Nous avons vu que tant l'approche de Gakidou et King (2006) que celle de Obermeyer et al. (2010) se basent sur tous les frères et sœurs (enfants y compris), ce qui n'est pas adéquat. Il faudrait idéalement tenir compte des âges des frères et sœurs, mais cela complique sensiblement l'estimation.

Ceci conclut cette discussion des biais de sélection qui caractérisent les données sur les frères et sœurs. Dans la section suivante, nous traitons de l'exploitation indirecte qui peut être faite de ces données pour estimer la mortalité adulte.

5.3 Exploitation indirecte des informations portant sur la survie des frères et sœurs

Pourquoi vouloir accorder une attention particulière aux méthodes indirectes quand une exploitation directe des données sur les frères et sœurs est possible ? Nous y voyons au moins deux raisons :

1. D'une part, comme c'est le cas pour la mortalité des enfants, la comparaison entre les estimations directes et indirectes devrait permettre de mieux juger de la *plausibilité des niveaux et des tendances obtenues*. Les écarts entre les deux séries d'estimations sont souvent le signe d'une mauvaise qualité des données dans au moins un des systèmes. Par exemple, les estimations directes sont très sensibles aux erreurs faites sur les âges des frères et sœurs, ce qui n'est pas le cas des estimations indirectes.
2. D'autre part, les méthodes indirectes sont beaucoup *moins exigeantes en termes de collecte*. Elles ne requièrent que 3 à 4 questions, alors que les biographies complètes pour l'approche directe reposent typiquement sur 6 à 7 questions pour chaque frère et sœur. Une série d'enquêtes, comme les MICS, se contentent donc souvent des questions de type indirect.

Deux méthodes sont disponibles pour exploiter indirectement les données sur les frères et sœurs : celle développée par Hill et Trussell (1977), qui intègre tous les frères et sœurs (en ce compris les enfants), et celle proposée plus récemment par Timaeus et al. (2001), qui s'en tient aux frères et sœurs ayant survécu à 15 ans. A ces deux approches s'ajoute la "méthode des sœurs", plus connue, et mise au point par Graham et al. (1989) pour estimer spécifiquement la mortalité maternelle. Mentionnons également une technique indirecte parallèle, destinée à estimer l'intensité de la migration de façon indirecte à partir des frères et sœurs (Hill 1981,

Zaba 1987).

Avant d'appliquer la méthode de Timaeus et al. (2001) aux enquêtes EDS, nous présentons brièvement la démarche qui sous-tend le développement de ces approches indirectes. L'idée est la même que pour la survie des parents : il s'agit de convertir des indicateurs de structure (des proportions de survivants) en indicateurs de mouvements (des probabilités de survie). Nous ne rentrerons donc pas dans le détail des simulations, car elles sont très proches de celles présentées dans le chapitre 3 pour la survie des parents.

5.3.1 Formalisation de la proportion de frères et sœurs survivants

Comme pour les méthode des orphelins, tout part de la formalisation des proportions de frères et sœurs survivants. Par facilité, on séparera les frères et sœurs en deux catégories : du point de vue d'un individu donné, nommé *Ego*, on distinguera ceux nés avant sa naissance et ceux nés après. En effet, le nombre de "petits" frères et sœurs (ceux nés après) dépend de la survie des mères après la naissance d'*Ego*, ce qui n'est bien sûr pas le cas des frères et sœurs nés avant.

Frères et sœurs nés avant Ego (les aînés)

Le cas des frères et sœurs nés avant *Ego* est le plus simple. En suivant les développements de Keyfitz et Caswell (2005), le nombre de ces frères et sœurs survivants s'exprime comme ceci :

- Supposons dans un premier temps que la mère d'*Ego* ait accouché de ce dernier à l'âge x.
- Elle avait à l'époque déjà donné naissance à $\int_\alpha^x m(y)dy$ enfants, si $m(y)$ correspond au taux de fécondité par âge.
- Afin de supprimer la condition sur l'âge à l'accouchement, il faut prendre une moyenne sur tous les âges de reproduction, en pondérant chaque âge par le nombre de naissances qui surviennent à cet âge. Dans le cas d'une population stable, cela revient à multiplier par $[e^{-rx}\ {}_xp_0\ m(x)dx]$ puis à intégrer sur tous les âges x compris dans la période de reproduction (*cf.* chapitre 3).
- Le nombre de frères et sœurs nés avant *Ego* vaut donc :

$$S^{old} = \int_\alpha^\beta \int_\alpha^x m(y)\ \left[e^{-rx}\ {}_xp_0\ m(x)dx\right] dy \tag{5.21}$$

- Le nombre de frères et sœurs nés avant *Ego* ne dépend nullement de l'âge d'*Ego* (qu'il ait 5 ou 50 ans, il aura toujours le même nombre de frères et sœurs aînés). Par contre, son âge intervient pour le calcul du nombre de frères et sœurs nés avant *Ego* qui *survivent* jusqu'à l'enquête. Si l'on note a l'âge d'*Ego* et $x - y$ l'intervalle d'âge entre frères et sœurs, le nombre de frères et sœurs survivants nés avant *Ego* s'obtient ainsi (Keyfitz et Caswell 2005) :

$$S^{old,surv}(a) = \int_\alpha^\beta \int_\alpha^x m(y)e^{-rx}\ {}_xp_0\ m(x)\ {}_{a+x-y}p_0 dxdy \tag{5.22}$$

Frères et sœurs nés après Ego (les cadets)

Voyons maintenant le cas des "petits" frères et sœurs :

– Imaginons toujours que la mère d'*Ego* lui ait donné naissance à l'âge x. Imaginons qu'elle ait ensuite donné naissance à d'autres enfants, à l'âge $x+u$. Dans ce cas, u varie entre o et a, a étant l'âge actuel d'*Ego*[63]. Il faudrait donc que la mère ait survécu jusqu'à $x+u$. Le nombre de frères et sœurs nés après *Ego* peut alors s'exprimer comme ceci :

$$S^{young}(a) = \int_{\alpha}^{\beta} \int_{0}^{a} m(x+u)\, {}_{u}p_{x} du\, e^{-rx}\, {}_{x}p_{0}\, m(x) dx \tag{5.23}$$

– Comme l'âge d'*Ego* vaut a, l'âge de ses frères et sœurs cadets vaut $a-u$, et le nombre de ces frères et sœurs qui ont survécu se note donc :

$$S^{young,surv}(a) = \int_{\alpha}^{\beta} \int_{0}^{a} m(x+u)\, {}_{u}p_{x}\, {}_{a-u}p_{0}\, du\, e^{-rx}\, {}_{x}p_{0}\, m(x) dx \tag{5.24}$$

A partir de ces équations, des proportions de frères/sœurs survivants peuvent être simulées et être introduites dans des régressions avec les probabilités de survie comme variables dépendantes.

5.3.2 Conversion des proportions de survivants en probabilités de survie

Probabilités de survie depuis la naissance

Hill et Trussell (1977) ont été les premiers à proposer d'estimer la mortalité adulte à partir des frères et sœurs. Au départ des équations précédentes, ils fournissent une première série de coefficients, qui s'intègrent dans l'équation suivante :

$${}_{n}p_{0} = \beta_0(n) + \beta_1(n)\, {}_{5}S_{n} \tag{5.25}$$

Il s'agit donc ici de prendre en compte tous les décès, même ceux survenus dans l'enfance, et d'obtenir des probabilités de survie non conditionnelles, en d'autres termes de la naissance à un âge adulte allant jusqu'à 55 ans. Les coefficients β de ces auteurs sont repris dans le tableau 5.12. Ils sont établis à partir de 324 simulations, dont les paramètres sont proches de ceux qu'ils ont utilisés pour la survie des parents (*cf.* p.124).

Signalons que les *écarts d'âge entre frères et sœurs* jouent un rôle déterminant dans cette méthode indirecte. Combinés à l'âge des enquêtés, ce sont eux qui vont permettre d'approximer la durée d'exposition. En l'occurrence, ces écarts d'âge sont ici entièrement déterminés par les schémas de fécondité (tous rangs) retenus pour mener les simulations.

Pour être tout à fait précis, il faut toutefois noter que les proportions simulées par Hill et Trussell (1977) ont en fait été calculées à partir d'une version modifiée des équations 5.22 et 5.24. Les auteurs vont y intègrer des

63. Sans tenir compte à ce stade des durées d'aménorrhée postpartum ni de gestation.

n	β_0	β_1	R^2	CV
10	0.0033	0.9998	0.9997	0.0029
15	0.0223	0.9839	0.9995	0.0041
20	0.0216	0.9847	0.9996	0.0041
25	0.0129	0.9920	0.9997	0.0036
30	0.0102	0.9954	0.9997	0.0038
35	0.0115	0.9972	0.9997	0.0045
40	0.0134	1.0009	0.9995	0.0057
45	0.0143	1.0090	0.9993	0.0073
50	0.0125	1.0242	0.9990	0.0094
55	0.0075	1.0488	0.9987	0.0118

TABLE 5.12 – *Coefficients utilisés pour convertir les proportions de frères/sœurs survivant(e)s en probabilités de survie - Hill et Trussell (1977)*

modifications pour tenir compte de l'*infécondité durant les périodes de gestation et d'aménorrhée*, ainsi que d'une *moindre fécondité après une naissance* [64]. Sans cela, les écarts d'âge entre frères et sœurs seraient sous-estimés [65].

Mortalité maternelle

Graham et al. (1989) ont revu la méthode de Hill et Trussell (1977) afin d'estimer plus spécifiquement la mortalité maternelle. Il s'agit de convertir la proportion de sœurs qui, ayant atteint l'âge adulte, sont décédées de cause maternelle. Cette proportion s'obtient à partir des quelques questions qui concernent donc uniquement les plus de 15 ans (*cf.* chap.1). Pour faire cette conversion entre proportions et probabilités, Graham et al. (1989) proposent de tenir compte de deux éléments : (1) de la distribution des différences d'âge entre une enquêtée et ses différentes sœurs, et (2) de la distribution par âge des risques de décès maternels.

Nous n'entrons pas ici dans les détails des calculs. Un descriptif plus complet de la méthode est disponible en annexe A.5.2 (p.439). Signalons uniquement que la méthode suppose que *la distribution des écarts d'âge entre l'enquêtée et ses sœurs ne varie pas selon l'âge des enquêtées, et qu'elle soit symétrique et de moyenne nulle* [66].

Par exemple, imaginons qu'une enquêtée soit âgée de 20 ans et qu'elle ait deux sœurs, l'une de 17 ans et l'autre de 25 ans. Du point de vue de l'enquêtée, la somme des différences d'âge vaudrait 2 (-3 + 5). Si l'on augmente le nombre d'individus sur lequel ce calcul est fait, cette

64. Pratiquement, ils font l'hypothèse qu'aucune naissance ne survient dans l'année qui suit ou qui précède celle de la naissance d'*Ego*, et ils supposent également que la fécondité est diminuée de moitié entre 1 et 2 ans avant et après cette naissance.

65. Notons que l'équation 5.22 n'est exacte que si $m(y)$ est "conditionnel" à la naissance d'un enfant à l'âge x, et il en va de même pour $m(x+u)$ dans l'équation 5.24 (Keyfitz et Caswell 2005). Or, les valeurs de $m(y)$ et $m(x+u)$ sont plus élevées quand elles sont *"conditionnelles" à cette naissance* que lorsqu'elles sont observées dans la population générale - notamment parce que les femmes infécondes sont éliminées quand elles sont conditionnelles. Mais ceci ne devrait pas avoir d'effet sur les estimations qui sont dégagées des coefficients.

66. Un ajustement est toutefois nécessaire pour les plus jeunes enquêtés - *cf.* annexe A.5.2.

somme des différences va tendre vers 0. Cela paraît assez intuitif : il y a en moyenne autant de petites sœurs que de grandes sœurs parmi les femmes enquêtées. D'où l'hypothèse de Graham et al. (1989) d'une distribution des écarts d'âge symétrique et de moyenne nulle [67]. Cette hypothèse revient en fait à considérer que l'âge des enquêtées est notre meilleure approximation pour l'âge de ses sœurs. Nous y reviendrons.

La méthode des sœurs est extrêmement simple à utiliser et elle recquiert peu de données. Par contre, il semble qu'elle procure des estimations peu fiables. A l'aide de simulations, Garenne et Friedberg (1997) ont pointé de sérieux problèmes : (1) les estimations indirectes s'avèrent systématiquement supérieures aux estimations directes, (2) elles s'accompagnent d'importantes marges d'erreurs [68], et (3) les probabilités de décéder de cause maternelle déduites des plus jeunes enquêtés semblent particulièrement biaisées. Garenne et Friedberg (1997) montrent que l'une des principales sources de biais provient de l'hypothèse faite par rapport aux différences d'âge. A l'aide de simulations, ils montrent que la distribution des écarts d'âge n'est en fait ni constante par âge, ni symétrique, ni de moyenne nulle.

En dépit de ces limites, un nombre considérable d'enquêtes centrées sur la mortalité maternelle ont posé les questions nécessaires à l'application de la méthode de Graham et al. (1989). C'est le cas d'enquêtes à petite échelle, mais également d'enquêtes nationales, comme les MICS (David et al. 1991, Walraven et al. 1994, Danel et al. 1996, Font et al. 2000, Lech et Zwane 2002, Mbaruku et al. 2003). Mais comme telle, la "méthode des sœurs" ne nous dit rien sur la mortalité générale. Et les coefficients de Hill et Trussell (1977) ne peuvent être utilisés lorsque les questions portent sur les frères et sœurs ayant atteint l'âge adulte.

Afin d'exploiter les données sur les frères et sœurs qui ne prennent pas la forme d'une histoire génésique complète, et pour fournir d'autres estimations pouvant être confrontées à celles calculées directement, Timaeus et al. (2001) ont donc développé une méthode spécifique pour les frères et sœurs adultes.

Probabilités de survie à l'âge adulte

Timaeus et al. (2001) commencent par mettre en question l'hypothèse faite par Graham et al. (1989) quant à la distribution des écarts entre l'âge de chaque individu et l'âge de ses frères et sœurs. Ils précisent que la moyenne de ces écarts d'âge n'est nulle que dans les populations stationnaires. Elle est négative lorsque le taux de croissance est positif, parce que

67. Lorsque les fratries sont complètes, c'est-à-dire que les mères des enquêtés ont atteint la fin de leur période reproductive.

68. "The sample variations in each age group were studied by computing linear regression coefficients between the direct estimate (independent variable) and the indirect estimates (dependent variable) among the 100 simulations. The variations were, in general, very large, so large that in one of the 14 regressions no correlation was found between the direct and the indirect estimates (age group 15-19), and two others had a low R^2 (age groups 25-29 and 35-39). In many other age groups, a marked slope indicated major interactions in addition to a systematic bias (age groups 20-24, 25-29, 30-34, 75-79, and 80-84).", Garenne et Friedberg (1997), p.136.

les enquêtés ont alors davantage de petits frères que de grands frères [69]. Elle est positive lorsque le taux de croissance est négatif. Les distributions des écarts d'âge présentent également des écarts-types variables, qui ne peuvent pas être fixés arbitrairement, comme l'ont fait Graham et al. (1989) [70].

Après avoir réinterrogé ces deux hypothèses et suggéré d'autres approches pour traiter de ces problèmes, Timaeus et al. (2001) procèdent par simulation pour suggérer de nouveaux coefficients (repris dans le tableau 5.13) [71]. Ces derniers s'appliquent comme précédemment à des proportions de survivants, mais celles-ci sont calculées à partir des frères et sœurs *qui ont survécu au moins jusqu'à leur quinzième anniversaire* ($_5S_n^{+15}$) [72] :

$$_{n-15}p_{15} = \beta_0(n) + \beta_1(n) \times_5 S_{n-5}^{+15} \tag{5.26}$$

n	β_0	β_1	R^2	CV
25	-0.0003	1.0011	0.9818	0.0042
30	-0.1546	1.1560	0.9950	0.0034
35	-0.1645	1.1660	0.9981	0.0029
40	-0.1388	1.1406	0.9984	0.0035
45	-0.1140	1.1168	0.9985	0.0042
50	-0.1018	1.1066	0.9986	0.0052

TABLE 5.13 – *Coefficients utilisés pour convertir les proportions de frères/sœurs survivant(e)s (parmi ceux qui ont survécu à 15 ans) en probabilités de survie - Timaeus et al. (2001)*

On observe dans le tableau 5.13 que la constante β_0 et le coefficient β_1 ont tendance à s'annuler mutuellement, ce qui traduit le fait que les ajustements à faire pour passer des proportions $_5S_{n-5}^{+15}$ aux probabilités $_{n-15}p_{15}$ sont assez faibles. La méthode est donc *peu sensible aux variations des schémas de fécondité et de mortalité*. Ceci signifie que ces coefficients

69. Prenons des individus du même âge. Plus les mères de ces individus sont jeunes, plus il s'agit d'aînés (plutôt que de cadets). En effet, comme leurs mères sont jeunes, leur période de reproduction est plus étendue *après* la naissance de ces individus qu'avant. Par contre, plus leurs mères sont âgées, plus il s'agit de cadets, puisque leurs mères ont déjà derrière elles plusieurs années de période de reproduction. Or, dans une population stable, pour ce même ensemble d'individus, *plus le taux de croissance est positif, plus il y aura de jeunes mères*. Dès lors, plus le taux de croissance est élevé, plus le rapport du nombre de cadets sur les aînés est élevé. Une série de travaux ont d'ailleurs été consacrés au problème de l'estimation du taux de croissance à partir du rapport entre le nombre moyen de plus jeunes sœurs et le nombre moyen de sœurs plus âgées (Goldman 1978, Wachter 1980, McDaniel et Hammel 1984).

70. En fait, la distribution des écarts d'âge ne peut pas être établie à partir de la connaissance du *calendrier de la fécondité* mais à partir de la *distribution des écarts entre les naissances*. Comme le notent Timaeus et al. (2001), si l'âge à la première naissance varie beaucoup mais que les premières naissances se suivent à un rythme rapide, alors la variance de la distribution des naissances sera plus faible que la variance du calendrier de la fécondité.

71. La mortalité est modélisée au moyen du standard de Brass, et du système des logits. La fécondité est appréhendée par le biais du standard de Booth (1984) et du modèle relationel présenté dans le chapitre 3.

72. Les données fournies par les moins de 20 ans ne sont pas utilisées, car les frères et sœurs de ces enquêtés sont en moyenne plus âgés que les enquêtés eux-mêmes, ce qui rend la méthode assez sensible aux choix faits pour modéliser la distribution des écarts d'âge entre un enquêté et ses frères et sœurs. Ce n'est plus le cas chez les plus de 20 ans.

devraient également procurer des estimations fiables dans les contextes perturbés par le VIH-sida. C'est ce que montrent Timaeus et al. (2001) à partir de simulations : les erreurs introduites sur les quotients ${}_{35}q_{15}$ sont négligeables chez des enquêtés âgés de 20-24 ans ainsi que chez les plus de 40 ans. La mortalité est par contre sous-estimée chez les 30-39 ans, mais de quelques pourcents seulement. En outre, à la différence de la méthode des orphelins, les problèmes posés par la transmission verticale n'affectent pas les estimations basées sur les frères et sœurs.

Afin de pouvoir dater les estimations, Timaeus et al. (2001) ont proposé des valeurs γ_0 et γ_1 (correspondant à chaque âge n) qui permettent d'estimer la durée écoulée entre la période de référence de l'estimation et l'enquête[73] :

$$T = \gamma_0(n) + \gamma_1(n) \times ln({}_5S_{n-5}^{+15}) \tag{5.27}$$

Un exemple de calcul est présenté dans le tableau 5.14, à partir de l'enquête EDS menée au Congo RDC en 2007.

Age	Surv.	Décédés	${}_5S_{n-5}$	n	β_0	β_1	${}_{n-15}p_{15}$	T
20-24	4236	215	0.952	25	-0.0003	1.0011	0.952	3.29
25-29	3498	251	0.933	30	-0.1546	1.1560	0.924	5.59
30-34	3036	319	0.905	35	-0.1645	1.1660	0.891	7.80
35-39	2489	249	0.909	40	-0.1388	1.1406	0.898	9.73
40-44	1971	347	0.850	45	-0.1140	1.1168	0.835	11.72
45-49	1399	295	0.826	50	-0.1018	1.1066	0.812	13.33

TABLE 5.14 – *Exemple de calcul indirect à partir de l'enquête Congo RDC 2007 - Calcul sur les **frères** ayant survécu à leur 15e anniversaire - Coefficients de Timaeus et al. (2001)*

Le problème du choix du standard

Nous avions noté dans le chapitre 3 que l'une des difficultés de la méthode des orphelins était qu'après avoir estimé des probabilités de survie de type ${}_np_{25}$, il fallait pouvoir *les convertir dans un indice commun* (du type ${}_{45}q_{15}$). Et pour faire cette conversion, le choix du standard est difficile à justifier quand on ignore comment les risques de décès se distribuent par âge. Ce choix a peu d'impact quand on s'en tient à des schémas "classiques" (tables de Princeton, etc.), mais il est par contre déterminant lorsqu'on utilise des standards perturbés par le VIH-sida, comme c'est le cas du standard INDEPTH-2. Pour rappel, dans le cas de la méthode des orphelins, ce standard va procurer des niveaux de mortalité *plus faibles* pour les valeurs de ${}_np_{25}$ issues des premiers groupes d'âge, et *plus élevés* pour les estimations de la probabilité ${}_np_{25}$ issue des enquêtés plus âgés ($n > 30$) (*cf.* page 121)[74].

Les estimations indirectes déduites des frères et sœurs doivent elles aussi être converties en indices de type ${}_{45}q_{15}$. Mais les probabilités de

73. Ces coefficients sont repris en annexe A.5.3

74. Dans ce cas, comme les plus jeunes enquêtés procurent les estimations les plus récentes, convertir les ${}_np_{25}$ au moyen de schémas par âge classiques va conduire à une surestimation de la hausse de mortalité.

survie se calculent sur des intervalles d'âge différents de celles des orphelins, et il s'avère en fait que l'*on va observer la situation inverse de celle des orphelins*. En effet, pour la mortalité féminine, si l'on retient le standard INDEPTH-2, les niveaux de mortalité déduits des plus jeunes enquêtés seront *supérieurs* à ceux qui auraient été obtenus à partir d'un schéma de mortalité plus classique. Pour des valeurs de n égales à 30, les deux schémas procurent globalement les mêmes niveaux, et au-delà, le schéma perturbé par le VIH va fournir des niveaux de mortalité féminine *inférieurs*. Ceci s'explique par le fait que l'excès de mortalité féminine dû au VIH est concentré entre 20 et 35 ans. On peut observer ces différentes relations entre les probabilités ${}_nq_{15}$ et ${}_{45}q_{15}$ sur le graphe 3.8 (voir p.121 pour le même schéma pour les orphelins). Notons que comme les plus jeunes enquêtés procurent les estimations les plus récentes, convertir les ${}_np_{15}$ au moyen de schémas par âge classiques va nous conduire à sous-estimer l'ampleur de la hausse de mortalité.

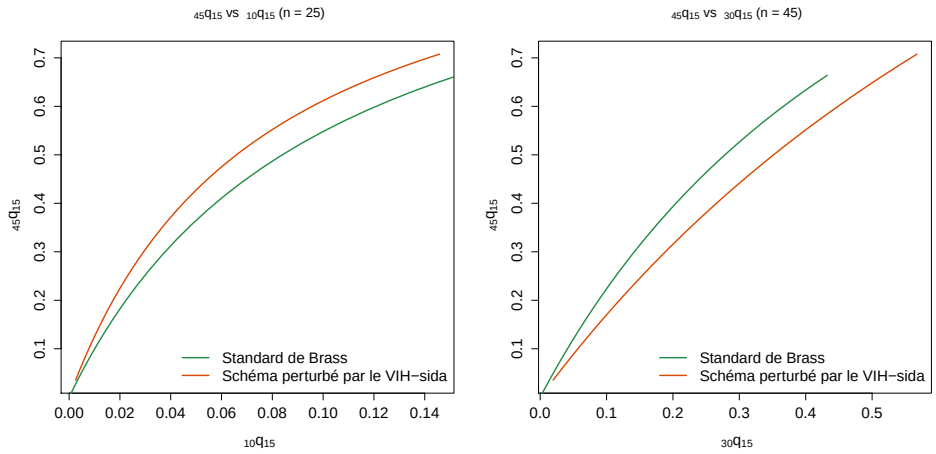

FIGURE 5.13 – *Relation entre les probabilités ${}_{10}q_{15}$, ${}_{30}q_{15}$ et ${}_{45}q_{15}$ selon le type de schéma de mortalité - Schémas INDEPTH-2 (INDEPTH 2004) et standard de Brass*

5.3.3 Application aux données d'enquêtes

Nous avons appliqué cette méthode indirecte basée sur la survie des frères et sœurs adultes aux enquêtes EDS disposant du module sur la mortalité maternelle. Nous convertissons ensuite les quotients de mortalité obtenus (${}_{n-15}q_{15}$) en quotients de mortalité de 15 à 60 ans, par le biais du standard général de Brass.

Nous isolons ici les cas rencontrés au Sénégal, à Madagascar, au Kenya et au Bénin (graphe 5.14).

1. On note un bon degré d'accord entre les deux séries d'estimations au Sénégal et au Bénin, là où la mortalité féminine est restée relativement stable sur la période considérée. Ceci nous permet de penser que les âges des frères et sœurs n'ont pas été systématiquement *surestimés*. Ceci se traduirait en effet par des niveaux "indirects" plus élevés que des niveaux "directs".

2. Par contre, à Madagascar, les estimations indirectes présentent une succession de tendances à la hausse incompatibles avec la baisse de la mortalité déduite des estimations directes. Dans ce cas, on peut supposer que les estimations fournies *par les enquêtés plus âgées* sont biaisées à la baisse. Ceci se traduit par une fausse impression de hausse de la mortalité lorsque celle-ci est calculée indirectement[75].
3. Dans le cas du Kenya, les divergences entre les deux sources d'estimation ont sans doute une autre explication. Les niveaux calculés indirectement font intervenir des décès qui sont survenus sur une longue période de temps, et l'approche indirecte n'est pas en mesure de saisir des changements rapides de mortalité. La datation des estimations pose alors problème. On retrouve ici une limite partagée avec la méthode des orphelins.

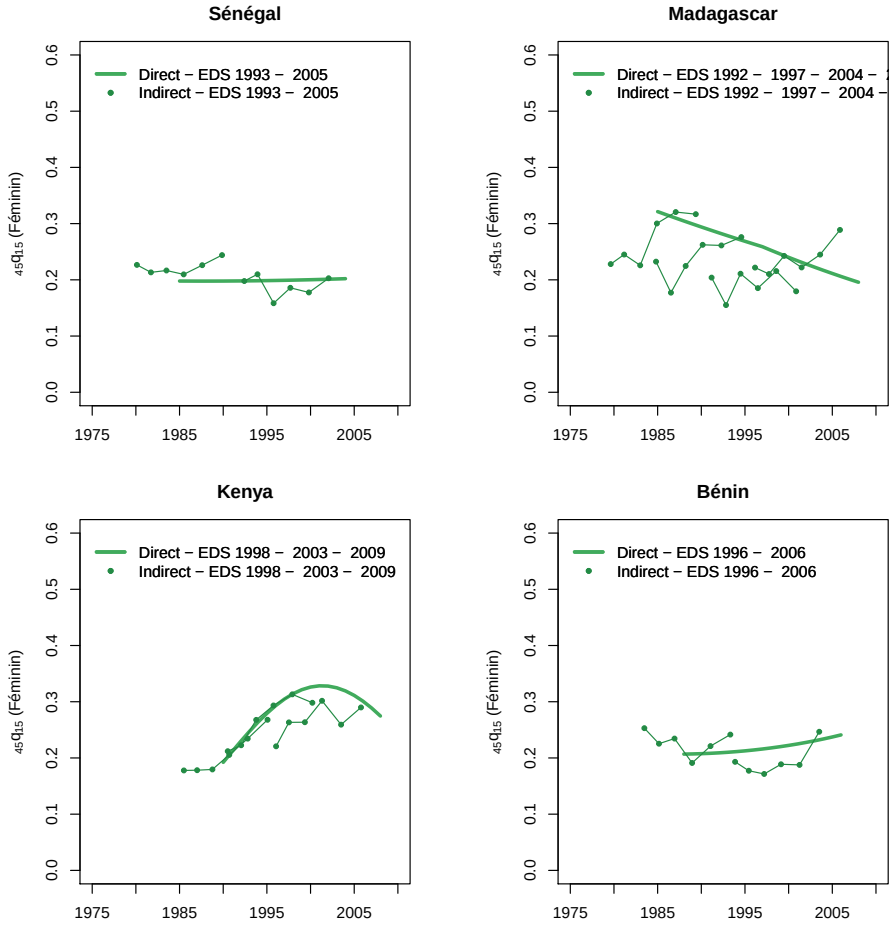

FIGURE 5.14 – *Quotients de mortalité féminins de 15 à 60 ans calculés à partir de la survie des frères et sœurs de façon biographique (à partir du modèle de Timaeus et Jasseh (2004) et de façon indirecte (à partir des coefficients de Timaeus et al. (2001)) - Enquêtes EDS menées au Sénégal, à Madagascar, au Kenya et au Bénin.*

Le chapitre 7 présentera en détail les différentes estimations nationales.

75. Cette sous-déclaration affecte moins les estimations directes, car les informations fournies par les différents groupes d'âge sont alors combinées entre elles dans les mêmes personnes-périodes.

Avant cela, le chapitre suivant est consacré à un examen de la qualité des données sur la survie des proches.

Qualité des données, principaux biais introduits et effets sur les estimations

6

Sommaire

6.1 Données portant sur les parents 278
6.2 Données portant sur la fratrie 290
Conclusion 308

Les données sur à la survie des proches, relativement nombreuses en Afrique subsaharienne, constituent une source incontournable pour estimer la mortalité adulte et en reconstituer les tendances. Ces données restent toutefois sous-utilisées, notamment en raison d'une certaine prudence face à la qualité variable des informations recueillies. Dans les pays en développement, les informations fournies par les individus *sur leurs propres caractéristiques démographiques* sont souvent déjà elles-mêmes entachées d'erreurs : erreurs sur les âges, sur les parités, sur les relations de parenté, etc. Or on peut imaginer que les erreurs d'âge, les omissions et les déclarations approximatives sont encore plus fréquentes *lorsqu'elles concernent un parent ou un frère*.

Il nous paraît important de pouvoir suggérer un ordre de grandeur des erreurs qui sont faites, d'identifier les catégories de répondants où ces erreurs sont les plus prononcées, et de tenter d'en estimer les effets sur les niveaux de mortalité obtenus.

Plus spécifiquement, ce chapitre concerne,

1. pour les données sur la survie des parents :
   - l'"effet d'adoption"
   - la présence de données manquantes
   - les erreurs sur l'âge des enquêtés
2. pour les données sur les frères et sœurs :
   - la présence de données manquantes
   - les erreurs sur les âges des frères et sœurs (âge à l'enquête ou âge au décès)
   - la sous-déclaration des frères et sœurs ainsi que de leurs décès.

6.1 Données portant sur les parents

En ce qui concerne la survie des parents, c'est surtout le *biais d'adoption* qui a retenu l'attention des chercheurs. Dès les premières applications de la méthode des orphelins, il est apparu que les niveaux de mortalité obtenus à partir des *plus jeunes enfants* étaient trop bas, et que les tendances indiquaient des *baisses de mortalité non plausibles*. Cela peut s'expliquer en partie par ce biais d'adoption (*cf.* chapitre 3).

6.1.1 Biais d'adoption

Le terme "biais d'adoption" renvoie au fait que l'information fournie sur les parents se réfère *à d'autres personnes que les parents biologiques*. Si la survie de ces personnes (souvent appelées "parents adoptifs") diffère de celle des parents biologiques, un biais est possible, et il conduit le plus souvent à sous-estimer la mortalité.

Cette erreur de déclaration peut provenir (1) des enquêtés eux-mêmes, (2) des parents adoptifs ou (3) des enquêteurs. Plusieurs éléments permettent d'expliquer la fréquence de cette erreur en Afrique subsaharienne :

1. Tout d'abord, les *termes associés à la filiation* (père, mère, etc.) sont rarement réservés aux parents biologiques dans les sociétés africaines, où ils sont souvent utilisés comme une marque de respect. Dans certains cas, la terminologie des liens de parenté ne comprend même pas de terme voulant spécifiquement dire "père" ou "mère" : Diop (1985) note qu'en *wolof*, par exemple, le terme "*Baay*" s'étend au père, aux frères, aux cousins, et par extension, à tous les hommes de la même génération [1]. Des erreurs peuvent donc être causées par une compréhension trop large des questions sur la survie des parents. Sur les questionnaires, il n'est malheureusement pas toujours fait mention du fait que ces questions concernent les parents biologiques uniquement [2].
2. Les enquêteurs peuvent également considérer les adultes qui cohabitent avec des enfants comme les parents biologiques de ces derniers, et *s'abstenir de poser les questions* sur la parenté réelle pour gagner un peu de temps. Cela va surtout affecter les données relatives aux plus jeunes enfants, comme le précise Blacker (1984) :

> "There is no doubt that the group of respondents most susceptible to the adoption effect are young children enume-

1. A l'exception de l'oncle maternel.

2. Lors du recensement sud-africain de 2001, la question était explicite : "Is (the person's) own biological father still alive ?". Mais ce n'est pas souvent le cas, et il faut plutôt se référer au manuel des agents recenseurs. Les instructions du recensement kenyan de 1999 sont par exemple très claires : "Tick the box under the appropriate column in respect of the survival status of the respondent's biological father and mother. Note that at times destitute children are brought up or adopted at a very young age by relatives. Such foster parents should not be considered as the biological parents of the respondent. Please always probe to establish the reality of the situation." Au Sénégal en 1988, le manuel des agents recenseurs précise également : "Assurez-vous qu'il s'agit du vrai père de l'enquêté et non de l'oncle ou du tuteur qui a élevé la personne."

> rated in the same household as adults who are shown, in answer to the question on relationship, as being the parents of the children ; in these circumstances, enumerators rarely bother to probe whether the children are the true offspring of these so-called "parents" and automatically record the fathers and mothers as being alive without asking the questions." (Blacker, 1984, p.85)

3. La *pratique du "confiage"* peut également être à l'origine du biais d'adoption. A cet égard, notons que le terme "adoption" est en fait mal adapté à la situation africaine, où l'adoption au sens occidental du terme est très rare, et où la pratique du "confiage" des enfants est par contre très fréquente. Cette pratique a fait l'objet de nombreux travaux anthropologiques et démographiques (Goody 1982, Isiugo-Abanihe 1985, Vandermeersch 2002, Zimmerman 2003, Ainsworth 1996).

   A la différence de l'adoption, le confiage est pratiquement toujours *temporaire* et *réversible*. Il ne modifie pas définitivement la relation qui unit les parents biologiques aux enfants. Les enfants sont placés dans le réseau familial (ou amical) des parents, notamment pour poursuivre leur scolarité, accéder à un apprentissage, renforcer les liens familiaux, participer aux tâches domestiques ou encore apporter un soutien aux personnes âgées ou aux femmes sans enfants [3].

   De la littérature consacrée au confiage, il ressort notamment qu'il s'agit d'une pratique surtout importante en Afrique de l'Ouest (Isiugo-Abanihe 1985), et plus fréquente pour les 10-14 ans que pour les enfants plus jeunes. A mesure que les enfants avancent en âge, les garçons semblent également plus fréquemment confiés que les filles (Renata 2009). Ces travaux indiquent par ailleurs que *le décès d'un conjoint* est un motif fréquent de confiage (Vandermeersch 2002). Dans ce cas, le confiage va introduire un *biais à la baisse* dans les estimations de mortalité, si les enfants confiés sont ensuite déclarés comme enfants biologiques. Dans les autres cas, l'ampleur du biais va dépendre de *la différence entre les risques de mortalité des parents qui confient les enfants et les risques de décès de ceux qui les accueillent*. L'ampleur des erreurs va donc varier avec les formes que prend le confiage. Par exemple, Antoine et Bocquier (1992) relèvent que les règles qui gouvernent l'allocation des enfants varient avec l'âge de ceux-ci :

   > "...on voit clairement que suivant l'âge de l'enfant, ce ne sont pas les mêmes personnes qui sont sollicitées. Au jeune âge l'enfant est confié plutôt à des aînés, parents de la mère. Par contre, après 10 ans, il est confié à des hommes, parents ou non, aptes à prendre en charge, sous une forme ou sous une autre, sa formation et son éducation." (Antoine et Bocquier, 1992, p.17)

3. Renata (2009) note à propos du confiage :"Fostering is ubiquitous, especially in West Africa, and represents the most visible aspect of the wider phenomenon of extended family involvement in all aspects of child rearing. The borderline between fostering proper and these other exchanges may not be obvious, making empirical research difficult and very context-dependent." (p.159)

4. Le quatrième élément pouvant expliquer l'importance des erreurs de déclaration est la *faible fréquence de la cohabitation des parents avec leurs enfants*. Ceci concerne bien sûr les enfants confiés, mais aussi de nombreux autres enfants, qui vivent avec leur mère sans vivre avec leur père, ou vice versa, ou qui ne vivent avec aucun des deux. A titre illustratif, nous avons calculé à partir des EDS, les proportions d'enfants qui ne vivent pas avec leur mère (parmi les non-orphelins de mère).

   Le graphe 6.1.A présente la distribution des enquêtes selon cette proportion, calculée sur les enfants de 5 à 9 ans. La même distribution est présentée pour les pères dans le graphe 6.1.B., parmi les enfants dont le père survit.

   En moyenne, *sur l'ensemble des enquêtes, 20% des enfants non-orphelins de mères âgés de 5 à 9 ans ne cohabitent pas avec leur mère*. Cette proportion s'élève à 30% pour les pères chez les enfants du même âge. Chez les 10-14 ans, ces proportions atteignent respectivement 26 et 33%. L'Afrique australe est la région où les enfants de 10 à 14 ans cohabitent le moins avec leur mère (parmi les non-orphelins) : seuls 70% des enfants de cet âge vivent avec leur mère, contre 77% en Afrique de l'Est[4]. Cela pourrait en partie expliquer les nombreuses confusions en réponse aux questions sur la survie des parents.

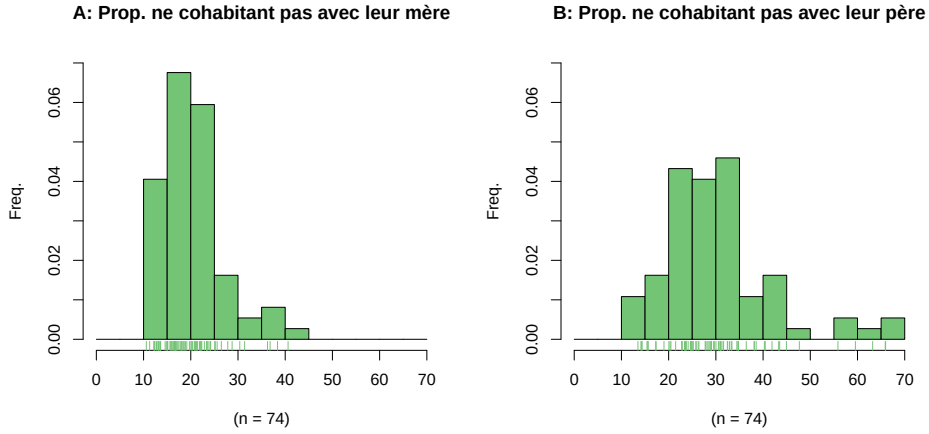

FIGURE 6.1 – *Proportions d'enfants de 5 à 9 ans dont la mère/le père est survivant(e) qui ne cohabitent pas avec leur mère/leur père - Enquêtes EDS africaines (n = 74)*

L'effet de ce biais d'adoption sur les estimations va varier *avec le niveau de mortalité* (Timaeus 1986). Par exemple, si 30% des mères décèdent et que 5% des orphelins sont déclarés comme non-orphelins, la proportion de non-orphelins est sur-estimée de 2.14%. Par contre, si 70% des mères décèdent et que *la même proportion d'orphelins est déclarée comme non-orphelins*, la proportion de non-orphelins est surestimée de 11.6%. Toutes choses égales par ailleurs, plus la mortalité est élevée, plus le biais est prononcé.

Palloni et al. (1984) ont proposé une modélisation de ce biais d'adoption, que nous ne présenterons pas ici en détail. Même s'il n'existe pas de

4. Pour les pères, les proportions d'enfants de 10 à 14 ans qui cohabitent avec leur père varient de 57% en Afrique australe à 71% en Afrique de l'Ouest.

données permettant de paramétriser précisément leur modèle pour en déduire des facteurs d'ajustement [5], ces auteurs aboutissent à des ordres de grandeur. Ils estiment que pour un niveau de mortalité élevé ($e_0 = 45$), le biais d'adoption conduit à une sur-estimation de la probabilité de survie ${}_np_{25}$ variant de 9% chez les 5-9 ans à 7% chez les 45-49 ans [6].

Une série de recherches viennent compléter leurs travaux en estimant l'ampleur du biais d'adoption à partir des données d'enquêtes :

1. Une première approche consiste à *comparer les proportions d'orphelins empiriques aux proportions issues de la modélisation*. C'est ce que font Grassly et al. (2004) à partir des enquêtes MICS et EDS, en les confrontant avec celles déduites des projections de population des Nations Unies (*cf.* chapitre 3, p. 176). Ils constatent que les proportions issues des modèles sont significativement *plus élevées* que celles dégagées de l'observation, ce qu'ils expliquent en partie par le biais d'adoption. Ils notent que les proportions de pères cadrent mieux avec les estimations issues des modèles que ce n'est le cas chez les mères. Dans le chapitre 3, nous avons également pu mettre en évidence un décalage important entre les proportions d'orphelins issues de nos micro-simulations et celles issues des enquêtes. La médiane du rapport entre les premières et les secondes se chiffre à 1.45 pour les enfants de 0 à 4 ans, à 1.3 pour les 5-9 ans et à 1.42 pour les 10-14 ans [7].

2. Une deuxième série de travaux ont cherché à estimer l'effet d'adoption *à partir d'opérations de collecte répétées*. Pison et Langaney (1988) ont ainsi analysé la cohérence des déclarations sur les orphelins auprès des Fula Bandé du Sénégal, quelque temps après le recensement de 1975. Ils ont constaté que *près d'un tiers des enfants de 10 à 14 ans qui étaient orphelins de pères avaient en fait été déclarés comme ayant toujours un père survivant*. Plus récemment, Robertson et al. (2008) se sont appuyés sur des données longitudinales issues de l'observatoire de Manicaland (au Zimbabwe) pour documenter cet effet d'adoption. Ils ont analysé la cohérence des déclarations faites sur la survie des parents entre 3 rounds successifs. Au premier round, 198 enfants de moins de 16 ans ont été déclarés comme orphelins de mère. *33% (27-40%) d'entre eux ont ensuite été déclarés comme non-orphelins lors d'au moins un des deux rounds qui ont suivi.* La sous-déclaration des orphelins semble donc massive, et elle s'avère plus prononcée chez les jeunes enfants, ainsi que chez les enfants vivant dans des zones moins favorisées. La proportion d'enfants

5. Il faudrait pour cela disposer d'informations sur l'âge des femmes qui accueillent les enfants, sur les différences de mortalité des mères adoptives et biologiques, sur la durée écoulée entre le décès des parents et l'éventuelle adoption, sur la capacité de l'enfant adopté à distinguer ses nouveaux parents d'adoption de ses véritables parents biologiques, etc.

6. Comme l'ampleur des erreurs semble diminuer légèrement avec l'âge des enquêtés, le biais peut avoir pour effet d'induire une fausse impression de baisse de la mortalité.

7. Blacker (1984) note également que le biais peut être estimé en comparant le nombre de non-orphelins et le nombre d'enfants survivants déclarés par les mères. En l'absence de migration importante, ces deux nombres devraient être proches l'un de l'autre.

déclarés orphelins de père au premier round puis non-orphelins dans les rounds suivants s'élèvait à 13% (11-16%), sur 689 enfants. Une fois encore, le biais d'adoption apparaît donc moins élevé chez les pères. Dans le cas des pères, Robertson et al. (2008) estiment que les incohérences sont davantage associées à une *incertitude sur l'identité ou la survie de ceux-ci*, tandis que les incohérences observées à propos des mères sont davantage associées à un *effet d'adoption* plus systématique. Ceci pourrait notamment s'expliquer par le fait que la prise en charge des orphelins varie selon que ce soit le père ou la mère qui survit. Les mères veuves sont beauvoup plus souvent en charge des orphelins de pères, tandis que les orphelins de mères sont moins susceptibles de résider avec leur père survivant (UNICEF 2006).

3. Rutstein (2008) utilise une troisième approche, en effectuant des *comparaisons internes* au sein des données EDS pour estimer l'ampleur du biais d'adoption. Les informations collectées dans le feuillet ménage sur la survie des parents ont été confrontées à d'autres variables disponibles ; (1) la relation entretenue avec le chef de ménage, (2) les histoires génésiques collectées dans l'enquête individuelle, et (3) les données relatives aux sœurs des femmes enquêtées et à leurs enfants. Cette approche révèle une importante sous-déclaration des orphelins dans le feuillet ménage [8].

En définitive, le biais d'adoption représente un problème majeur pour la méthode des orphelins. Cela est apparu dès les premières applications de la méthode sur les données africaines (Hill 1984, Blacker et Mukiza-Gapere 1988, Timaeus 1986; 1991a;b). Plusieurs auteurs ont rapidement relevé l'intérêt du recours aux cohortes synthétiques pour atténuer les problèmes posé par ce biais d'adoption (*cf.* chap.3). Bien qu'il soit difficile d'en estimer l'ampleur avec précision, ce biais est tel que les estimations obtenues ne peuvent souvent que servir de "bornes inférieures". *Des niveaux de mortalité qui seraient en-deça de ceux dégagés des orphelins devraient être considérés avec beaucoup de prudence.* C'est déjà une information en soi.

6.1.2 Données manquantes sur la survie des parents

Globalement, les proportions de réponses manquantes sur la survie des parents restent faibles, dans les enquêtes EDS comme dans les recensements. Toutefois, comme elles sont souvent du *même ordre de grandeur* que les proportions d'orphelins, le traitement de ces réponses manquantes peut avoir un impact important sur les estimations de mortalité.

8. Mais les hypothèses de Rutstein (2008) nous paraissent assez larges. Par exemple, tous les enfants qui sont référés à une mère dans le feuillet ménage sans être repris par celle-ci dans sa propre biographie génésique sont considérés comme de potentiels orphelins non-déclarés.

Enquêtes EDS

Sur les 75 enquêtes EDS disposant d'informations sur la survie des parents, les réponses manquantes[9] représentent en moyenne 0.76% des informations fournies sur la *survie des mères*, pour les enfants de moins de 15 ans. Cette proportion varie dans la plupart des enquêtes de 0.2 à 1.53% (il s'agit là des 10e et 90e percentiles). Le maximum est atteint avec l'enquête menée au Nigeria en 1999 (4.57%), une enquête connue pour sa moindre qualité [10].

Pour les *pères*, la complétude des informations est légèrement plus faible (pères inconnus, contacts perdus). Les proportions de réponses manquantes s'élèvent à 1.46% en moyenne. Les 10e et 90e percentiles des proportions de réponses manquantes sur la survie des pères se chiffrent respectivement à 0.5 et 3.11% [11]. *Ces réponses manquantes sont majoritairement concentrées chez des enfants pour lesquels il manque à la fois des données sur la mère et sur le père.* Cette situation concerne en moyenne 0.66% des enfants.

Les proportions de réponses manquantes que nous avons calculées par enquête sont reprises dans le tableau A.19 (en annexe A.6.1). Le graphe 6.2 présente ces proportions sous forme de box-plots. On observe une hausse de la fréquence des réponses manquantes avec l'âge des enfants. On note également une plus grande variation entre enquêtes lorsque l'on examine la complétude des données sur les pères.

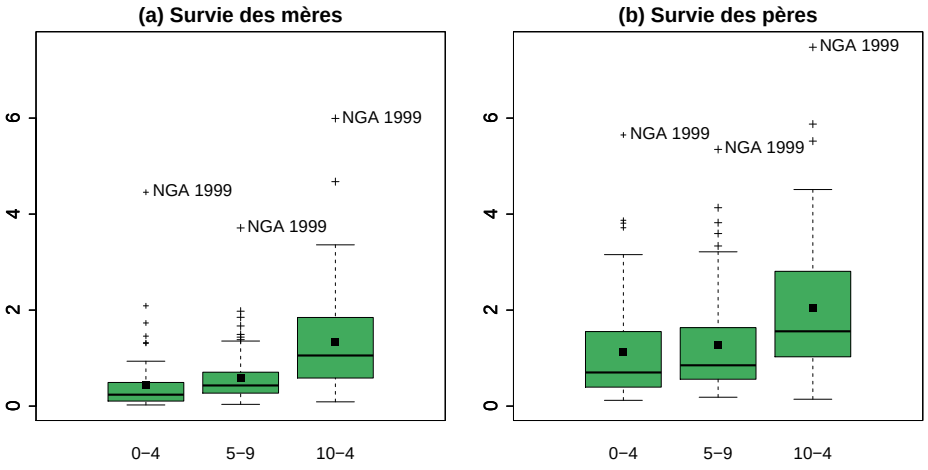

FIGURE 6.2 – *Proportions de réponses manquantes (par enquête et en %) sur la survie des mères (a), et la survie des pères (b), enquêtes EDS africaines (n = 75)*
LECTURE : Pour chaque âge, une barre de couleur comprend la moitié des enquêtes. Un quart des enquêtes présentent des proportions de réponses manquantes plus élevées que la limite supérieure de la barre de couleur (c-à-d. le dernier quartile), un autre quart présente des valeurs plus faibles que la limitée inférieure de cette barre. La médiane est notée par un trait noir, la moyenne par un carré. Les droites en pointillés s'étendent jusqu'aux points qui ne sont pas séparés de la boîte de plus de 1.5 fois l'intervalle interquartile.

9. Nous regroupons sous ce terme la modalité "ne sait pas" et l'absence de réponse.
10. *Cf.* note 11, p.231
11. C'est à nouveau l'enquête Nigeria 1999 qui présente la proportion de réponses manquantes la plus élevée (6%).

Monasch et Boerma (2004) ont mené une analyse comparable. Ils indiquent que les enquêtes menées en Afrique australe présentent en moyenne les proportions les plus élevées de réponses manquantes (pour au moins l'un des parents). Ils associent cela au fait que cette sous-région est également celle où *les enfants co-habitent le moins avec leurs parents lorsque ceux-ci survivent* [12]. Cette hypothèse est intéressante, mais elle ne convainc pas tout à fait :

1. D'une part, l'Afrique australe se caractérise effectivement par des réponses manquantes plus fréquentes pour les pères, mais de peu [13]. La survie des mères n'y est par contre pas déclarée de façon moins complète : l'Afrique centrale et l'Afrique de l'Ouest présentent même des proportions de réponses manquantes légèrement plus élevées.
2. D'autre part, les informations recueillies sur la survie des parents ne semblent pas plus complètes dans les pays où les *enfants co-habitent davantage avec leurs parents*.

Pour le vérifier, nous avons croisé, dans 75 enquêtes EDS africaines, les *proportions de réponses manquantes sur la survie des mères* avec la *proportion d'enfants qui co-habitent avec leur mère lorsque celle-ci survit*. Comme on l'observe sur le graphe 6.3.a (en haut à gauche), il n'y a pas de relation claire entre ces deux séries de proportions [14]. Il en va de même pour les pères : les réponses manquantes sur leur survie ne sont pas plus fréquentes dans les pays où les enfants co-habitent moins avec leur père (lorsque celui-ci survit). Le graphe en bas à gauche (figure 6.3.c) indique qu'*il n'y a pas non plus de relation nette entre, d'une part, la proportion de non-orphelins qui co-habitent avec leur mère et, d'autre part, la proportion d'orphelins de mère*. A nouveau, on peut faire le même constat pour les pères. En conclusion, le fait de ne pas cohabiter avec ses enfants (qu'il s'agisse de confiage des enfants ou d'autres formes d'organisations familiales) ne semble pas associé à une moindre qualité des données ni à une tendance à la sur- ou sous-estimation de la proportion d'orphelins.

12. "As children are least likely to live with their parent(s) in this sub-region, these higher levels of unknowns might be because absent parent(s) have been out of touch with their children, and current caretakers of the children are not sure about their survival status." - Monasch et Boerma (2004), p.S62

13. En retenant le découpage régional UNECA, les proportions de réponses manquantes sur la survie des pères se chiffrent à 1.92% en Afrique australe, 1.71% en Afrique de l'Est, 1.5% en Afrique Centrale et 1.2% en Afrique de l'Ouest.

14. L'enquête rwandaise de 1992 n'est pas retenue pour ce graphe, car les données sur la co-habitation des enfants avec les parents n'ont pas été collectées.

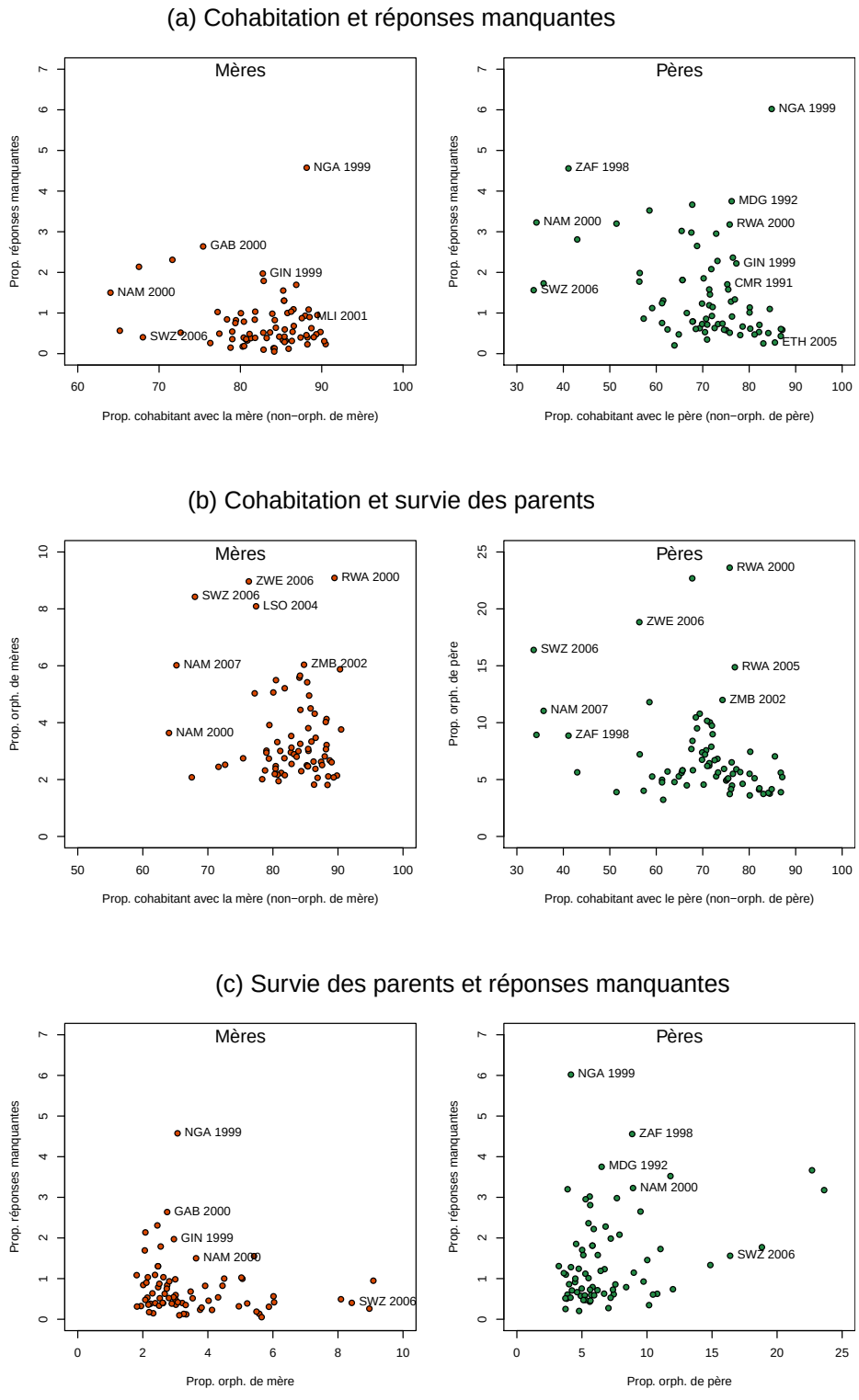

FIGURE 6.3 – *Enquêtes EDS (n= 75) :*
(a) Proportions de non-orphelins cohabitant avec leur mère(père) et proportions de réponses manquantes sur la survie des mères(pères),
(b) Proportions de non-orphelins de mère(père) cohabitant avec leur mère(père) et proportions de non-orphelins de mère(père),
(c) Proportions de non-orphelins de mère(père) cohabitant avec leur mère(père) et proportions de réponses manquantes sur la survie des mères(pères)

Recensements

La complétude des informations recueillies dans les recensements sur la survie des parents est comparable à celle des données d'enquête. Le tableau 6.1 détaille les proportions de réponses manquantes sur la survie des *mères* pour une dizaine de recensements africains (à partir des fichiers IPUMS analysés dans le chapitre 3) [15]. A l'exception du recensement rwandais de 2002, où les données manquantes se chiffrent à 2.7% chez les moins de 15 ans, la survie des mères est enregistrée de façon aussi complète que dans les enquêtes (en tous cas chez les moins de 15 ans, pour lesquels la moyenne des proportions de réponses manquantes se chiffre à 0.68%). Comme dans les enquêtes, l'importance des données manquantes augmente légèrement avec l'âge.

	Pays	Année	0-15	15-35	35-50
1	Mali	1987	0.23	0.34	0.46
2	Sénégal	1988	0.04	0.46	0.67
3	Tanzanie	1988	0.09	0.46	0.84
4	Kenya	1989	0.53	1.32	1.25
5	Rwanda	1991	0.26	0.43	0.44
6	Ouganda	1991	0.03	0.07	0.19
7	Afrique du Sud	1996	1.15	1.42	1.85
8	Mali	1998	0.71	0.93	0.89
9	Kenya	1999	1.57	1.12	1.00
10	Rwanda	2002	2.69		
11	Afrique du Sud	2007	0.17	0.21	0.21

TABLE 6.1 – *Proportions de réponses manquantes sur la survie des mères dans quelques recensements africains (en %) - 0-15 ans, 15-35 ans et 35-50 ans*

Pour quel type d'imputation opter ?

Il n'y a pas actuellement de méthode standard pour traiter de ces réponses manquantes, mais plutôt une certaine diversité d'approches :

1. Certains recourent à des solutions assez "brutes" : considérer par exemple les enfants pour lesquels la survie des mères est inconnue comme étant non-orphelins (Rutstein 2008). Une autre option consiste à reclasser toutes les réponses manquantes comme se réfé-

15. Les individus qui ne sont pas éligibles n'interviennent pas dans les calculs. C'est le cas des personnes en institutions pour le recensement sud-africain de 1996, par exemple. Nous ne présentons pas les résultats pour le recensement sud-africain de 2001, ni pour les recensements ougandais et tanzaniens de 2002 car aucune observation n'a de données manquantes. Comme il est clair que "ne sait pas" est une réponse valide et qu'une proportion de répondants, même minime, doit avoir opté pour cette modalité, cela signifie que ces fichiers de données ont déjà fait l'objet d'imputations. Cela est confirmé par la lecture de Dorrington et al. (2004) à propos du recensement sud-africain de 2001. Ces auteurs donnent plusieurs raisons pour lesquelles les données devraient être utilisées *sans les corrections* qui leur ont été apportées par Statistics South Africa pour éliminer les réponses manquantes. Le rapport final du recensement ougandais de 2002 mentionne également 0.2% de réponses manquantes sur la survie des parents, ce qui n'apparaît plus dans les données IPUMS.

rant à des orphelins. Beegle et al. (2010) optent pour une troisième approche : ils ignorent simplement les réponses manquantes.

2. D'autres travaux déterminent *a priori* comment reclasser les réponses manquantes. Ainsi, Case et al. (2004) analysent par exemple l'impact du décès d'un parent sur la scolarisation des orphelins (à partir d'une vingtaine d'enquêtes EDS). Ils recodent comme doubles orphelins les enfants pour lesquels l'information sur la survie manque pour *les deux parents*. Les enfants pour lesquels l'un des deux parents survit mais où la réponse est manquante pour l'autre parent sont exclus des analyses [16]. Bicego et al. (2003) procèdent de la même façon, mais en se basant sur la survie de *l'autre* parent. Par exemple, pour les données manquantes sur la survie des mères, la *moitié* des enfants orphelins de pères sont reclassés comme double orphelins, et un *tiers* des enfants dont le père survit sont recodés comme orphelins de mère. [17]

L'approche de Bicego et al. (2003) nous paraît être la plus rigoureuse. Elle tient compte du fait que la probabilité d'être orphelin de père et celle d'être orphelin de mère ne sont pas indépendantes. Les seuils retenus (une moitié, un tiers, etc.) apparaissent toutefois un peu arbitraires. La proportion d'orphelins parmi ceux qui ignorent si l'un de leurs parents survit va logiquement varier avec la mortalité.

Plutôt que de définir des seuils a priori, une alternative consisterait à se baser sur les proportions *observées* à partir des réponses non-manquantes (par exemple, pour les orphelins de père qui ignorent si leur mère est survivante, utiliser la proportion de doubles orphelins parmi les orphelins de père). Cette correction, plus proche des données, reviendrait à supposer que *la méconnaissance de la survie de son père ou de sa mère n'est pas associée à une plus grande probabilité d'être orphelin de père ou de mère*. C'est en somme la même hypothèse que lorsque les données manquantes sont simplement ignorées. Par contre, cette approche tiendrait compte de l'information fournie sur la survie de l'autre parent, comme le suggèrent Bicego et al. (2003).

Tout cela complique toutefois sensiblement l'analyse. N'est-il pas préférable d'ignorer simplement les réponses manquantes ? Le choix des différentes approches a-t-il un impact important sur les estimations ? Nous discutons de cette question en annexe A.6.2 (p. 443). *Notre conclusion est que la meilleure option consiste tout simplement à ignorer les réponses manquantes.* Cela a l'avantage de la simplicité, et c'est également défendable sur le plan théorique. D'une part, on pourrait supposer que les données manquantes sont plus fréquentes parmi les orphelins [18], mais d'autre part, il est pos-

16. Les enfants chez qui cette information manque pour l'un des deux parents, sont recodés comme double orphelins si l'autre est décédé.

17. Dans le cas d'une réponse manquante sur la survie des pères, 50% des enfants orphelins de mères sont codés comme double orphelins, et 25% des enfants non-orphelins de mère sont codés comme orphelins de père. Quant aux enfants pour lesquels l'information manquent sur *les deux* parents (qui représentent la majorité des données manquantes), ils sont exclus des analyses.

18. Les enfants qui auraient perdu contact avec leurs parents ne peuvent renouer contact avec eux quand ils sont décédés.

sible également qu'un certain nombre de personnes ayant perdu contact avec leurs parents les déclarent comme décédés. On supposera ici que ces deux types d'erreurs se compensent.

6.1.3 Erreurs sur les âges

La méthode des orphelins nécessite que l'âge des répondants soit bien déclaré, car l'estimation de la durée d'exposition en dépend. Dans les recensements, cela concerne l'âge des individus repris dans les feuilles ménages, et dans les enquêtes rétrospectives (type EDS), l'âge des enfants de moins de 15 ans.

Classiquement, deux types de problèmes se posent (Ewbank 1981, Pullum 2006) : certains âges ronds (se terminant par 0 et 5) "attirent" très souvent, certains âges sont également erronés en raison de vieillissements ou de rajeunissements artificiels.

Attirance sur les âges ronds

Le premier de ces problèmes va surtout induire une *sous-estimation de la mortalité*. Comme les chiffres ronds marquent l'entrée dans une nouvelle classe d'âge (5 pour 5-9 ans, 10 pour 10-14 ans,...), les déplacements vers le bas auront moins d'effet que les déplacements vers le haut [19]. Par exemple, les proportions de non-orphelins ne sont pas modifiées si une certaine proportion d'enfants de 6 et 7 ans sont enregistrés comme ayant 5 ans. Par contre, si un certain nombre d'enfants âgés de 3 à 4 ans sont déplacés vers 5 ans, ils vont transférer vers un autre groupe d'âge des risques de mortalité plus faibles. En effet, comme ces enfants sont plus jeunes, leurs parents ont été moins longtemps exposés à la mortalité. Il conviendrait de mesurer les effets de ces déplacements sur la mortalité adulte.

Vieillissements et rajeunissements

Les vieillissements et rajeunissements d'âge sont communs en Afrique subsaharienne, et on observe généralement les deux phénomènes en même temps. Selon les estimations anciennes de Gibril (1979), basées sur le recensement gambien de 1973, les hommes ont tendance à se vieillir de 2.5 ans en moyenne entre 30 et 49 ans et d'au moins 5 ans au delà de 50 ans. Comme les proportions de non-orphelins chutent rapidement avec l'âge, de tels vieillissements artificiels vont provoquer une sous-estimation importante de la mortalité (Blacker 1984), puisqu'ils conduisent à surestimer la durée d'exposition des parents au risque de mourir.

Chez les femmes, Gibril (1979) note un rajeunissement en-dessous de 40 ans, puis une exagération des âges au-delà. La présence d'erreurs sur les âges peut donc s'observer en comparant les proportions de parents survivants déclarée par sexe : ces dernières s'avèrent souvent *plus élevées* lorsqu'elles sont rapportées par les hommes que lorsqu'elles sont rapportées par les femmes (Hill 1984, Blacker 1984). Le graphe 6.4, construit à

19. Pour qu'ils introduisent une erreur, les attirances vers le bas devraient se faire à plus de 5 ans d'intervalle.

partir des données du recensement ougandais de 2002 [20], confirme ce résultat.

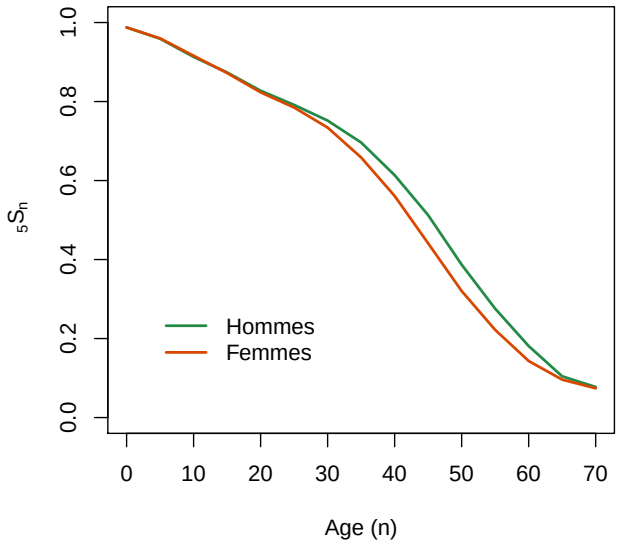

FIGURE 6.4 – *Proportions de mères survivantes déclarés par âge et par sexe - Recensement ougandais de 2002*

Les proportions de mères survivantes déclarées par les hommes et les femmes sont quasiment identiques jusqu'à 20-25 ans. Ensuite, les hommes se déclarent plus nombreux à avoir une mère survivante. Entre 50 et 65 ans, *les proportions de mères survivantes déclarées par les hommes sont supérieures de plus de 20% à celles déclarées par les femmes*. En raison de ces erreurs, il est préférable d'exploiter les données sans répartir les enquêtés par sexe. Il devrait y avoir une compensation au moins partielle entre le vieillissement des hommes et le rajeunissement des jeunes femmes.

Il est toutefois difficile d'évaluer l'effet "net" de ces deux types d'erreurs d'âge sur les estimations de mortalité. Cet effet va dépendre de l'intensité des erreurs d'âge, ainsi que de la forme de la courbe de survie. Plus celle-ci sera incurvée, plus la sous-estimation de la mortalité sera élevée. Ewbank (1981) a procédé par simulation pour analyser l'effet des erreurs d'âge sur la méthode des orphelins (à partir du standard de Brass). Il montre qu'une exagération de 2.5 ans sur les âges des enquêtés d'au moins 30 ans se traduit par un biais à la hausse dans les espérances de vie de 2.5 ans environ.

***

Après avoir présenté les principaux problèmes de qualité des données sur la survie des parents, nous discutons dans la section suivante de la qualité des données relatives aux frères et sœurs.

20. Il s'agit du fichier IPUMS, qui reprend 10% des microdonnées sur recensement. https ://international.ipums.org/international/

6.2 Données portant sur la fratrie

Hill et Trussell (1977), qui ont mis au point la première méthode permettant d'utiliser les données sur les frères et sœurs, ont tout de suite mis en garde contre une éventuelle sous-déclaration des décès. Selon eux, la principale limite de ces données était qu'un certain nombre de frères et sœurs ont pu mourir avant la naissance des enquêtés - ou durant leur enfance. Compte tenu de la taille des fratries africaines, il est clair que certains naissances et décès sont survenus loin dans le passé, ce qui rend leur restitution difficile.

Afin de déceler ces problèmes, plusieurs indicateurs sont régulièrement présentés dans les rapports d'enquêtes EDS. Ils concernent surtout le *nombre* de frères et sœurs, et pas directement leurs décès. L'hypothèse implicite est que si des frères et sœurs sont sous-déclarés, il s'agit sans doute davantage de frères et sœurs décédés. Notons toutefois que la sous-déclaration des décès n'affecte les estimations que s'il s'agit de décès *adultes* [21].

Les principaux indicateurs de qualité des données sont les suivants :
- l'examen du nombre de frères et sœurs déclaré par âge permet de suspecter une *sous-déclaration plus importante par les enquêtés âgés* (*cf.* section 6.2.4)
- les rapports de masculinité à la naissance des frères et sœurs permettent de suspecter *une sous-déclaration différentielle par sexe* (*cf.* section 6.2.2)
- les écarts entre la médiane des années de naissance des enquêtés et la médiane des années de naissance des frères et sœurs permettent de suspecter *une sous-déclaration des frères et sœurs plus importante lorsqu'ils sont nés avant les enquêtés* (*cf.* section 6.2.3).

Il y a une dizaine d'années, Stanton et al. (2000) ont proposé une première synthèse de ces indicateurs à partir des enquêtes EDS disponibles en 1997 (8 en Afrique subsaharienne). Ils ont également examiné la distribution des données manquantes. Nous reprenons ici leurs principales conclusions avant de revenir en détail sur certains points.

Stanton et al. (2000) font les constats suivants :
- Les données manquantes sur la *survie* des frères et sœurs restent rares.
- A l'exception de quelques pays, les *âges à l'enquête* des frères et sœurs survivants sont enregistrés à plus de 95%.
- La complétude des informations (date des décès, etc) ne semble pas varier par sexe, ni avec le nombre d'années avant l'enquête.
- Les *rapports de masculinité à la naissance* des frères et sœurs restent proches des niveaux attendus (103) et relativement stables. Une diminution de ces rapports avec le nombre d'années avant l'enquête traduirait une sous-déclaration des sœurs nées longtemps avant l'enquête.
- l'examen des dates de naissance des enquêtées et de celles des frères

21. Seule la méthode indirecte de Hill et Trussell (1977) fait intervenir les décès dans l'enfance.

et sœurs (c-à-d. la comparaison de leurs médianes respectives) indique que les aînés ne semblent pas moins bien déclarés que les cadets.

La qualité des données paraît donc satisfaisante à partir de ces indicateurs. Mais les mêmes auteurs relèvent toutefois quelques problèmes ;

- une forte concentration des années écoulées entre le décès et l'enquête sur *les valeurs 5, 10 et 15*,
- une concentration des décès dans les années précédant directement l'enquête - ce qui résulte de *déplacements* et/ou d'une *sous-déclaration des décès* dans le passé plus lointain,
- un nombre de frères et sœurs déclarés par les enquêtées *plus âgés* parfois équivalent - voire plus faible - que le nombre déclaré par les plus jeunes enquêtées, ce qui semble être en contradiction avec la tendance globalement à la baisse de la fécondité (*cf.* section 6.2.4),
- un enregistrement des *âges au décès* bien moins complet que celui des âges des survivants, avec de grandes variations d'une enquête à l'autre. Selon Stanton et al. (2000), cela résulte davantage de variations dans la qualité des enquêtes que de véritables différences en termes de capacité des enquêtés à fournir les informations nécessaires [22].

Nous allons approfondir certains de ces aspects. Nous nous concentrons d'abord sur les *données manquantes* au sein des enquêtes EDS, avant de revenir sur certains *indicateurs* déjà mentionnés, utilisés régulièrement pour juger de la qualité des données. Ces derniers se réfèrent en fait à une certaine forme de "cohérence interne" des informations sur les frères et sœurs. Mais ils nous disent que peu de choses sur la sous-déclaration des décès eux-mêmes. Pour tenter d'estimer l'ampleur de cette sous-déclaration, il faut plutôt chercher du côté des niveaux de mortalité obtenus, en comparant *les niveaux estimés pour une même période par plusieurs enquêtes successives*. Nous présenterons plus loin cette démarche, et nous l'appliquerons aux EDS pour rehausser les probabilités de décès estimées par la méthode directe.

6.2.1 Données manquantes

Le tableau 6.2 reprend une série d'enquêtes EDS, classées par ordre selon *la proportion de frères et sœurs dont les enquêtés ignorent le sexe* [23]. Le tableau reprenant toutes les enquêtes est en annexe A.6.3 (p.447). Ces proportions sont calculées successivement sur tous les frères et sœurs, sur ceux qui sont décédés, ceux qui ont survécu, et ceux dont l'enquêté ignore également s'ils ont survécu ou non. Elles sont calculées séparément pour les enquêtes menées auprès des hommes et des femmes. Dans l'ensemble, à l'exception de l'enquête menée au Sierra Leone en 2008, la proportion

22. Stanton et al. (2000) relèvent aussi des problèmes au niveau des informations sur les décès maternels.

23. Il s'agit en fait des 5 enquêtes où la proportion de réponses manquantes est *la plus faible* (de la Centrafrique au Mozambique), de 5 enquêtes où cette proportion est proche de la moyenne observée sur toutes les enquêtes (du Swaziland au Zimbabwe) et des 5 enquêtes où cette proportion est la plus élevée (de l'Ouganda au Sierra Leone).

	Total	Décédés	Survivants	NSP	Pays	Année	Type
1	0.0	0.0	0.0	0.0	Centrafique	1995	♀
2	0.0	<0.1	<0.1	<0.1	Tchad	2004	♀
3	<0.1	<0.1	<0.1	<0.1	Côte d'Ivoire	1994	♀
4	<0.1	<0.1	<0.1	<0.1	Madagascar	1992	♀
5	<0.1	<0.1	<0.1	<0.1	Mozambique	1997	♀
32	0.3	0.3	<0.1	66.3	Swaziland	2006	♀
33	0.3	0.7	0.1	60.5	Tanzanie	1996	♂
34	0.3	0.2	0.1	58.9	Zimbabwe	2006	♂
35	0.4	0.4	<0.1	65.8	Mozambique	2003	♀
36	0.4	0.8	0.1	82.5	Zimbabwe	1994	♂
63	1.2	0.5	0.1	62.3	Ouganda	1995	♂
64	1.2	0.9	0.1	77.8	Ouganda	2001	♀
65	1.8	4.3	0.1	77.0	Burkina Faso	1999	♀
66	1.9	3.1	0.7	94.7	Cameroun	1998	♀
67	1.9	2.8	0.4	83.2	Afrique du Sud	1998	♀
68	9.0	0.1	0.1	96.1	Sierra Leone	2008	♀

TABLE 6.2 – *Proportion d'observations dont le sexe est inconnu - par statut de survie (en %) - enquêtes EDS*

de frères et sœurs dont le sexe est inconnu reste faible, avec une médiane à 0.35% pour l'ensemble des frères et sœurs. Cette proportion est par contre beaucoup plus élevée parmi les frères et sœurs pour lesquels le répondant ignore également s'ils ont survécu ou non (avec une médiane à 65%) - mais ces derniers sont rares (environ 0.5%).

La proportion de frères et sœurs mentionnés par l'enquêté *pour lesquels celui-ci ignore s'ils ont survécu ou non* est en effet assez marginale. Ceci s'observe dans le tableau 6.3 pour une série d'enquêtes et en annexe A.6.3 pour le tableau complet. Sur l'ensemble des frères et sœurs, sa médiane vaut 0.2% et sa moyenne 0.5%. Ces valeurs sont inférieures à 0.2% tant chez les frères que chez les sœurs et s'élèvent par contre à plus de 50% parmi les frères/sœurs dont le répondant ignore également le sexe. Les données manquantes sur la survie des frères et sœurs mentionnés n'apparaissent pas plus fréquentes lorsque les enquêtes sont menées auprès des hommes. Notons les proportions plus élevées de réponses manquantes au Rwanda (2000) et surtout au Sierra Leone (2008), deux situations où les conflits sont associés à une plus grande incertitude quant à la survie des proches.

Enfin, le tableau 6.4 présente les *proportions de frères et sœurs survivants pour lesquels il manque une information sur l'âge à l'enquête*. Le tableau complet est repris en annexe A.6.3. Dans plus de la moitié des enquêtes, cette proportion s'établit à moins de 0.55%, mais une vingtaine d'enquêtes présentent des proportions supérieures à 1%, parfois jusqu'à 8-9% (Mozambique 1997 et Madagascar 1992). En moyenne, cette proportion s'élève à 1.2%. La complétude des informations ne semble pas varier par sexe.

	Total	Frères	Soeurs	NSP	Pays	Année	Type
1	<0.1	<0.1	<0.1	49.8	Tchad	2004	♀
2	<0.1	<0.1	<0.1	88.2	Côte d'Ivoire	1994	♀
3	<0.1	<0.1	<0.1	<0.1	Gabon	2000	♀
4	<0.1	<0.1	<0.1	<0.1	Mali	1996	♀
5	<0.1	<0.1	<0.1	92.0	Mozambique	1997	♀
32	0.2	<0.1	<0.1	44.0	Bénin	2006	♀
33	0.2	<0.1	0.1	42.5	Tchad	1997	♀
34	0.2	0.1	0.1	85.2	Congo RDC	2007	♀
35	0.2	0.1	0.1	64.4	Guinée	1999	♀
36	0.2	<0.1	<0.1	42.9	Madagascar	2004	♀
63	1.2	1.4	0.6	86.4	Rwanda	2005	♀
64	1.2	0.8	0.8	84.3	Ouganda	2006	♀
65	1.4	0.3	0.2	62.4	Afrique du Sud	1998	♀
66	1.6	0.8	0.4	83.1	Ouganda	1995	♂
67	2.1	2.1	1.0	71.7	Rwanda	2000	♀
68	9.3	0.4	0.4	99.0	Sierra Leone	2008	♀

TABLE 6.3 – *Proportion d'observations dont le statut de survie est inconnu - par sexe (en %) - enquêtes EDS*

	Total	Frères	Soeurs	NSP	Pays	Année	Type
1	<0.1	<0.1	<0.1	100.0	Centrafique	1995	♀
2	<0.1	<0.1	<0.1	37.0	Ethiopie	2000	♀
3	<0.1	<0.1	<0.1	<0.1	Guinée	1999	♀
4	<0.1	<0.1	0.1	<0.1	Malawi	2000	♀
5	0.1	0.1	<0.1	<0.1	Tanzanie	2004	♀
32	0.5	0.5	0.5	<0.1	Congo RDC	2007	♀
33	0.5	0.5	0.5		Guinée	2005	♀
34	0.5	0.5	0.5	<0.1	Zambie	2007	♂
35	0.6	0.6	0.6		Congo	2005	♀
36	0.6	0.6	0.6	<0.1	Zimbabwe	2006	♀
63	4.0	3.9	4.1	6.0	Namibie	1992	♀
64	4.7	4.8	4.7	2.5	Afrique du Sud	1998	♀
65	6.6	6.3	6.8	23.6	Madagascar	1997	♀
66	6.6	6.5	6.6	34.8	Namibie	2000	♀
67	8.0	8.0	7.9	94.4	Mozambique	1997	♀
68	9.6	9.5	9.7	66.8	Madagascar	1992	♀

TABLE 6.4 – *Proportion de frères et soeurs survivants dont l'âge à l'enquête est inconnu - par sexe (en %) - enquêtes EDS*

Venons-en aux *données manquantes sur la datation des décès*. Il peut manquer une information (1) soit sur l'âge au décès, (2) soit sur le nombre d'années écoulées depuis le décès, (3) soit sur les deux.

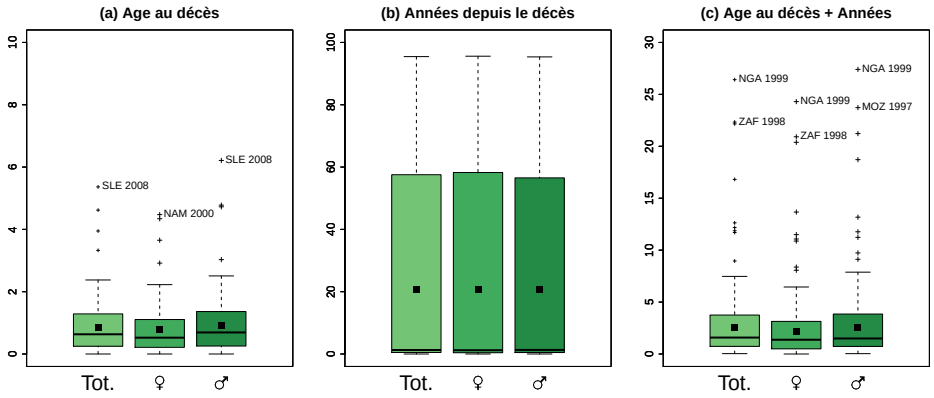

FIGURE 6.5 – *Proportions de frères et sœurs décédés (par enquête) pour lesquels il manque l'âge au décès (a), le nombre d'années écoulées depuis le décès (b) et les deux informations (c)*

Les boxplots 6.5 présentent sur plus de 65 enquêtes les proportions de données manquantes sur ces deux variables et leur combinaison. Les moyennes (pondérées par le nombre de frères et sœurs décédés observés dans chaque enquête) sont indiquées par les carrés [24]. Le tableau correspondant, qui reprend par sexe les proportions de frères et sœurs pour lesquelles ces données sont manquantes, est présenté en annexe A.6.3 (p. 450).

Comme ont pu le montrer Stanton et al. (1997), il apparaît que les années écoulées depuis le décès sont nettement moins bien enregistrées que les âges au décès. Il n'y a par contre pas de nette différence par sexe.

1. En moyenne, *l'information sur les années écoulées depuis le décès est manquante pour environ 20% des frères et sœurs décédés*, quel que soit leur sexe. Dans la majorité des enquêtes, elle est pourtant inférieure à 1.3%, mais elle est supérieure à 55% dans un quart des enquêtes. D'une enquête à l'autre, il y a donc une très grande variation dans la complétude des informations receuillies.
2. *Les proportions de données manquantes sur l'âge au décès sont nettement moins élevées*, surtout pour les sœurs. Tous sexes confondus, elles sont inférieures à 1% dans 3/4 des enquêtes.
3. *La proportion de frères et sœurs décédés pour lesquels il manque à la fois un âge au décès et une localisation dans le temps est par contre un peu plus élevée* (2.5% en moyenne). Certaines enquêtes se distinguent par des proportions très élevées : Nigeria 1999 (♀), Afrique du Sud 1998 (♀) ou Mozambique 1997 (♀) avec plus de 20% des frères et sœurs décédés pour lesquels aucune information n'est disponible ni sur l'âge au décès ni sur les années écoulées depuis. Dans ces cas, l'imputation repose notamment sur l'ordre de naissance des frères

24. Pour ces calculs ainsi que pour le tableau en annexe, dans le cas où le répondant ne connaît pas l'âge au décès mais sait établir que son frère ou sa sœur est décédé(e) avant l'âge de douze ans, nous considérons les données comme non manquantes.

et sœurs déclarés (Stanton et al. 1997). Les estimations obtenues sont donc à prendre avec prudence.

En définitive, la fréquence des données manquantes sur les décédés varie énormément d'une enquête à l'autre. Si l'on classe les enquêtes selon la fréquence des données manquantes sur *au moins* l'une des variables (âge au décès ou années depuis le décès), on peut établir que dans une enquête sur 4, la proportion de décédés pour lesquels il manque l'une des deux informations s'élève à 2.2% seulement. Dans une enquête sur deux, cette proportion est inférieure à 5.2%. Par contre, il y a toute une série d'enquêtes où une imputation est nécessaire pour une large fraction des décédés. *Dans une enquête sur quatre (les moins bonnes), au moins l'une des informations est manquante pour plus de 56% des décédés*. Pire, dans plus d'une quinzaine d'enquêtes, une imputation est nécessaire pour plus de 80% des décédés. Précisons que pour certaines EDS, la fréquence des réponses manquantes nous semble très sous-estimée par les rapports d'enquêtes, en raison du fait qu'ils ne comptabilisent que les réponses codées comme manquantes ("98"), alors qu'il y a parfois des proportions élevées de réponses manquantes non codées comme telles (" "). Ceci est discuté dans l'annexe A.6.3 (p.455).

Quel est l'impact de ces données manquantes sur les estimations ? Les niveaux obtenus indirectement ne sont en fait que faiblement influencés par les procédures d'imputations. En effet, il s'agit uniquement d'établir si les frères/sœurs ont survécu jusqu'à leur 15e anniversaire. Par contre, les estimations directes en dépendent ; il s'agit à la fois de localiser le décès dans le temps et par groupe d'âge. Dans certaines enquêtes, les proportions importantes de décédés pour lesquels une imputation est nécessaire invitent donc à la prudence. Il faut toutefois bien garder à l'esprit que les estimations directes ne sont pas influencées par les erreurs ou par les omissions lorsque celles-ci concernent les frères et sœurs dans l'enfance. Elles ne sont pas non plus influencées par les décès anciens, survenus plus de dix ans avant l'enquête.

6.2.2 Sous-déclaration différentielle par sexe

Les informations données sur les sœurs sont-elles aussi fiables que celles fournies sur les frères ? La section précédente a montré que *la proportion des informations manquantes ne varie pas par sexe*. Mais l'un des deux sexes pourrait être davantage sous-déclaré que l'autre, sans que ceci n'apparaisse au niveau de données manquantes.

A priori, il y a des risques de sous-déclarer davantage les frères, puisque ces derniers sont plus susceptibles de migrer, par exemple (Bicego 1997). Mais une sous-déclaration des sœurs est possible également, notamment dans les contextes où les femmes mariées habitent dans le lignage des époux (patrilocalité), les éloignant de leur fratrie d'origine (Stanton et al. 1997).

Pour rechercher une éventuelle sous-déclaration différentielle par sexe, les rapports des EDS rapportent le nombre de frères déclarés au nombre de sœurs déclarés. Nous présentons dans le graphe 6.6 la distribution des rapports de masculinité à la naissance calculés sur les enquêtes (menées

auprès des femmes). La moyenne de ces rapports s'établit à 1.03. Elle est inférieure à la moyenne mondiale de 1.055, mais correspond tout à fait à la valeur attendue en Afrique subsaharienne (UN 2009c).

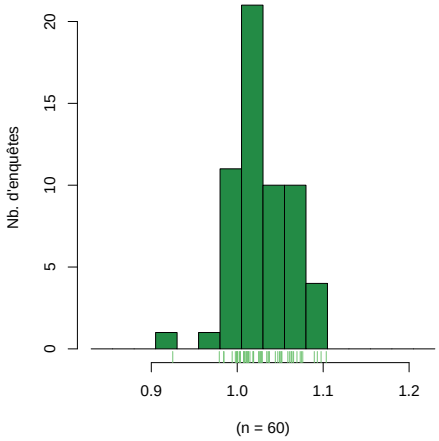

FIGURE 6.6 – *Rapports de masculinité à la naissance des frères et sœurs déclarés - 60 enquêtes EDS*

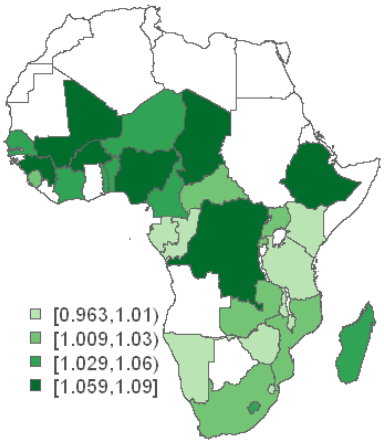

FIGURE 6.7 – *Rapports de masculinité des frères et sœurs déclarés (moyennes des valeurs observées par pays) - Enquêtes EDS*

On observe également une grande variation dans les rapports de masculinité : dans une enquête sur deux, ces rapports se situent en-deça de 1.008 ou au-delà de 1.052. Cette variation correspond en fait à une certaine distribution régionale. Ceci n'a jamais été relevé dans la littérature, mais cela apparaît assez nettement sur la carte en figure 6.7. Cette carte combine différentes enquêtes en prenant une moyenne des rapports observés par pays. On observe des rapports assez faibles en Afrique de l'Est et en Afrique australe. Au Malawi, en Namibie et au Zimbabwe, les enquêtées déclarent même plus de sœurs que de frères. Les rapports observés en Afrique de l'Ouest sont par contre plutôt situés autour de 1.06. Les pays qui présentent les rapports de masculinité les plus élevés sont le Nigeria et l'Éthiopie (1.09).

Nous suggérons plusieurs explications à cette distribution régionale :

1. Les rapports plus faibles en Afrique australe et en Afrique de l'Est pourraient résulter de *comportements migratoires masculins plus importants*, qui amèneraient les enquêtées à sous-déclarer davantage les frères éloignés, parfois depuis longtemps.
2. Cette répartition régionale pourrait refléter aussi une *géographie des inégalités sexuelles de mortalité* infanto-juvénile. L'Afrique de l'Ouest se caractérise davantage par une surmortalité des petites filles, tandis que la surmortalité masculine concerne plutôt l'Afrique australe (Tabutin et al. 2007). Comme les omissions les plus fréquentes sont les décès survenus dans l'enfance, ces variations régionales pourraient participer à ce qui est observé ici.
3. Plutôt que de renvoyer à des différences régionales dans la qualité des données, il pourrait s'agir simplement d'*une réelle variation dans les rapports de masculinité à la naissance*. La géographie observée ici

recoupe en effet celle qui peut être dégagée des histoires génésiques des femmes dans les enquêtes EDS. A partir de ces données sur les enfants, Garenne (2002) isole trois sous-groupes de pays africains : un premier aux rapports peu élevés, essentiellement dans les populations bantoues de l'Afrique australe et de l'Afrique de l'Est (avec des rapports qui avoisinent 1.01), un second sous-groupe comprenant des pays aux rapports moyens (103.5), et enfin, un troisième sous-groupe présentant des rapports élevés (107), en particulier au Nigeria et en Ethiopie. On peut faire quasiment les mêmes constats à partir des frères et sœurs, en ce compris pour le Nigeria et l'Ethiopie. A titre indicatif, nous avons croisé les rapports de masculinité calculés chez les enfants et ceux calculés chez les frères à partir des mêmes enquêtes [25]. Le coefficient de corrélation entre les deux séries s'élève à 0.67. Il se pourrait qu'une partie de l'explication vienne de là.

4. Enfin, un quatrième type d'explication peut être avancé. Nous avons vu (dans le chapitre 1) que plusieurs sœurs pouvaient être interrogées dans le même ménage. Dès lors, les fratries où les *sœurs sont proportionnellement plus nombreuses* sont davantage susceptibles d'être mentionnées à plusieurs reprises que les fratries où *les frères sont proportionnellement plus nombreux*. Ces fratries plus féminines sont donc surreprésentées dans les fichiers. Ceci a pour effet de biaiser les rapports de masculinité à la baisse. D'ailleurs, si nous calculons les rapports de masculinité à partir des mêmes enquêtes, mais auprès des hommes enquêtés (lorsque le module a été intégré au questionnaire homme), nous obtenons des valeurs bien supérieures. Par exemple, en Ouganda en 1995, les femmes enquêtées mentionnent 1.01 frère pour une sœur, mais les hommes enquêtés en mentionnent plus de 10% en plus (1.13). La distribution régionale des rapports de masculinité observée sur la carte 6.7 reflète donc en partie la fréquence de la cohabitation avec les sœurs. Et précisément, nous avions montré dans le premier chapitre que les proportions d'enquêtées cohabitant avec d'autres sœurs éligibles étaient les plus élevées en Afrique australe et en Afrique de l'Est. Cela se traduit par des rapports de masculinité plus faible.

En définitive, ces différents éléments ne nous permettent pas de conclure dans le sens d'une sous-déclaration plus importante d'un sexe par rapport à l'autre.

6.2.3 Sous-déclaration des frères et sœurs nés avant les enquêtés

Nous l'avons relevé, une des principales craintes mentionnée dans la littérature est que la sous-déclaration des événements soit particulièrement forte pour les périodes *précédant la naissance de l'enquêté* et lorsque celui-ci est tout jeune (Hill et Trussell 1977, Gakidou et al. 2004, Stanton et al. 2000). En d'autres termes, les enquêtés pourraient omettre un certain nombre de frères aînés, et *a fortiori*, sous-déclarer leur décès. En

25. En partant des estimations des rapports de masculinité à la naissance de Garenne (2002).

fait, cette crainte n'est pas justifiée dans le cas d'une exploitation *directe*, car on analyse généralement les données pour la période des 10 ans au maximum qui précédent l'enquête. Donc l'enquêté est âgé d'au moins cinq ans quand débute la période d'observation. Par contre, il est vrai que cette sous-déclaration peut avoir un impact sur les estimations *indirectes*, si elle concerne des frères et sœurs ayant atteint l'âge adulte.

Pour mesurer l'ampleur de cette éventuelle sous-déclaration, les rapports d'enquêtes EDS présentent le résultat d'un calcul assez intuitif : comme chacun a à peu près autant de petits frères que de grands frères, on peut considérer que l'année de naissance d'un enquêté est très proche de la médiane des années de naissance de ses frères et sœurs. Si les individus nés *avant* l'enquêté(e) sont particulièrement sous-déclarés, la médiane des dates de naissances des frères et sœurs va être *trop élevée* ; elle va dépasser de plusieurs années celle des enquêté(e)s.

Sur l'ensemble des enquêtes EDS, nous observons en fait que la distribution des écarts entre ces deux médianes est concentrée autour de zéro (résultats non présentés ici). L'écart moyen entre ces deux médianes s'élève à 2 mois et demi seulement.

Bien qu'utilisée de façon routinière dans les rapports d'enquête, cette méthode de détection des erreurs nous semble en fait assez peu utile, car il est difficile de savoir à partir de quel seuil on peut considérer que la sous-déclaration des aînés est particulièrement forte [26].

6.2.4 Taille des fratries et descendance finale

Dans la plupart des enquêtes EDS analysées par Stanton et al. (1997; 2000), les plus jeunes enquêtés (15-19 ans) semblent déclarer des tailles de fratries plus élevées que les enquêtés plus âgés. Or, selon ces auteurs, on devrait *a priori* observer la situation inverse. Leur raisonnement est le suivant : comme la fécondité a globalement baissé dans une majorité de pays africains, les fratries des plus jeunes se sont constituées à un moment où la fécondité était plus faible. Le nombre de frères et sœurs déclarés par les plus jeunes devrait donc être *inférieur* aux tailles de fratries déclarés par les plus âgés. Stanton et al. (1997; 2000) pensent avoir identifié là l'un des principaux problèmes de qualité des données sur les frères et sœurs. Le fait que les tailles de fratries baissent avec l'âge des enquêtés alors qu'elles devraient augmenter (ou tout au plus rester stables) indique que *les frères et sœurs sont particulièrement sous-déclarés par les enquêtées plus âgées* :

> "The data quality assessment ... suggests some omission of siblings from reports of older DHS respondents. Without valid external data sets for comparative purposes, determining the effect, if any of such omission on the adult or maternal mortality estimates is not possible." - Stanton et al. (2000), p. 120

Stanton et al. (1997; 2000) pointent en particulier du doigt les enquêtes menées à Madagascar en 1992 et au Sénégal en 1992-93. Les tailles

26. De plus, le rapport du nombre d'aînés sur le nombre de cadets dépend de la croissance de la population (Goldman 1978), ce qui brouille quelque peu l'analyse.

moyennes de fratries observées par âge dans ces deux enquêtes sont présentées dans le graphe 6.8. On constate effectivement que dans ces deux cas (les enquêtes de 1992 et 1993 sont reprises en gras), la taille des fratries diminue rapidement avec l'âge des enquêtées. Les 45-49 ans déclarent environ 20% de frères et sœurs de moins que les 15-19 ans. Mais cette baisse avec l'âge ne se retrouve pas dans toutes enquêtes. Par exemple, on observe à Madagascar des schémas assez différents d'une enquête à l'autre. Dans l'enquête menée en 2009 un nombre de frères et sœurs est plus élevé chez les enquêtées plus âgées, alors que c'était l'inverse en 1992.

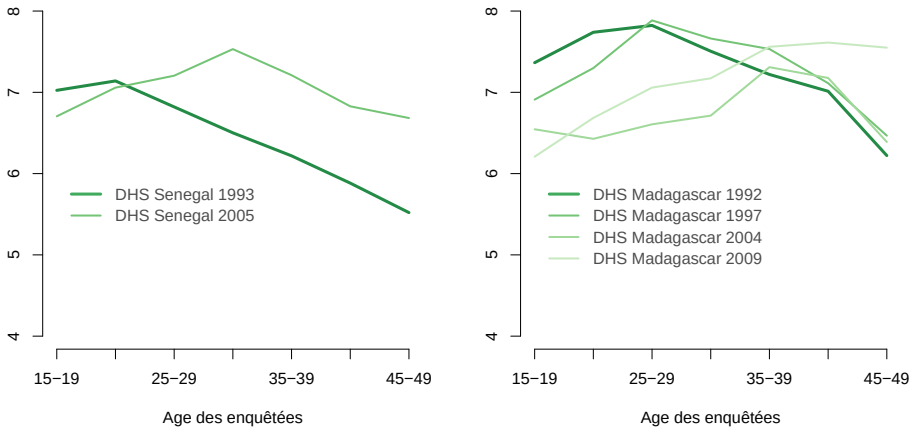

FIGURE 6.8 – *Tailles des fratries déclarées selon l'âge des enquêtées - EDS menées au Sénégal et à Madagascar*

Cet indicateur de qualité des données a ses limites, car il y a toute une série de pays africains où la fécondité a d'abord augmenté dans la période 1960-1980 avant d'amorcer une baisse (c'est le cas au Bénin, au Burkina Faso, au Congo, etc.). Dans ces pays, la baisse de la taille des fratries avec l'âge des enquêtés pourrait bien être un résultat des tendances passées de la fécondité plutôt qu'un véritable problème de qualité des données.

En fait, pour pouvoir creuser cette question de façon un peu plus rigoureuse, nous proposons une nouvelle approche basée sur des données relatives aux *parités atteintes déclarées par les femmes*. Cette approche part du principe que la taille moyenne des fratries est étroitement liée aux parités atteintes (Keyfitz et Caswell 2005). En procédant à quelques ajustements, le nombre d'enfants nés vivants déclarés par les femmes dans un passé plus ou moins lointain devrait être proche de la taille des fratries déclarées quelques années plus tard par la génération suivante.

Pour tester cette hypothèse sur les données d'enquêtes, la démarche est la suivante :

1. pour chaque groupe d'âge, nous identifions d'abord la cohorte des mères des enquêtées, en partant de l'âge des enquêtées et de l'âge moyen à la naissance des enfants qui prévalait à leur naissance. Par exemple, si cet âge moyen vaut 25 ans, les mères des enquêtées de 15-19 ont aujourd'hui 40-45 ans.

2. une fois que nous connaissons la cohorte des mères, nous pouvons rassembler des données sur les enfants nés vivants qu'elles ont eus. Par exemple, si ces mères sont nées en 1950, nous cherchons les parités atteintes des femmes âgées de 45 ans en 1995.
3. nous comparons ensuite le nombre d'enfants nés vivants de ces femmes (que nous avons identifiées comme les mères des enquêtées), aux tailles des fratries déclarées par ces enquêtées.

Prenons un exemple concret. Si nous calculons l'âge moyen à la naissance des enfants au Sénégal à partir des enquêtes EDS, nous pouvons constater qu'il n'a pas changé énormément sur les 20 dernières années. Il passe de 27.5 ans en 1986 à 27.9 ans en 2005. Supposons pour le moment que cet âge moyen soit resté stable, autour de 27 ans sur les 50 dernières années. Cela signifie que les enquêtées âgées de 45 à 49 ans en 2005 sont nées en moyenne de mères qui étaient âgées de 45-49 ans en 1978. Ceci est illustré dans le graphe 6.9 avec deux femmes, la mère née en 1931 et sa fille née en 1958. Si nous supposons qu'à 40-45 ans, les femmes ont atteint leur descendance finale, la taille des fratries déclarée par les enquêtées de 40-45 ans dans l'enquête de 2005 devrait être proche du nombre moyen d'enfants nés vivants des femmes des mêmes groupes d'âge en 1978. Or précisément, nous disposons de ces données grâce à l'enquête de fécondité de 1978 (WFS). Nous pouvons donc comparer les deux estimations (*cf.* infra).

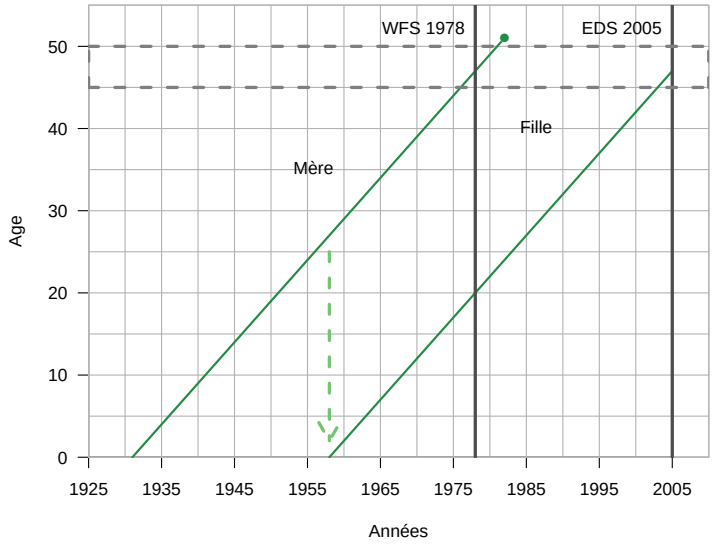

FIGURE 6.9 – *Exemple du choix des générations utilisées pour comparer la taille des fratries et le nombre d'enfants nés vivants - mère née en 1931, fille née en 1958*

Jusque là nous n'avons exploité que les données sur la fratrie des femmes de 40-45 ans en 2005. Les femmes âgées de 35 à 39 ans en 1978 n'avaient pas encore atteint leur descendance finale, ce qui empêche de confronter la taille des fratries des enquêtés du même âge en 2005 avec le nombre d'enfants nés vivants observés dans l'enquête de 1978. Mais il suffit de suivre la cohorte de ces mères un peu plus loin pour qu'elles puissent procurer des informations sur leur parité totale. Au moment où la première enquête EDS est organisée au Sénégal (1986), ces mères sont

âgés de 43 à 47 ans. Par le biais de quelques interpolations, on peut donc utiliser l'enquête de 1986. Et on procède de la même façon pour l'enquête de 1992, puis de 2005.

Nous avons appliqué cette approche aux enquêtes EDS disposant de données sur les frères et sœurs. Les tendances de l'âge moyen à la naissance des enfants sont calculées à partir des schémas de fécondité et des structures par âge des Nations Unies. Pour pouvoir être strictement comparable aux tailles des fratries, le nombre d'enfants nés vivants est légèrement rehaussé, en tenant compte de la variance de la distribution des parités atteintes [27]. Les informations sur le nombre d'enfants nés vivants proviennent essentiellement des EDS, des enquêtes WFS et des annuaires statistiques des Nations Unies (UN 1997; 1966) [28].

Le graphe 6.10 présente pour 8 enquêtes bien documentées les résultats de cette comparaison. En haut à gauche, nous retrouvons l'enquête sénégalaise de 1993 : le nombre d'enfants nés vivants suit globalement la même tendance que celle esquissée par les tailles de fratries. Ce qui nous était apparu comme un problème de données peut donc s'expliquer au moins partiellement par l'historique de la fécondité. Nous pouvons faire les mêmes constats au Burkina Faso (1999), en Côte d'Ivoire (1994) ou au Malawi (2004). On note un degré d'accord assez satisfaisant entre les différentes estimations ici, ce qui peut paraître étonnant compte tenu du caractère exploratoire de la méthode, mais cela la conforte.

Si dans les rapports d'enquêtes, le nombre de frères et sœurs déclarés par âge est régulièrement utilisé comme un indicateur de qualité, nous montrons finalement qu'*il apporte peu d'informations en soi, et qu'il est préférable de croiser l'évolution de la taille des fratries avec celle relative aux nombres d'enfants nés vivants.*

27. Sans procéder à aucune correction, et en l'absence complète d'erreurs, comparer les tailles de fratries déclarées aux parités atteintes nous amènera toujours à conclure que les premières sont légèrement plus élevées que les secondes. Keyfitz et Caswell (2005) notent : "The answer lies in the manner of selection : the number of daughters of a randomly selected mother is decidedly smaller than the number of sisters plus one of a randomly selected girl." En effet, dès qu'il y a une variation dans les schémas de fécondité, il y a un certain nombre de femmes qui auront plus de frères et sœurs, et un certain nombre de femmes qui en auront moins. Celles qui en auront plus seront plus fréquemment observés, ce qui hausse la taille des fratries par rapport au nombre d'enfants nés vivants. Nous renvoyons à la démonstration de Keyfitz et Caswell (2005) pour davantage de détails. Ce qui est important de noter ici, c'est que pour comparer le nombre d'enfants nés vivants (noté G), à la taille des fratries (notée F), il faut faire intervenir la variance de la distribution des descendances finales (σ^2) :

$$F = G + \frac{\sigma^2}{G} \tag{6.1}$$

Pour approximer σ^2, nous avons calculé la variance des parités atteintes par les femmes de 45-49 ans telles qu'elles sont déclarées dans les EDS. Cette variance s'élève en moyenne à 5.9 [4.7 - 7.4] sur les 88 enquêtes considérées ici. Nous appliquons cette valeur à toutes les enquêtes.

28. Certaines valeurs reprises ici nous ont été fournies par Patrick Gerland.

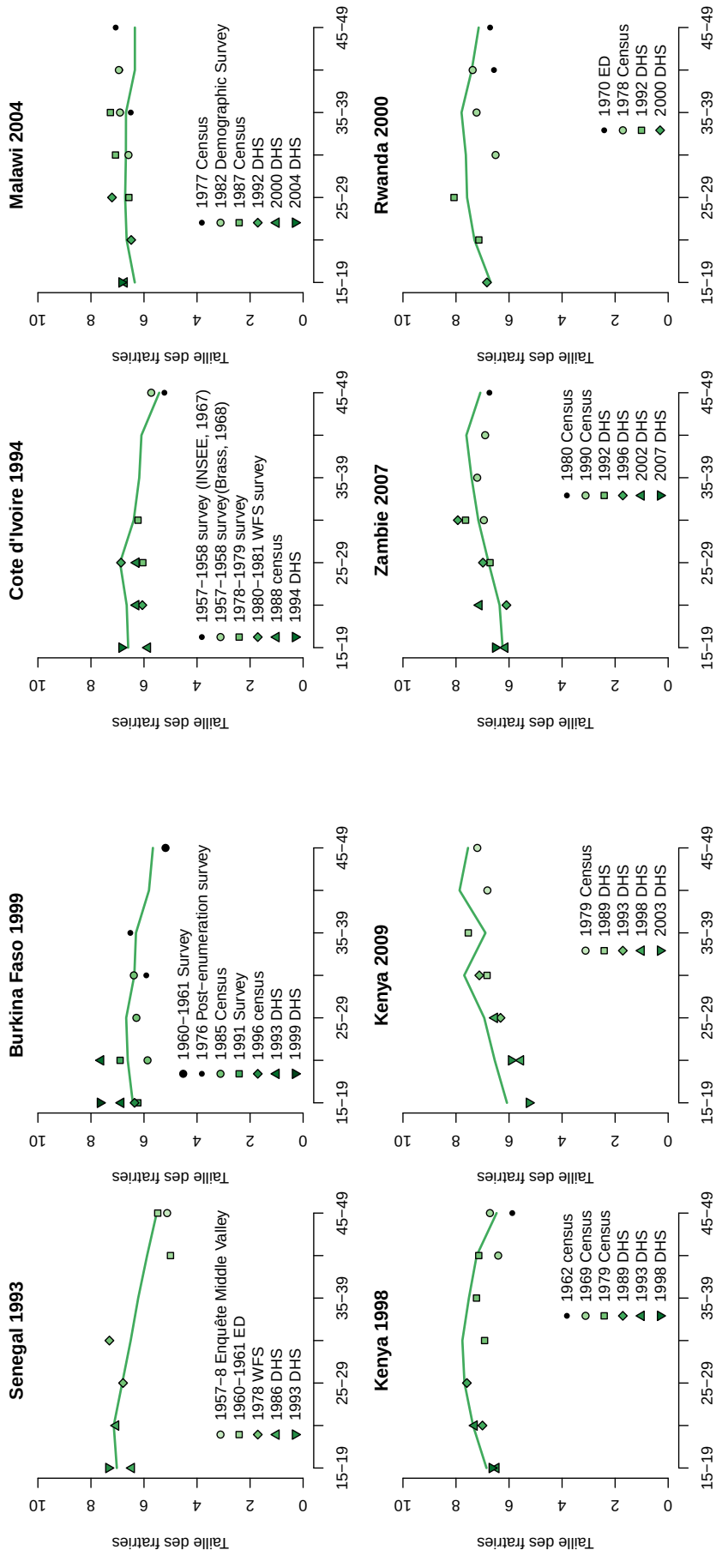

FIGURE 6.10 – *Tailles des fratries selon l'âge des enquêtées (EDS), et comparaison avec le nombre d'enfants nés vivants des femmes des générations antérieures (diverses sources)*

FIGURE 6.11

6.2.5 Erreurs faites sur les âges des frères et sœurs

Nous n'aborderons que rapidement la question des erreurs faites sur les âges des frères et sœurs, même s'il s'agit d'une question importante, qui a été relativement peu traitée dans la littérature. Il conviendrait d'appliquer *aux âges des frères et sœurs* les tests généralement utilisés pour détecter des erreurs au niveau des *âges des enquêtés* (indices de Myers, etc.). Nous pourrions également introduire dans les micro-simulations des *schémas typiques d'erreurs d'âge* (par exemple ceux mis à jour par Caldwell et Igun (1971) ou Bhat (1990)), afin de mesurer la sensibilité des méthodes directes et indirectes à ces erreurs. Un troisième type d'approche consisterait à croiser les informations fournies par différents enquêtés *sur une même fratrie* pour en analyser la cohérence, notamment des âges déclarés. Nous présentons ici simplement les premiers résultats d'un croisement de ce type.

Une enquête EDS spécifique (*Sumve Survey on Adult and Childhood Mortality - SACM 1995*) menée dans le nord de la Tanzanie offre la possibilité de confronter les déclarations faites par les enquêtées sur leur âge avec l'âge estimé pour ces mêmes femmes par leurs sœurs. Rappelons [29] que cette enquête a été conçue pour évaluer si l'histoire génésique des femmes pouvait être collectée de façon fiable en interrogeant leurs sœurs (Bicego et al. 1997). Elle repose donc sur le croisement de différentes déclarations faites par les sœurs d'une même fratrie. Au cours d'une première phase, les femmes ont été enquêtées sur la survie de leurs sœurs, puis elles ont fourni des informations permettant d'identifier et de retrouver les différentes sœurs survivantes. Dans une seconde phase, une grande majorité des sœurs identifiées ont été interrogées à leur tour, avec des questions quasiment identiques.

Il est ainsi possible de calculer pour chaque enquêtée l'écart entre *l'âge de cette femme tel qu'elle l'a déclaré* et *son âge tel qu'il a été déclaré par l'une de leurs sœurs* [30]. La moyenne de ces écarts absolus s'élèvant à deux ans et demi et leur écart-type à trois ans, on peut dire que les erreurs faites sur l'âge des sœurs restent assez faibles. Dans plus de soixante pour cent des déclarations, les écarts d'âge ne dépassent pas deux ans. En regroupant les âges par classe de cinq ans, nous obtenons une concordance entre les âges auto-déclarés et les âges déclarés par les sœurs, qui, sans être parfaite, n'est pas mauvaise [31].

Le graphe 6.12 présente l'évolution des erreurs d'âges faites sur l'âge des femmes (*proxy-report*), en fonction de l'âge de ces femmes tel qu'elles-

29. Cette enquête a été présentée dans la section 3.5.1, page 163.

30. Ces calculs sont menés sur le fichier comprenant l'appariement de Bicego et al. (1997). Nous sommes repartis de ce travail pour apparier également les déclarations qui portaient sur des sœurs n'ayant pas répondu à l'enquête. Au total, 13700 déclarations ont été appariées, essentiellement au départ du nom et prénom de la mère des enquêtées et de ceux de ses filles.

31. Le coefficient kappa est utilisé pour estimer cette concordance. Il permet de tenir compte de la "quantité d'accord" qui serait observée sous l'effet du hasard. Il varie entre 0 (concordance mauvaise) et 1 (concordance parfaite). Il s'élève ici à 0.493, ce qui correspond à une concordance "moyenne". Notons que nous sous-estimons ici l'ampleur des erreurs, car l'appariement est basé partiellement sur l'examen des écarts d'âge.

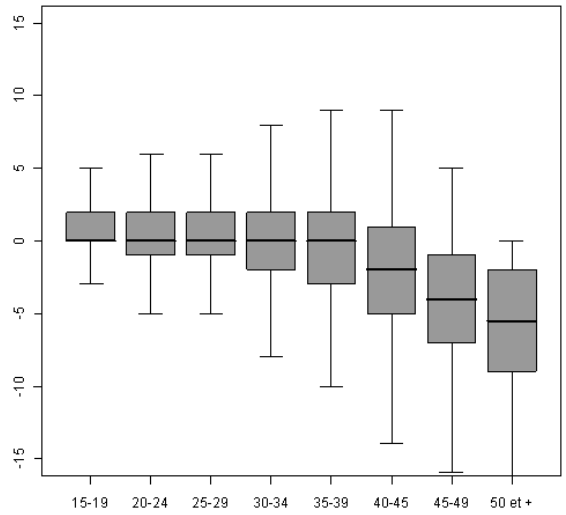

FIGURE 6.12 – *Ecarts d'âge entre les âges auto-déclarés et les âges déclarés par les sœurs enquêtées - nos calculs à partir de l'enquête SUMVE.*

mêmes le déclarent (*own-report*). Les problèmes d'erreurs d'âge concernent plus particulièrement les âges déclarés *pour* les femmes plus âgées (à partir de quarante ans). Une tendance au rajeunissement des femmes par leurs sœurs se dessine à mesure que ces femmes avancent en âge. Ceci devrait se traduire par de légers biais à la hausse sur les estimations de mortalité.

Il conviendrait de creuser davantage cette question, notamment pour chercher à savoir si les erreurs d'âge vont dans le même sens quand il s'agit de frères. Cela serait possible à partir des enquêtes où les questions sur la fratrie ont été posés aux hommes comme aux femmes.

6.2.6 Sous-déclaration des décès dans le passé lointain

Nous abordons maintenant le dernier test de qualité des données sur les frères et sœurs, et sans doute celui qui conduit aux corrections les plus importantes.

Du fait que certains pays ont organisé plusieurs enquêtes EDS successives et qui ne sont séparées que de quelques années, il est possible de comparer les niveaux de mortalité estimés *au sein de chaque enquête pour un même pays*, dans les cas où les périodes de référence se recoupent partiellement.

Timaeus et Jasseh (2004) ont été les premiers à procéder à cette analyse. Ils examinent la situation pour six enquêtes menées au Malawi (1992 et 2000), au Zimbabwe (1994 et 1999) et en Ouganda (1995 et 2001). Pour chacun de ces pays, ils introduisent dans leur modèle[32] une variable qui spécifie s'il s'agit de la *première* ou de la *seconde* enquête considérée. L'objectif est de détecter une éventuelle sous-déclaration des décès qui serait plus prononcée dans le passé lointain (par rapport aux années qui précèdent directement les enquêtes). Ceci est illustré dans le graphe 6.13.

32. Détaillé au chap.5, p.229.

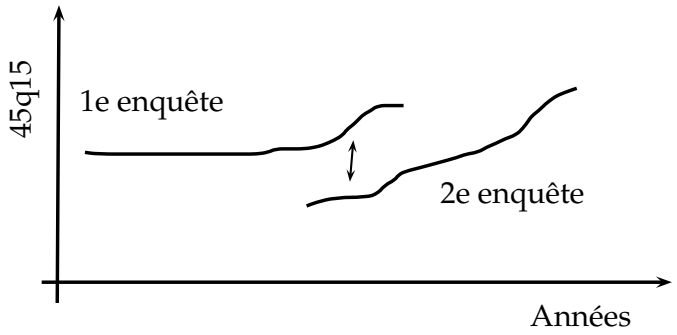

FIGURE 6.13 – *Exemple de sous-déclaration possible des décès dans le passé lointain*

A partir des résultats de l'ajustement de leur modèle, Timaeus et Jasseh (2004) tirent une conclusion peu encourageante :

> "The coefficients suggest that adult mortality in Malawi and Zimbabwe is underestimated by about 16% in the survey reports for *six to eight years before the second survey*, compared with *zero to two years before the first survey*. In Uganda the relative level of omissions in the more-distant reports from the second survey is 32%. These findings are worrying. If the completeness of reporting of dead siblings decays by about 20% as the interval since deaths increases by six years, even the most recent reports may be somewhat incomplete, which would imply that the absolute level of omissions is even higher". (p.762)

Lorsque Timaeus et Jasseh (2004) ont mené cette analyse, relativement peu de pays avaient déjà conduit plusieurs enquêtes EDS successives en collectant des données sur les frères et sœurs. Plus récemment, Obermeyer et al. (2010) ont adopté la même approche, afin d'évaluer comment la qualité des déclarations évolue à mesure que l'on remonte dans le temps [33]. Ils introduisent simplement la durée écoulé entre chaque observation et la date de l'enquête, et recodent cette durée en blocs de deux ans.

> "For males, the all-country coefficient is -0.0165, representing a 1.64% decrease in reported deaths for each additional year prior to the survey. For females, the annual decrease is lower at 1.09%." Obermeyer et al. (2010), p. 6

Cela signifie que huit ans avant l'enquête, les décès masculins seraient donc sous-estimés d'environ 12% de plus que directement avant l'enquête [34]. La sous-déclaration correspondante des décès féminins se chiffrerait à 8% de plus que directement avant l'enquête [35]

Nous avons inclus une variable du même type dans le modèle de régression appliqué à l'ensemble des données EDS africaines sur la survie des frères et sœurs (voir le chapitre suivant, et Reniers, Masquelier, et Gerland (forthcoming)). La valeur de cette variable est calculée comme *la*

33. Leur modèle a été détaillé au chap.5, p.233.

34. En effet : $1 - 1 \times exp(-0.0165)^8 = 0.1236$.

35. La sous-déclaration paraît donc moindre que celle observée par Timaeus et Jasseh (2004) au Malawi, au Zimbabwe et en Ouganda. Ceci s'explique sans doute par le fait que Obermeyer et al. (2010) utilisent également des données relatives à d'autres régions en développement - or la qualité des données africaines est généralement moins bonne.

différence en années entre le moment de l'enquête et la période à laquelle l'observation fait référence. Dans la moitié des pays, cette variable ne peut être calculée (soit parce qu'il n'existe qu'une enquête comprenant des données sur la survie des frères et sœurs, soit parce que les différentes enquêtes sont séparées de plus de 9 ans). Seuls 16 pays sur les 31 sont retenus dans les analyses [36]. Cette variable est entrée sous forme catégorielle, en retenant comme référence la période entre 1 et 2 ans (exacts) avant l'enquête. Le graphe 6.14 présente l'exponentielle des coefficients de cette variable, ce qui procure des rapports de taux par rapport à cette référence. Ce graphe présente en fait trois modèles différents : le premier est estimé en retenant *toutes* les observations (à gauche), le second lorsque nous retenons les données sur les *sœurs* uniquement (au centre) et le troisième lorsque nous retenons les données sur les *frères* uniquement (à droite). Les points correspondent à l'exponentielle des coefficients, et les barres en traits pleins aux intervalles de confiance à 95%.

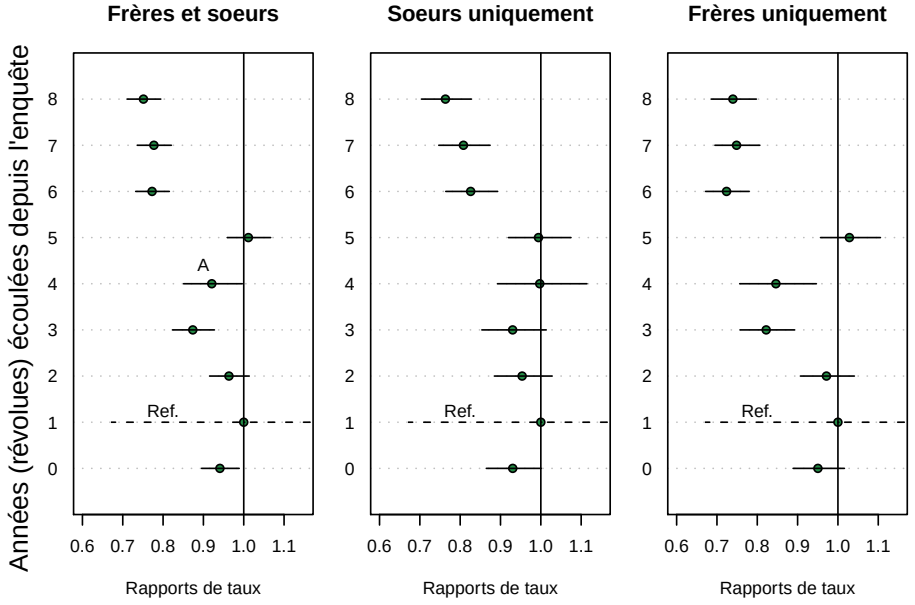

FIGURE 6.14 – *Estimation de la sous-déclaration des décès des frères et sœurs selon le temps écoulé depuis l'enquête (en années)*
LECTURE : Le point A indique que par rapport à la référence (entre 1 et 2 ans exacts avant l'enquête), les décès des frères et sœurs survenus entre 4 et 5 ans exacts sont sous-déclarés d'environ 10%.

Nous pouvons faire les constats suivants :

1. Par rapport à la référence (entre 1 et 2 ans exacts avant l'enquête), l'année qui précède directement l'enquête est caractérisée par une légère sous-déclaration des décès, quel que soit le sexe des frères et sœurs. Cette différence se chiffre à 6% environ sur l'ensemble des

36. Le Burkina Faso, le Cameroun, l'Ethiopie, la Guinée, le Kenya, Madagascar, le Malawi, le Mali, le Mozambique, la Namibie, l'Ouganda, le Rwanda, la Tanzanie, le Tchad, la Zambie et le Zimbabwe. Pour cette raison, le modèle est ajusté en combinant tous les pays - c'est-à-dire qu'il n'est pas ajusté par région comme nous le faisons pour estimer les taux de mortalité.

frères et sœurs. Ceci est sans doute dû à un déplacement de certains décès très récents vers le passé.

2. Entre 3 et 5 ans avant l'enquête, les décès *masculins* semblent sous-déclarés, de 10 à 13% environ, et cette sous-déclaration est significative. Chez les sœurs, on observe également une sous-déclaration, mais non significative.
3. Les décès ne semblent pas sous-déclarés quand ils surviennent 5 ans avant l'enquête. Ceci reflète certainement une attirance sur 5 (en réponse à la question sur le nombre d'années écoulées depuis le décès).
4. Ensuite, la qualité des informations chute rapidement, et la sous-déclaration va dépasser les 20% à partir de 6 ans avant l'enquête.
5. Ces chiffres montrent également que, dans l'ensemble, la sous-déclaration des décès est un peu plus importante pour les frères que pour les sœurs.

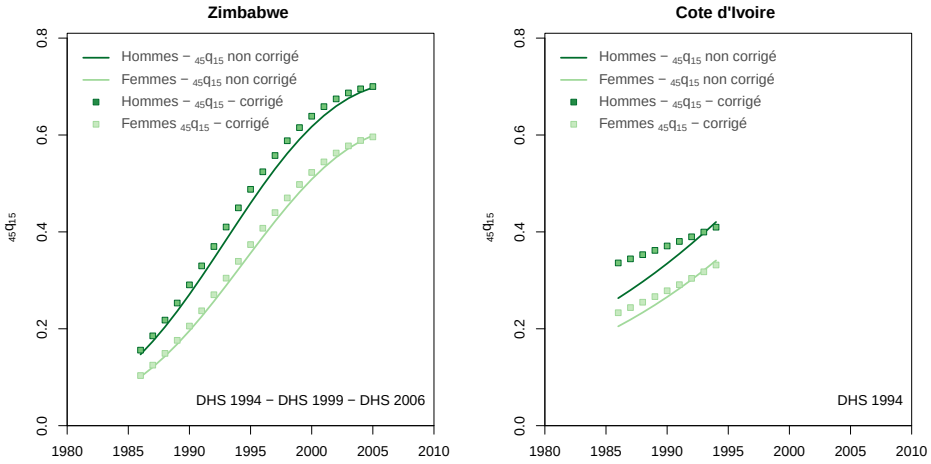

FIGURE 6.15 – *Quotients de mortalité de 15 à 60 ans calculés à partir de la survie des frères et sœurs et du modèle de Timaeus et Jasseh (2004)- Zimbabwe et Côte d'Ivoire : comparaison des estimations ajustées et non-ajustées.*

A partir de ces estimations, nous pouvons multiplier les décès survenus dans le passé lointain par l'inverse du coefficient de la variable TIPS pour leur donner plus d'importance. Nous ajustons ensuite un nouveau modèle, avec les décès re-pondérés. Cette correction sera appliquée à toutes les enquêtes, en supposant que l'ampleur de la sous-déclaration est constante par âge et par pays [37]. Ceci aura pour effet de relever les niveaux de mortalité dans le passé. Sans cette correction, nous serions amenés à surestimer l'ampleur des reprises de mortalité. On l'observe avec le cas de la Côte d'Ivoire dans le graphe 6.15 [38].

37. Il semble qu'Obermeyer et al. (2010) fixent le coefficient de la variable "TIPS" à zéro pour calculer les probabilités prédites de décès et générer des estimations à partir de leur modèle de régression. Nous ne savons pas si ceci a pour effet de ne corriger que les estimations pour les pays où ce coefficient peut être calculé ou si ces auteurs corrigent toutes les estimations.

38. Les estimations semblent davantage 'corrigées' pour ce pays que ce n'est le cas pour

Conclusion

L'impression générale qui se dégage de ce chapitre est celle d'une *qualité des données très imparfaite*, avec notamment de sérieux problèmes de sous-déclaration des décès, ainsi que des difficultés à collecter des informations qui se réfèrent bien aux parents *biologiques* et aux frères et sœurs *nés de même mère*.

En matière de survie des parents, le biais d'adoption constitue le problème principal. Son ampleur est telle que les estimations obtenues ne peuvent souvent que servir que de "bornes inférieures". Dans le cas des frères et sœurs, de nombreux éléments peuvent expliquer que les informations collectées soient approximatives, erronées ou incomplètes. Le contact n'a pas nécessairement été maintenu entre frères et sœurs ne cohabitant plus ensemble. La qualité de ce contact entre frères et sœurs va d'ailleurs varier par sexe et en fonction des contextes ; âge à la décohabitation parentale, migrations de travail, réallocation des épouses dans la famille du mari dans certaines sociétés patrilocales, etc. Compte tenu de la taille des fratries, certains naissances et décès sont survenus loin dans le passé, et un certain nombre de frères et sœurs ont pu mourir avant la naissance des enquêtés - ou durant leur enfance (Hill et Trussell 1977). Précisons également qu'il est possible que l'enquêté s'inclue lui-même dans la liste de ses frères et sœurs, en raison d'une erreur de compréhension qui peut concerner tant l'enquêté que l'enquêteur. Dans ce cas, un biais à la baisse sera introduit dans les estimations. Les enquêtés peuvent en outre préférer garder secrets certains décès, notamment ceux qui sont peu acceptables socialement (mort violente, VIH-sida,...). On peut même imaginer qu'un biais d'adoption comparable à celui observé pour les parents puisse jouer un rôle : il est possible que certains enquêtés se situent dans une fratrie "d'adoption" qui n'est pas celle formée par les frères et sœurs nés de leur mère biologique.

L'impact de ces différentes erreurs sur les estimations n'est pas évident. Inclure un cousin ou omettre un frère dans les déclarations ne va pas se traduire nécessairement par une sur- ou sous-estimation de la mortalité, car tout dépend des âges des personnes omises ou ajoutées, et des risques de décès auxquels elles font face. Et encore, la sous-déclaration des frères et sœurs et de leurs décès n'affecte pas les estimations si ces frères et sœurs n'ont jamais atteint l'âge adulte. S'il s'agit de décès adultes, cette sous-déclaration aura un impact sur l'estimation indirecte. Par contre, elle n'affectera les estimations directes que si les décès non-déclarés sont des décès récents. Cela étant, il est clair qu'il existe une sous-déclaration importante des décès survenus dans un passé même relativement récent. A cette sous-déclaration des décès s'ajoutent des problèmes au niveau des âges à l'enquête et des âges au décès (qui méritent un examen plus attentif).

Ces différents problèmes de qualité signifient notamment qu'il est bien

le Zimbabwe, mais cela tient au fait que les estimations ivoiriennes ne sont basées que sur une enquête, tandis que les décès et les personnes-années calculés pour le Zimbabwe combinent trois enquêtes différentes.

plus prudent de s'en tenir aux quelques années avant l'enquête, et de ne pas exploiter les informations qui font référence à un passé trop lointain. Quelques auteurs se sont pourtant risqués à des analyses sur de longues périodes. Ainsi, à partir des frères et sœurs, de Walque (2005) estime des taux de mortalité à partir de l'enquête EDS menée au Cambodge en 2000, se concentrant sur la période khmère de 1975-79, soit 25 ans avant l'enquête. Obermeyer et al. (2008) analysent également la mortalité due à la violence dans 13 pays à partir de données WHS sur les frères et sœurs, en poussant l'analyse jusqu'à 1965-1974 au Zimbabwe et 1955-1964 au Vietnam, alors que les enquêtes ont été menées en 2002/03 [39]. Ce type d'analyse nous paraît tout à fait surestimer le potentiel des données sur la survie des proches. Si la sous-déclaration des décès peut atteindre 20% dans les 10 ans qui précèdent l'enquête, on peut difficilement imaginer que les données collectées à partir des années 2000 puissent nous fournir des informations fiables sur ce qui s'est passé avant 1980, 20 à 30 ans plus tôt.

Nous avons pu montrer également que *les problèmes de données ne se situent pas toujours là où on les attend*. Par exemple, les rapports d'enquête EDS nous paraissent sous-estimer l'ampleur des réponses manquantes. A l'inverse, l'examen de la taille des fratries par âge ne conduit pas toujours à la mise en évidence d'une sous-déclaration des frères et sœurs par les enquêtés âgés. Nous l'avons illustré en recourant à une comparaison avec le nombre d'enfants nés vivants.

39. Ils notent : "The salience of the WHO survey data ... might indicate that war deaths are less subject to recall bias than other deaths, perhaps because of the young age of victims and strong emotional associations for family members", p.6

7 Estimations nationales basées sur la survie des proches

Sommaire

7.1 Afrique de l'Ouest 320
7.2 Afrique Centrale 326
7.3 Afrique de l'Est 329
7.4 Afrique australe 334
Conclusion 340

Ce dernier chapitre est consacré à la présentation des niveaux de mortalité adulte dégagés de la survie des parents et des frères et sœurs. L'objectif consiste à montrer que ces données, malgré les nombreuses limites dont nous avons déjà longuement discuté, peuvent améliorer grandement notre connaissance de la mortalité adulte africaine.

Nous combinons ici les différents types de données (survie des frères, des sœurs et des parents), les différentes sources (enquêtes et recensements), et les différentes méthodologies (exploitation directe et indirecte, cohortes synthétiques, etc). Ces estimations empiriques sont confrontées à celles issues de la modélisation, c'est-à-dire celles qui sont fournies par les Nations Unies et l'Organisation mondiale de la santé.

Rappelons que pour quatre pays africains sur cinq, les estimations des Nations Unies sont toujours obtenues en combinant des niveaux de mortalité infanto-juvénile et les tables de Princeton (*cf.* encadré 4, p.66). L'OMS utilise plutôt un modèle relationnel appelé le "système des logits modifiés" (inspiré du modèle de Brass). Celui-ci est construit à partir d'un panel de données plus large que les tables de Princeton, puisqu'il repose sur 1802 tables de mortalité issues de 64 pays du monde entier (Murray et al. 2003). Cela étant, il s'agit toujours de pays à mortalités modérées ou faibles, dans laquelle l'Afrique n'est pas bien représentée. Avec ce système des logits modifiés, l'OMS estime les risques de décès adultes à partir de

ceux observés chez les enfants [1]. Les seuls pays africains pour lesquels le système des logits intègre directement des niveaux de mortalité adulte sont l'Afrique du Sud, le Zimbabwe et la Tanzanie (Bradshaw et Timaeus 2003). Ensuite, les risques de décès sont revus à la hausse pour intégrer la surmortalité causée par le VIH-sida ou par les conflits.

Avant de comparer nos estimations à celles des Nations Unies et de l'OMS, il est utile de souligner que ces deux agences internationales fournissent des niveaux de mortalité qui diffèrent sensiblement. Ceci est discuté dans l'encadré 11.

Encadré 11 Comparaison des estimations de l'ONU et de l'OMS

Les graphes 7.1 et 7.2 opposent, pour chaque sexe, les probabilités de décès adultes ($_{45}q_{15}$) estimées par les nations Unies (en abscisses) et les mêmes probabilités estimées par l'OMS (en ordonnées). Les différentes estimations pour les hommes et pour les femmes se réfèrent aux années 1990, 2000 et 2006 pour 50 pays subsahariens [2]. Afin de voir si les divergences entre les deux séries sont liées à une modélisation du sida, les points qui correspondent à une année où la prévalence nationale du VIH était supérieure à 3% sont repris en plus fonçé.

Il est frappant d'observer l'ampleur des divergences entre les estimations des deux agences qui pourtant partent des mêmes probabilités de survie des enfants [3]. Il ne s'agit pas de différences minimes, puisqu'on trouve même des estimations qui diffèrent de plus de 40%.

Pour tenter d'expliquer ces écarts, nous avons calculé des coefficients de corrélations entre les deux séries selon la prévalence atteinte, selon le sexe et selon l'année considérée (tableau 7.1). Globalement, les estimations de l'ONU et de l'OMS sont davantage corrélées pour 2006 et 2000 que pour 1990, quel que soit le sexe ou la prévalence atteinte. Les désaccords sur les niveaux de mortalité féminine semblent également un peu moins prononcés. Mais il ressort surtout du tableau 7.1 que les écarts les plus importants concernent les estimations où la prévalence est faible. C'est donc vraisemblablement les niveaux de mortalité "sans sida" qui cadrent le moins les uns avec les autres.

En fait, on y regardant de plus près, on peut isoler trois sous-groupes de points dans le graphe 7.1 :

1. "Adult mortality rates were derived from either surveys or censuses where available ; otherwise the most likely corresponding level of adult mortality was estimated based on regression models of child versus adult mortality as observed in the set of approximately 1800 life tables. These estimated child and adult mortality rates were then applied to a global standard, defined as the average of all the life tables, using the modified logit model to derive the estimates for the reference years." OMS, *The World Health Statistics 2010*. Geneva : WHO, 2010. (http ://www.who.int/healthinfo/statistics/).

2. Les estimations des Nations Unies sont tirées de UN (2009d). Nous interpolons linéairement entre les valeurs quinquennales pour obtenir des estimations pour 1990, 2000 et 2006. Les estimations OMS sont téléchargées depuis le site http ://www.who.int/whosis/en/ (visité en août 2010).

3. Les deux organisations participent à l'*Inter-agency Group for Child Mortality Estimation* (IGME) (qui reprend l'UNICEF, l'OMS, la Banque Mondiale et la Division de la Population).

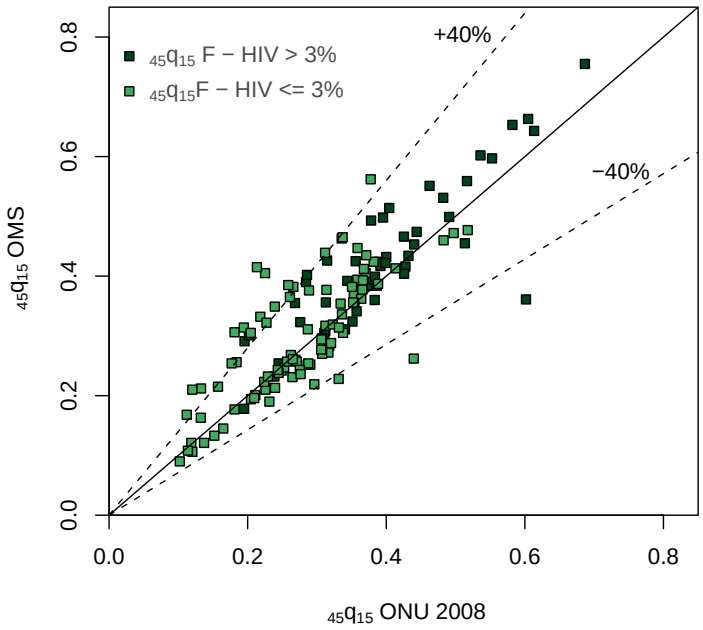

FIGURE 7.1 – *Probabilités de décès **féminins** de 15 à 60 ans estimés par les Nations Unies et par l'OMS pour 1990, 2000 et 2006, selon le niveau de la prévalence atteinte à ces dates - 50 pays africains.*

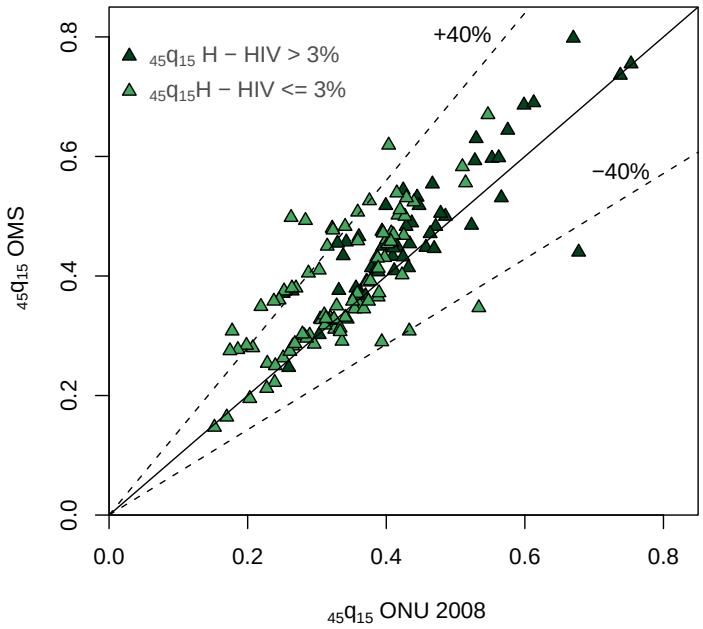

FIGURE 7.2 – *Probabilités de décès **masculins** de 15 à 60 ans estimés par les Nations Unies et par l'OMS pour 1990, 2000 et 2006, selon le niveau de la prévalence atteinte à ces dates - 50 pays africains.*

		Année		
		1990	2000	2006
HIV <= 3%	Mortalité féminine	0.76	0.79	0.79
	Mortalité masculine	0.68	0.79	0.72
HIV > 3%	Mortalité féminine	0.75	0.90	0.91
	Mortalité masculine	0.74	0.90	0.92

TABLE 7.1 – *Coefficients de corrélation entre les ${}_{45}q_{15}$ estimés par l'OMS et ceux estimés par les Nations Unies, selon le sexe, l'année et le niveau de la prévalence du VIH-sida.*

1. Le premier regroupe la majorité des valeurs du quotient ${}_{45}q_{15}$ autour de la diagonale. Dans ces cas-là, les deux agences procurent donc des niveaux de mortalité comparables. On note cependant qu'à mesure que la mortalité augmente (au delà de 0.4), l'OMS tend à fournir des estimations supérieures à celles des Nations Unies.
2. Le second groupe d'estimations rassemble quelques points où l'ONU procure des niveaux inférieurs d'environ 40% à ceux de l'OMS. Il s'agit en fait des cas du Rwanda, et de l'Erythrée, deux pays aux tendances de mortalité très perturbées[4].
3. Le troisième groupe d'estimations pose davantage question. *A des niveaux de mortalité compris entre 0.1 et 0.4 correspond une série de valeurs de ${}_{45}q_{15}$ qui se concentrent sur une droite où les niveaux de mortalité de l'OMS sont supérieurs de 40% à ceux des Nations Unies.* Ceci concerne tant les hommes que les femmes, et il s'agit essentiellement de situations où la prévalence est faible.

Le fait que les écarts entre ces estimations se concentrent autour de 40% indique que les divergences entre les deux séries d'estimations sont systématiques. En d'autres termes, leur source est à chercher au niveau des méthodes, et non des données. Qui plus est, la mortalité des enfants, qui sert de point de départ aux deux séries, est censée être la même. C'est donc vraisemblablement le choix des tables de mortalité qui joue ici un rôle important. Cela est confirmé par l'examen du graphe 7.3, qui reprend uniquement les estimations qui correspondent à des prévalences faibles. Ce graphe distingue les estimations qui ont été calculées par la Division de la Population par le biais du modèle *Nord* de Princeton et celles qui sont obtenues par le biais du modèle *Sud*. Ces deux modèles sont en fait utilisés dans 7 pays sur dix par les Nations Unies (tandis que le modèle Ouest concerne seulement une dizaine de pays). Ce graphe permet de pointer une des principales sources de divergences entre les estimations de l'ONU et de l'OMS. *Toutes les valeurs du quotient ${}_{45}q_{15}$ qui se situent sur la droite des 40% sont en fait calculées par l'ONU par le biais du modèle Sud.* Ce dernier se caractérise par une moindre mortalité adulte par rapport à la mortalité infanto-juvénile. Pour un même quotient ${}_5q_0$, il va donc procurer des niveaux nettement moins élevés que ceux prévus par le modèle Nord ou par l'OMS.

Ceci illustre l'importance déterminante du choix du standard de mortalité. Ces écarts témoignent également de la difficulté à modéliser la mortalité adulte *à partir de la survie des enfants*, et par là, ils montrent l'importance des données sur la survie des adultes.

4. Ces écarts pourraient donc résulter du fait que nous avons interpolé les estimations de l'ONU entre les intervalles quinquennaux.

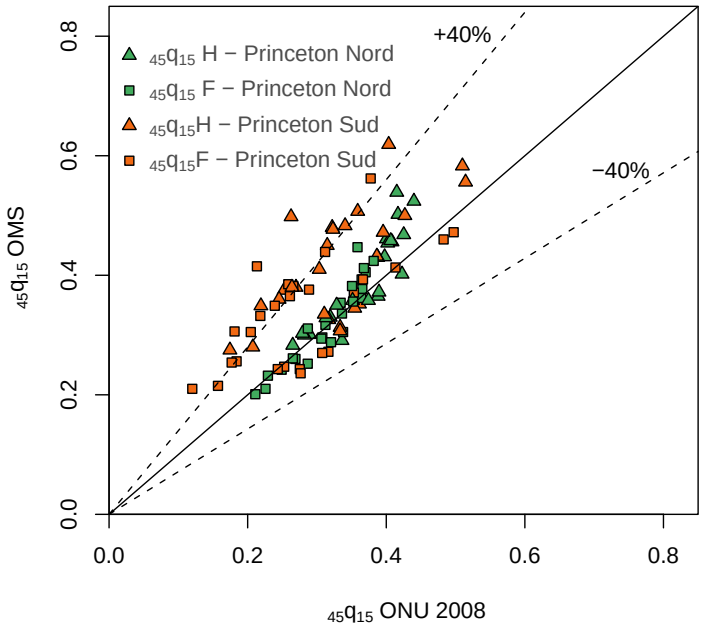

FIGURE 7.3 – *Probabilités de décès de 15 à 60 ans estimés par les Nations Unies et par l'OMS pour 1990, 2000 et 2006, dans le cas où la prévalence du VIH est inférieure à 3%, selon la table de mortalité retenue.*

Dans les sections qui suivent, nous proposons une comparaison systématique - au niveau national - entre différentes séries d'estimations issues des enquêtes EDS et des recensements.

1. Les deux premières séries sont les quotients $_{45}q_{15}$ fournis par *les Nations Unies et l'OMS*. Il s'agit des valeurs publiées par intervalles quinquennaux dans le cas des Nations Unies[5], et des estimations pour 1990, 2000 et 2006 dans le cas de l'OMS[6].
2. La troisième série d'estimations est basée sur les *données relatives aux frères et sœurs*, que nous avons exploitées de façon *biographique*, en utilisant le modèle de régression proposé par Timaeus et Jasseh (2004) (p. 229). Les tableaux de synthèse 7.2 et 7.3 regroupent les différentes estimations nationales que nous avons pu dégager de ces informations sur les fratries, pour 1990, 1995, 2000 et 2005. Nous les commenterons plus loin par pays. Rappelons que les données ne sont pas pondérées pour tenir compte des biais de sélection, et qu'aucune correction n'est apportée pour les fratries sans survivants, car ceci nous paraît être la meilleure option à ce stade (*cf.* chapitre 5). Nous supposons que les risques de décès adultes ne varient pas avec la taille des fratries, et donc que les différents biais de sélection s'annulent. Par contre, nous ajustons les données pour corriger la tendance observée

5. A partir de 1980 pour les pays affectés par le VIH-sida et à partir de 1995 pour les autres.
6. *Cf.* note 2, p.312.

à sous-déclarer les décès survenus dans les périodes plus anciennes (p.304).

	1990	1995	2000	2005
Afrique du Sud	0.102	0.171		
Benin	0.208	0.213	0.222	0.237
Burkina Faso		0.283	0.281	
Cameroun	0.222	0.275	0.321	
Centrafrique	0.364			
Congo			0.404	0.249
Congo RDC			0.29	0.274
Cote d' Ivoire	0.278			
Ethiopie		0.391	0.322	0.221
Gabon		0.279	0.293	
Guinee		0.249	0.295	
Kenya	0.192	0.279	0.326	0.311
Lesotho			0.454	
Madagascar	0.294	0.269	0.24	0.211
Malawi	0.28	0.468	0.545	
Mali	0.24	0.25	0.26	0.272
Mozambique	0.173	0.21	0.254	
Namibie	0.16	0.244	0.368	
Niger	0.25	0.238	0.227	0.217
Nigeria			0.303	0.272
Ouganda	0.374	0.418	0.421	0.382
Rwanda		0.537	0.39	
Senegal	0.198	0.199	0.2	
Sierra Leone			0.234	0.267
Swaziland			0.505	0.625
Tanzanie	0.223	0.278	0.328	
Tchad	0.248	0.253	0.258	
Togo	0.203	0.246		
Zambie	0.315	0.537	0.609	0.495
Zimbabwe	0.206	0.374	0.523	0.596

TABLE 7.2 – *Quotients ${}_{45}q_{15}$ estimés pour la mortalité **féminine** à partir de la survie des sœurs - nos calculs à partir des enquêtes EDS en retenant le modèle de régression de Timaeus et Jasseh (2004) (valeurs corrigées pour la sous-déclaration des décès survenus dans le passé lointain)*

3. Les niveaux de mortalité obtenus *indirectement à partir des frères et sœurs* sont présentés également. Ils sont calculés à partir des coefficients de Timaeus et al. (2001) sur les données EDS. Les différentes probabilités de survie sont converties en ${}_{45}q_{15}$ par le biais du standard de Brass. Cette question du choix du standard est approfondie dans l'encadré 12.
4. La cinquième série d'estimations reprend les niveaux de mortalité calculés à partir de la *survie des mères* qui est déclarée dans les enquêtes EDS par les enfants de 5 à 9 ans. Nous utilisons les coefficients de Timaeus (1992) et la méthode de datation de Brass et Bamgboye

	1990	1995	2000	2005
Afrique du Sud	0.285	0.367		
Benin	0.302	0.3	0.304	0.314
Burkina Faso		0.348	0.335	
Cameroun	0.3	0.364	0.419	
Centrafrique	0.464			
Congo			0.558	0.308
Congo RDC			0.362	0.336
Cote d' Ivoire	0.371			
Ethiopie		0.477	0.376	0.248
Gabon		0.401	0.418	
Guinee		0.273	0.328	
Kenya	0.212	0.321	0.391	0.392
Lesotho			0.606	
Madagascar	0.361	0.327	0.296	0.27
Malawi	0.308	0.51	0.596	
Mali	0.247	0.275	0.308	0.344
Maroc	0.149	0.13	0.113	
Mozambique	0.254	0.282	0.315	
Namibie	0.278	0.382	0.518	
Niger	0.256	0.238	0.222	0.208
Nigeria			0.34	0.296
Ouganda	0.479	0.534	0.544	0.505
Rwanda		0.768	0.517	
Senegal	0.242	0.239	0.237	
Sierra Leone			0.256	0.309
Swaziland			0.625	0.684
Tanzanie	0.317	0.373	0.418	
Tchad	0.274	0.286	0.299	
Togo	0.259	0.303		
Zambie	0.359	0.593	0.666	0.549
Zimbabwe	0.291	0.488	0.639	0.7

TABLE 7.3 – *Quotients ${}_{45}q_{15}$ estimés pour la mortalité **masculine** à partir de la survie des frères - nos calculs à partir des enquêtes EDS en retenant le modèle de régression de Timaeus et Jasseh (2004) (valeurs corrigées pour la sous-déclaration des décès survenus dans le passé lointain)*

(1981). Nous procédons de la même façon à partir des données d'une *quinzaine de recensements*[7]. Ces données censitaires permettent de retracer de plus longues tendances.

Pour les estimations issues des enfants, nous utilisons plutôt les coefficients corrigés de Timaeus et Nunn (1997), lorsque la prévalence au moment de la naissance de ces enfants atteint 5%. Les proportions d'orphelins sont alors ajustées à la baisse (comme détaillé dans le chapitre 4, p. 198).

5. A partir des enquêtes EDS et des recensements, nous recourons également à la méthode de Preston, basée sur les *cohortes synthétiques*, pour exploiter les données relatives aux parents[8]. Ceci constitue donc une sixième série d'estimations. Cette méthode nous permet de faire intervenir la mortalité des pères, qu'il n'est pas possible d'analyser à partir d'une seule enquête[9].
6. Nous avons vu qu'aucune méthode ne permet à ce jour d'exploiter les données sur les pères dans les contextes perturbés par le VIH-sida. Ceci a motivé le développement de nouveaux coefficients (*cf.* chapitre 4) qui font intervenir les prévalences passées. Nous appliquons ici ces coefficients sur les proportions de pères survivants des enfants de 5 à 9 ans observées dans les EDS, en recourant toujours à la méthode des cohortes synthétiques, et dans les cas où la prévalence du VIH atteint 5% au moins. Dans les autres cas, nous utilisons les coefficients plus classiques, fournis par Timaeus (1992) pour la mortalité paternelle[10].

Lorsqu'elles sont disponibles, les estimations rassemblées par Timaeus (1993), issues des données de recensements et des données d'enquête, sont présentées également. Elles fournissent un aperçu des niveaux observés de façon indirecte durant les années 1970 et 1980.

Encadré 12 Quel(s) schéma(s) retenir pour pouvoir comparer les différentes estimations indirectes ?

L'indice de mortalité adulte le plus utilisé est le quotient de mortalité de 15 à 60 ans, même s'il est très influencé par les décès qui surviennent au dessus de 30-35 ans, et qu'il est de ce fait moins sensible aux changements rapides de mortalité qui affectent les jeunes adultes (Bradshaw et Timaeus 2003). Le problème de ce quotient est qu'aucune méthode indirecte ne procure des probabilités de décès qui couvrent l'entièreté de l'intervalle de 15 à 60 ans. C'est toutefois celui que nous avons retenu ici, pour faciliter les comparaisons avec d'autres estimations (notamment celles de l'OMS).

---

7. Minnesota Population Center. Integrated Public Use Microdata Series, International : Version 6.0. Minneapolis : University of Minnesota, 2010.
https ://international.ipums.org/international/.

8. Il s'agit des points "Orph. (cohortes) - EDS", et "Orph. (cohortes) recensements".

9. Rappelons que pour dater les estimations, nous devons disposer de la proportion ${}_{10}S_n$. Or la seule qu'il est possible de calculer à partir des enquêtes est celle basée sur les 0-4 ans (${}_{10}S_5$). Mais il n'est pas possible de calculer la probabilité de survie correspondante (${}_5p_{35}$).

10. Cette fois à partir des 10-14 ans.

Nous avons vu qu'il est très délicat de convertir les probabilités de survie obtenues indirectement ($_np_{15}$, $_np_{25}$, etc.) en indices du type $_{45}q_{15}$. Il s'agit d'une opération où le *choix du standard de mortalité* peut avoir un impact déterminant, surtout dans les contextes perturbés par le sida. C'est donc un des aspects de l'estimation indirecte où il faut faire usage d'une certaine dose de bon sens... et il faut bien l'admettre, les choix relèvent également un peu de l'arbitraire.

Une des difficultés tient au fait qu'il n'existe actuellement pas de standard qui permet de passer de façon souple d'un schéma par âge "classique" à un schéma "perturbé" par le VIH-sida. Et si l'on veut par exemple utiliser *d'abord* le standard de Brass (pour les périodes où la prévalence du sida est faible), puis opter *ensuite* pour un schéma tel que le standard INDEPTH-2 (caractérisé par des risques de décès très élevés aux premiers âges adultes), des problèmes de continuité vont se poser au moment du passage d'un standard à l'autre. En d'autres termes, on observera des "cassures" dans les tendances, et parfois même des baisses de mortalité là où celle-ci est sensée augmenter. Il serait donc très utile de pouvoir disposer dans le futur de schémas de mortalité par âge qui tiendraient compte de l'excès de mortalité associé au VIH-sida et que l'on pourrait faire évoluer dans le temps (au moyen d'un système logit, par exemple) en intégrant des informations sur la prévalence du VIH.

Pour éviter ces perturbations dans les tendances, nous avons utilisé pour quasiment toutes les estimations indirectes présentées ici (orphelins, frères et sœurs) le standard de Brass afin d'obtenir une valeur de $_{45}q_{15}$. Une exception est toutefois faite pour les probabilités de décès de type $_{10}q_{25}$ et $_{15}q_{25}$ qui sont issues de la *survie des parents* dans les EDS, pour les pays fortement affectés par le VIH-sida [11]. Dans ces cas-là, nous retenons plutôt le standard INDEPTH-2, ce qui a pour effet de diminuer les niveaux de mortalité observés (*cf.* p.120) [12].

Sans revenir sur les détails, rappelons que le choix du standard n'a pas le même impact selon que l'on exploite les données sur les *parents* ou sur les *frères et sœurs*. Cela est dû au fait que les probabilités de survie ne sont pas estimées sur les mêmes intervalles d'âges (*cf.* chapitre 3, p.120 et chapitre 5, p. 274).

11. Plus précisément là où la prévalence du VIH-sida s'élève à plus de 5% (à la date correspondante à l'estimation).

12. Sans cela, nous obtenons des quotients $_{45}q_{15}$ qui s'avèrent beaucoup trop élevés et nous surestimons les hausses de mortalité dues au sida.

7.1 Afrique de l'Ouest

L'Afrique de l'Ouest se caractérise par les niveaux de mortalité adulte les plus faibles de l'Afrique subsaharienne. Ce sont également ceux qui sont restés les plus stables durant les dernières décennies, après une baisse rapide de la mortalité adulte entre les années 1960 et 1980 (Timaeus 1993).

Cette stabilité s'observe notamment à partir les tendances estimées pour le *Bénin*, le *Burkina Faso*, le *Mali* ou encore le *Sénégal* à partir des données sur les frères et sœurs exploitées de façon biographique. Ces différentes tendances sont présentées pour 10 pays et par sexe dans le graphe 7.4. Les pages suivantes détaillent davantage les estimations obtenues pour le Mali, le Niger et le Sénégal. Tous les autres pays de la sous-région pour lesquels nous disposons d'au moins une source de données empiriques sont repris en annexe A.7. Dans les graphes qui présentent les situations nationales, une droite verticale (en pointillés) indique l'année au cours de laquelle la prévalence du VIH a atteint au moins 5%.

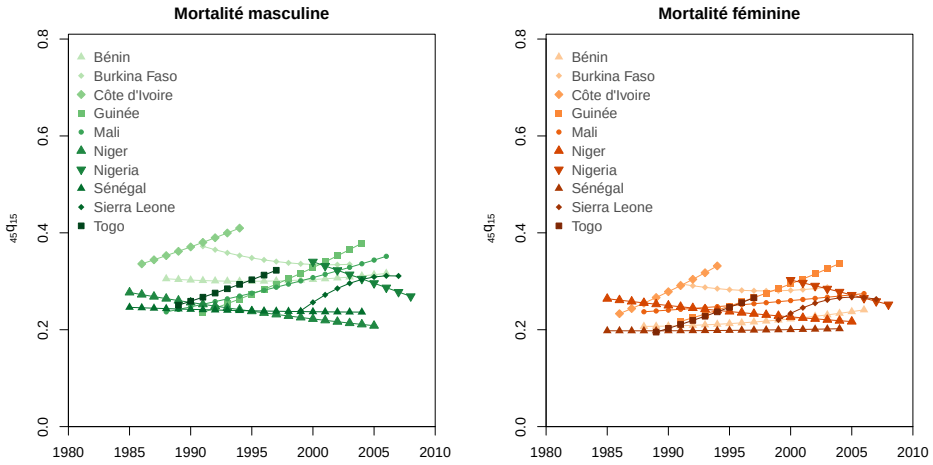

Figure 7.4 – *Quotients de mortalité de 15 à 60 ans calculés à partir de la survie des frères et sœurs et du modèle de Timaeus et Jasseh (2004)*

Relevons simplement ici les principaux enseignements de ces estimations nationales.

Au *Bénin* (p. 457), la mortalité semble avoir légèrement baissé depuis le début des années 1980, même si la survie des mères et celle des sœurs laissent entrevoir une légère reprise à partir des années 2000, que d'autres estimations devraient venir confirmer.

La *Côte d'Ivoire* se distingue clairement de ses voisins par des risques de mortalité plus élevés, et par une prévalence du VIH-sida elle aussi plus élevée. Cette dernière était estimée à plus de 5% à la fin des années 1990 (UNAIDS 2009), alors qu'elle n'a jamais dépassé ce seuil dans les autres pays de la sous-région. Il faut toutefois être prudent avec des tendances dégagées d'une seule enquête, notamment parce que le timing de cette reprise de mortalité n'est pas corroboré par les autres estimations. Cette hausse survient dès la fin des années 1980 si l'on en croit les données sur

les frères et sœurs, mais ceci ne cadre pas avec les estimations onusiennes, ni avec la survie des parents (graphe A.32, p. 463) [13].

La cas de la *Guinée* pose également question, puisque les estimations calculées en biographique et de façon indirecte à partir des frères et sœurs indiquent une reprise de mortalité durant la seconde partie des années 1990, ce qui ne correspond pas du tout avec les tendances esquissées par les Nations Unies ou l'OMS (p.467). Il pourrait d'agir là d'un problème de qualité des données [14]. C'est vraisemblablement le cas également au *Nigeria*, où la mauvaise qualité de l'enquête de 1999 nous invite à rester prudent (et à ne pas prendre pour acquis les niveaux de mortalité déduits des proches). Il en va de même pour la Sierra Leone, où nous avons vu que l'enquête de 2008 se caractérise par une proportion très importante de réponses manquantes sur les frères et sœurs.

Au niveau méthodologique, ce sont surtout *les pays sahéliens* qui retiennent notre attention. Les graphes repris en pages 323 à 325, présentent les cas du *Mali*, du *Niger* et du *Sénégal*. Ils montrent trois situations où les niveaux obtenus à partir de la survie des proches sont nettement *en-deçà* de celles de l'ONU et de l'OMS (bien que la tendance soit comparable). Précisons ici que pour ces trois pays, les Nations Unies utilisent le modèle Sud des tables de Princeton. Le recours au modèle Nord aurait pour effet d'accentuer les écarts entre les estimations onusiennes et celles issues des frères et sœurs.

Les divergences sont particulièrement prononcées au *Niger* : le quotient ${}_{45}q_{15}$ calculé en biographique à partir des sœurs s'élève à 0.24 en moyenne, ceux estimés par l'ONU avoisinent 0.33 et ceux fournis par l'OMS 0.49 !

Au *Mali*, les deux agences procurent des estimations comparables, mais elles restent sensiblement plus élevées que celles que nous obtenons à partir des enquêtes EDS. Elles cadrent davantage avec l'estimation fournie pour 1986 par Timaeus (1993), à partir de la survie des parents et des décès déclarés dans le recensement de 1987. On notera par ailleurs que les tendances récentes issues des enquêtes prolongent celle esquissée par la survie des parents dans les recensements de 1987 et de 1998.

Au *Sénégal* (p. 325), la mortalité estimée à partir des frères et sœurs vient s'inscrire dans la tendance dégagée des orphelins de mère dénombrés lors du recensement de 1988. Cette tendance cadre également avec l'estimation fournie par Timaeus (1991c) pour 1978 à partir d'une enquête à passage repétée (Cantrelle et al. 1986). Le graphe 7.7 reprend également l'estimation calculée directement des décès déclarés dans les douze derniers mois précédant le recensement de 1988 (Pison et al. 1997). Dans la pratique, ce recensement s'est déroulé presque un an jour pour jour après la Korité, une importante fête musulmane, ce qui a permis de limiter les problèmes généralement associés aux déplacements d'une fraction des décès hors de la période de référence [15]. Pour le Sénégal, on trouvera une

13. Nous n'avons pas exploité les données de l'enquête AIS 2005, qui pourraient permettre de mieux éclairer cette question.

14. Par exemple, des tendances incompatibles d'une enquête à l'autre pourraient introduire des distorsions dans les estimations que l'on déduit de leur combinaison.

15. L'estimation de la survie entre deux recensements a par contre procuré des estimations hautement implausibles à partir des recensements de 1976 et 1988, en raison d'erreurs d'âges systématiques (Pison et al. 1997). Pour la même raison, l'évaluation du taux

discussion plus complète et une comparaison plus systématique des différentes estimations dans Reniers, Masquelier, et Gerland (forthcoming), notamment au moyen des estimations tirées de l'état civil de Saint Louis.

de déclaration des décès s'est révélée impraticable. Timaeus et Jasseh (2004) discutent de cette estimation de 1988, dont ils pensent qu'elle sur-estime légèrement la mortalité.

7.1.1 Mali

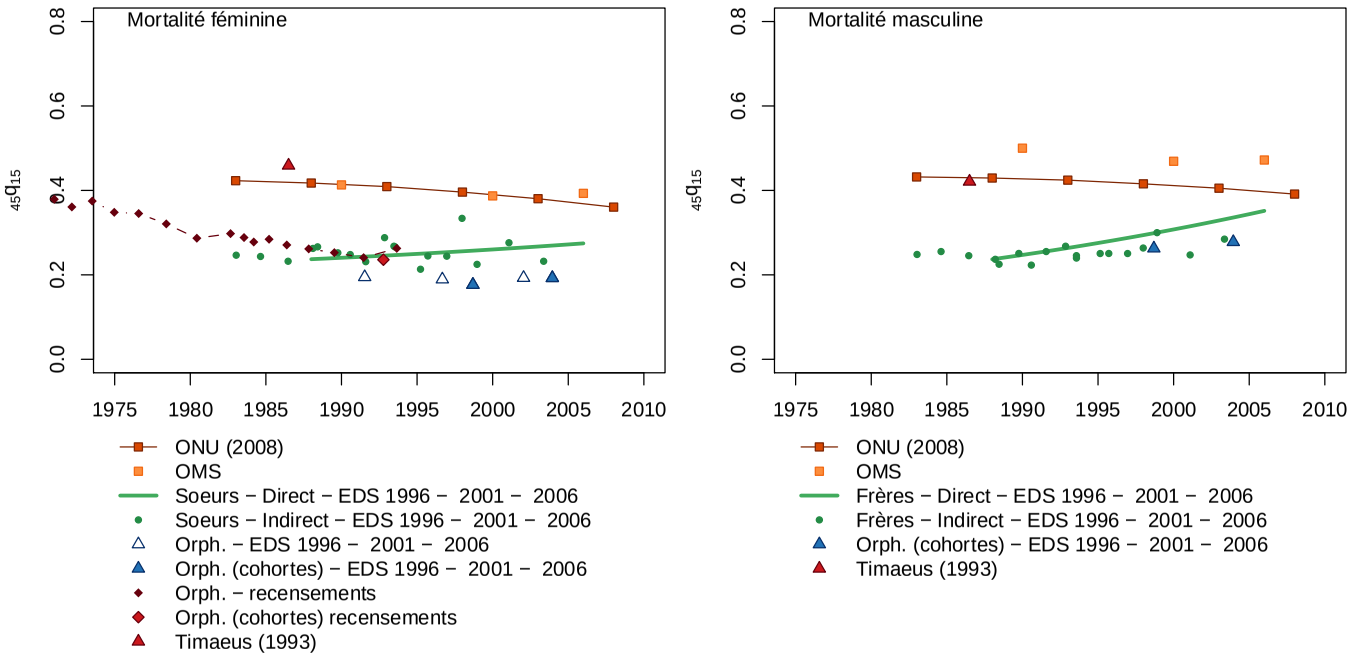

FIGURE 7.5 – *Mali : Estimations du quotient ${}_{45}q_{15}$ à partir de la survie des frères et soeurs et de la survie des parents, comparaison avec les estimations UN et OMS*

7.1.2 Niger

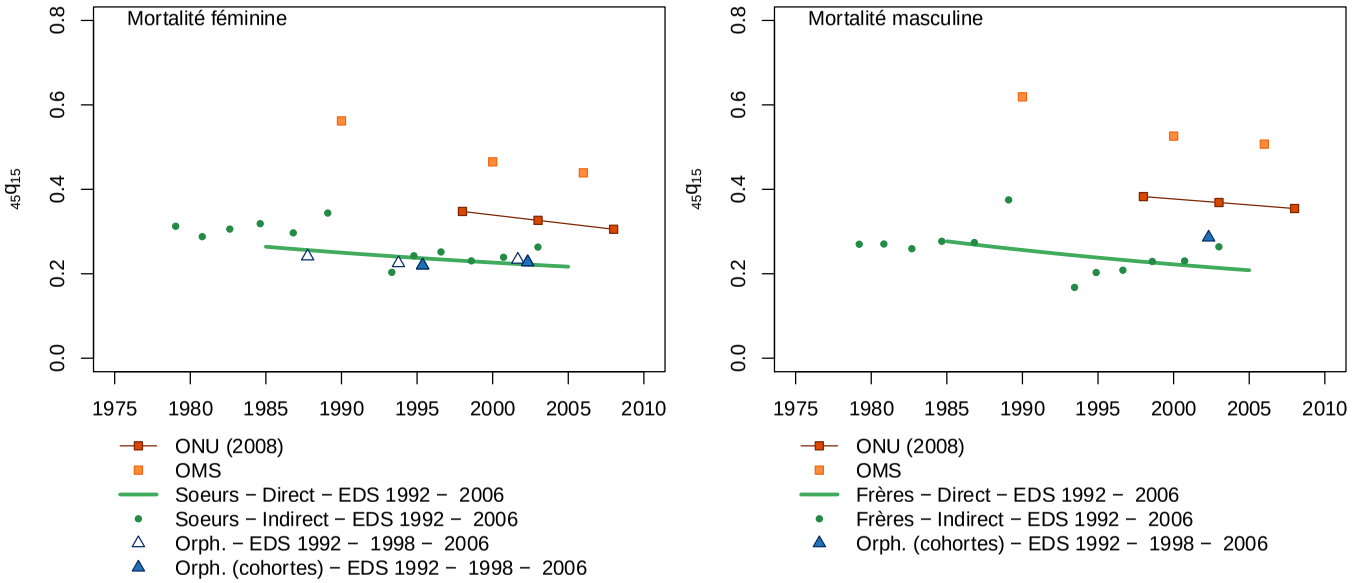

FIGURE 7.6 – *Niger : Estimations du quotient $_{45}q_{15}$ à partir de la survie des frères et soeurs et de la survie des parents, comparaison avec les estimations UN et OMS*

7.1.3 Sénégal

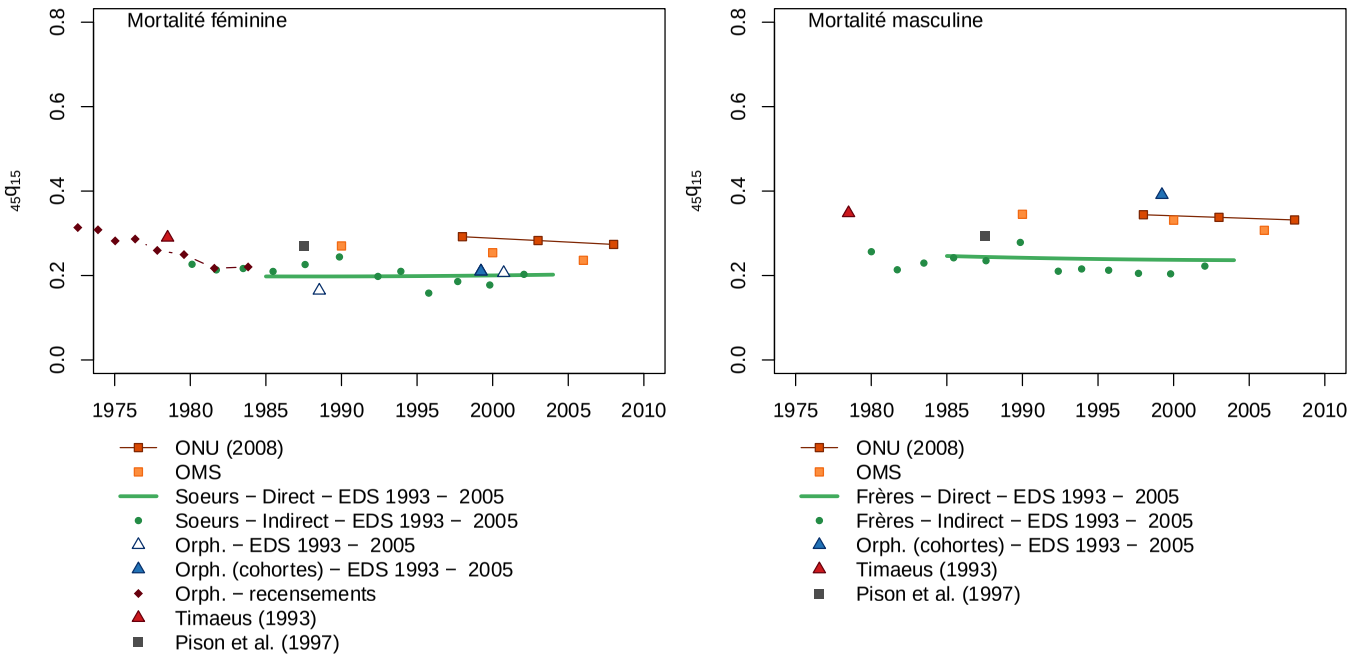

FIGURE 7.7 – *Sénégal : Estimations du quotient ${}_{45}q_{15}$ à partir de la survie des frères et soeurs et de la survie des parents, comparaison avec les estimations UN et OMS*

7.2 Afrique Centrale

L'Afrique Centrale (5 pays présentés ici) occupe une position intermédiaire par rapport aux évolutions qu'ont connues les autres sous-régions. On y trouve en effet :

1. des cas où la mortalité semble avoir augmenté de façon assez nette sous l'effet du VIH-sida - comme c'est le cas du *Cameroun* (graphe 7.9, page suivante),
2. d'autres où elle a globalement stagné sur les deux dernières décennies, comme on l'observe au *Tchad* (p. 328) ou au *Gabon* (p. 465),
3. et au moins un cas où les estimations démographiques récentes portent la marque de la violence et des conflits des années 1990, celui du *Congo-Brazzaville* (p. 461). Même si elle paraît trop prononcée, on peut sans doute associer cette baisse rapide de la mortalité observée dès lors à partir des frères et sœurs à une diminution des tensions consécutive à la fin de la guerre civile en 1997.

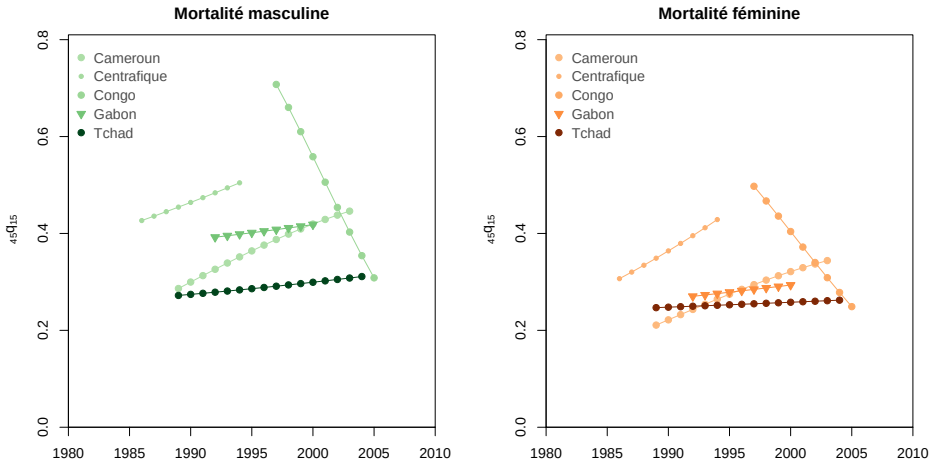

Figure 7.8 – *Quotients de mortalité de 15 à 60 ans calculés à partir de la survie des frères et sœurs et du modèle de Timaeus et Jasseh (2004)*

On relèvera également les cas du *Gabon* (p. 465) et de la *République Centrafricaine* (p. 459), deux pays qui restent peu documentés mais où l'on observe une légère hausse des risques de décès. Notons à propos de ces deux pays que les estimations issues des frères et sœurs sont dans les deux cas *supérieures* à celles des agences internationales. Si l'on convient que les données sur les proches fournissent plutôt des "bornes inférieures", alors on pourrait penser que la mortalité pourrait avoir été sous-estimée par l'ONU et l'OMS dans ces deux pays.

7.2.1 Cameroun

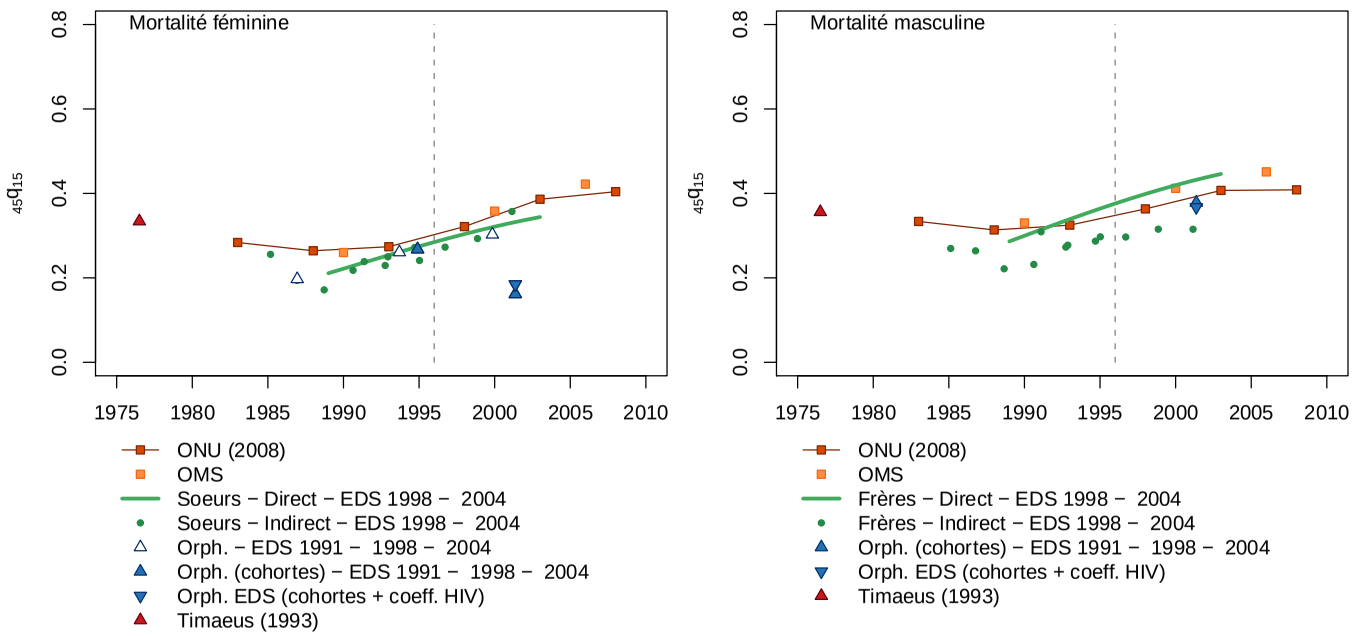

FIGURE 7.9 – *Cameroun : Estimations du quotient ${}_{45}q_{15}$ à partir de la survie des frères et soeurs des parents, estimations UN et OMS*

7.2.2 Tchad

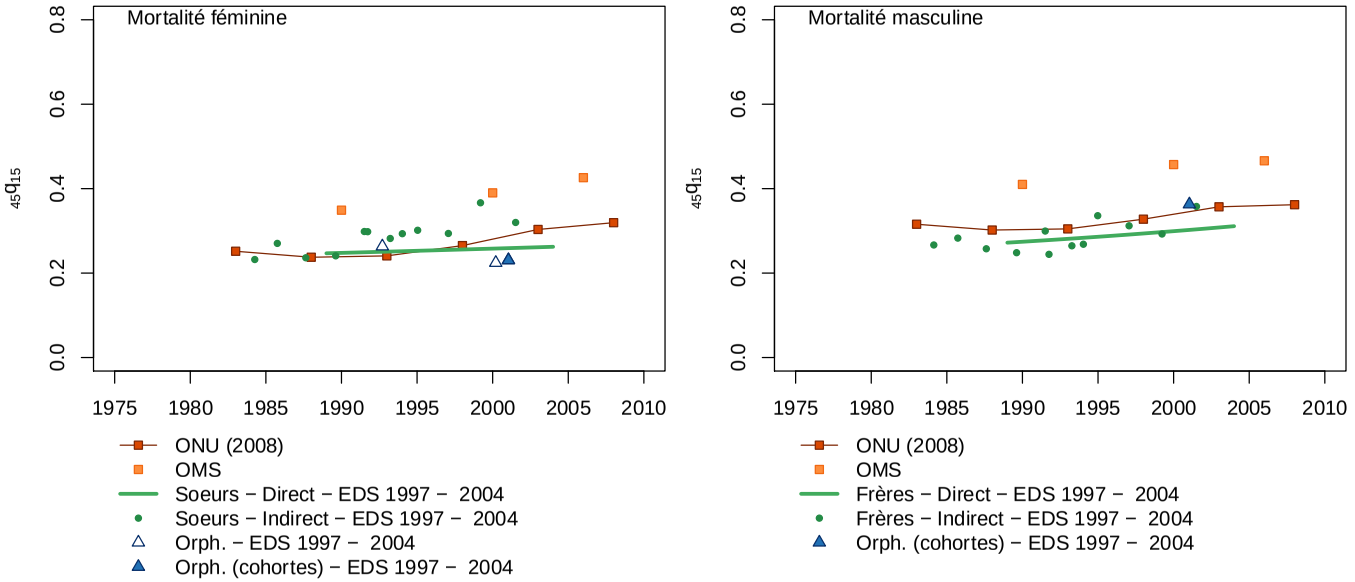

FIGURE 7.10 – *Tchad : Estimations du quotient $_{45}q_{15}$ à partir de la survie des frères et soeurs des parents, estimations UN et OMS*

7.3 Afrique de l'Est

Les différents pays de l'Afrique de l'Est ont été très inégalement affectés par le VIH-sida. Les prévalences nationales de l'Ethiopie, de l'Érythrée, du Burundi ou de Madagascar n'ont jamais dépassé le seuil des 5%, alors que le Kenya voisin a été confronté à ce niveau de prévalence dès le début des années 1990, puis à 10% autour de 1995. L'épidémie a concerné l'Ouganda bien plus tôt, puisque sa prévalence atteint 5% autour de 1985, puis plafonne à 13-14% cinq ans plus tard (UNAIDS 2009). Ces variations de l'intensité de l'épidémie se traduisent par une très grande diversité des niveaux de mortalité estimés (graphe 7.11).

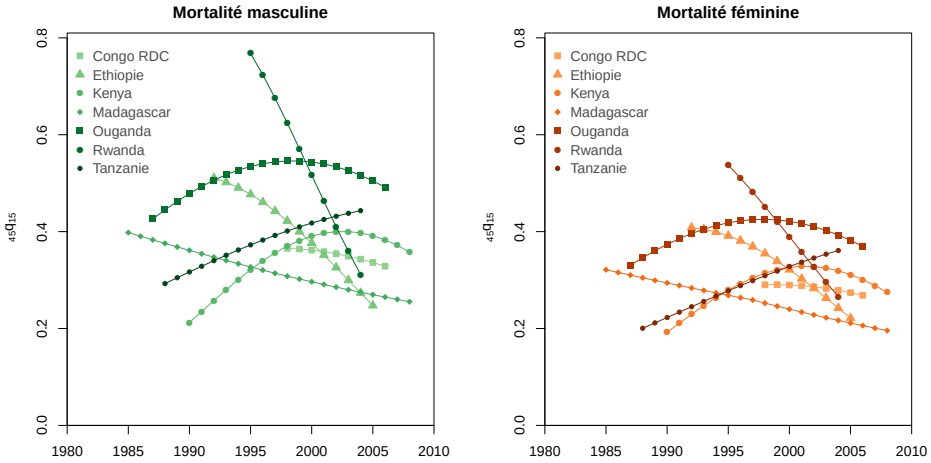

Figure 7.11 – *Quotients de mortalité de 15 à 60 ans calculés à partir de la survie des frères et sœurs et du modèle de Timaeus et Jasseh (2004)*

Le graphe 7.12 (p.331) présente par exemple les différentes estimations que nous avons calculées pour *Madagascar*. On retrouve la tendance quasi linéaire de la baisse de la mortalité déjà présentée précédemment. Cette tendance pourrait certainement être discutée à la lumière des travaux de Waltisperger et Meslé (2005), basés sur l'exploitation des registres de décès de la capitale, Antananarivo. Ces travaux témoignent notamment de hausses de la mortalité adulte en 1985-1987, liées à une augmentation de la mortalité par maladies infectieuses et respiratoires, carences alimentaires et morts violentes. Ces pics conjoncturels n'apparaissent pas ici, et on observe un très bon accord entre les estimations issues de la survie des parents et celles déduites des frères et sœurs [16]. Ces différents niveaux "empiriques" cadrent également cette fois avec les tendances estimées par les Nations Unies et l'OMS, qui toutes indiquent que le pays a enregistré des progrès continus depuis la fin des années 1985. En remontant à 1975, on observe des niveaux de mortalité bien supérieurs, estimés à partir des décès déclarés dans les ménages (Waltisperger et Rabetsitonta 1988, Timaeus 1993).

On observe aussi un degré d'accord assez satisfaisant entre les estimations pour l'*Ouganda*, du moins à partir des données sur la survie des

16. Sauf quand les données sur la fratrie sont exploitées de façon indirecte, surtout pour les hommes. Ceci pourrait peut-être s'expliquer par les proportions importantes de réponses manquantes sur les âges aux décès, ou bien par des erreurs dans les âges des enquêtées.

proches (p. 332), même si pour les femmes, les Nations Unies procurent des niveaux de mortalité plus élevés.

Au *Kenya* (p. 468), nous disposons de plusieurs sources d'informations, et elles témoignent toutes d'une hausse de la mortalité depuis la fin des années 1980. Les estimations dégagées des jeunes orphelins dans les EDS semblent procurer des niveaux de mortalité trop bas, mais ceci est certainement associé au choix du standard pour convertir les valeurs de ${}_np_{25}$ en ${}_{45}p_{15}$. Pour exploiter les données sur la survie des pères, nous avons utilisé les coefficients que nous avions suggérés dans le chapitre 4 pour les pays fortement affectés par le sida. On constate ici qu'ils procurent des niveaux de mortalité masculine plausibles.

Enfin, la mortalité du *Rwanda* (p. 333) est certainement la plus difficile à estimer, puisque les tendances résultent à la fois d'une mortalité violente extrêmement importante et d'une épidémie du VIH-sida qui est loin d'être négligeable. Le Rwanda permet d'illustrer les problèmes rencontrés par les méthodes indirectes quand il s'agit de saisir les variations brutales de la mortalité. La tendance à la baisse observée *de façon directe* à partir des frères et sœurs paraît plausible, si on la compare aux estimations de l'ONU et de l'OMS [17]. En revanche, les estimations *indirectes* basées sur la fratrie ne sont d'aucune utilité, car elles vont lisser les changements abrupts de la mortalité. Notons également les divergences importantes entre les quotients de l'ONU est ceux issus des orphelins (recensements et EDS 1992) pour la période précédant le génocide. Les estimations onusiennes cadrent davantage avec celle obtenue à partir des décès déclarés dans les ménages pour 1978 (Waltisperger et Rabetsitonta 1988, Timaeus 1993).

17. Les niveaux estimés sont inférieurs chez les femmes, mais très proches chez les hommes dans les années qui suivent le génocide de 1994.

7.3.1 Madagascar

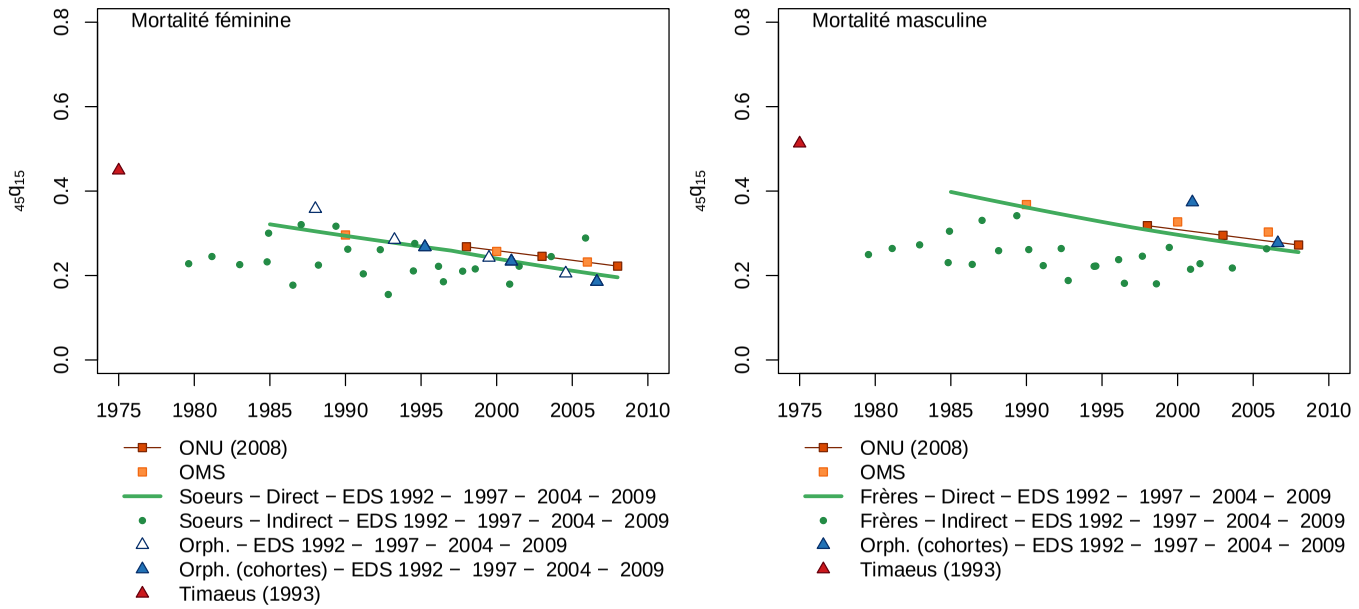

FIGURE 7.12 – *Madagascar : Estimations du quotient $_{45}q_{15}$ à partir de la survie des frères et soeurs des parents, estimations UN et OMS*

7.3.2 Ouganda

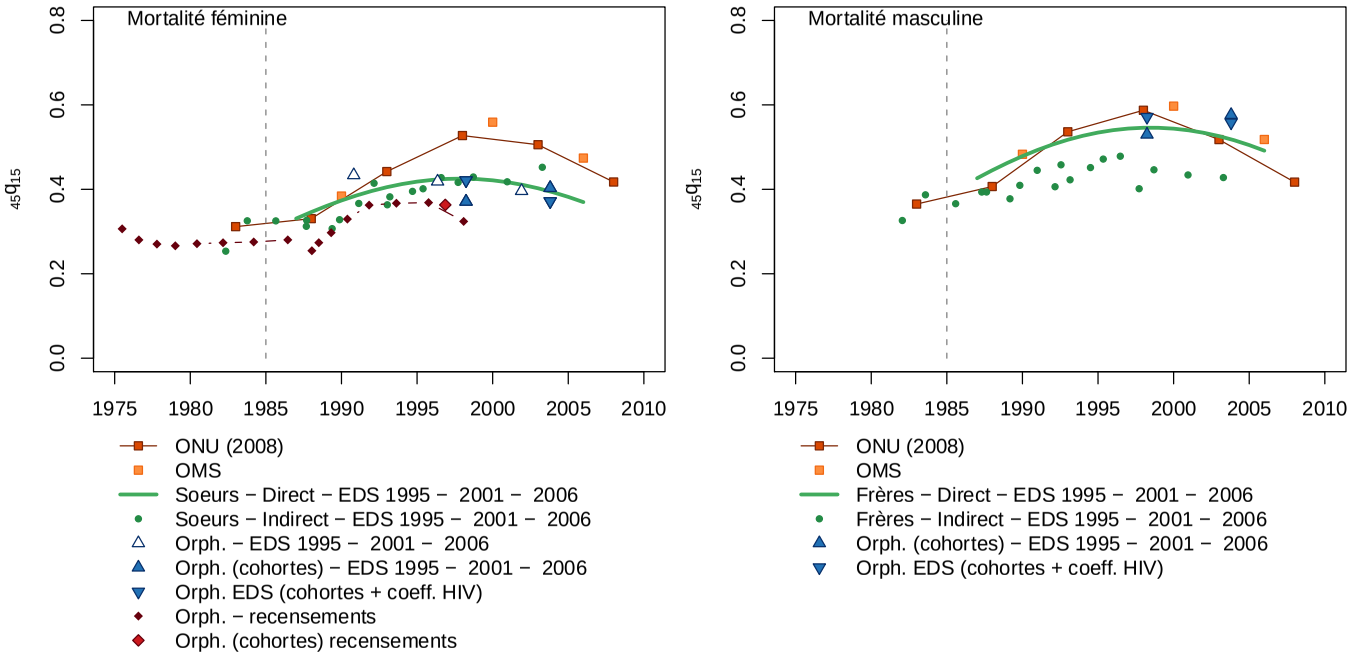

FIGURE 7.13 – *Ouganda : Estimations du quotient $_{45}q_{15}$ à partir de la survie des frères et soeurs des parents, estimations UN et OMS*

7.3.3 Rwanda

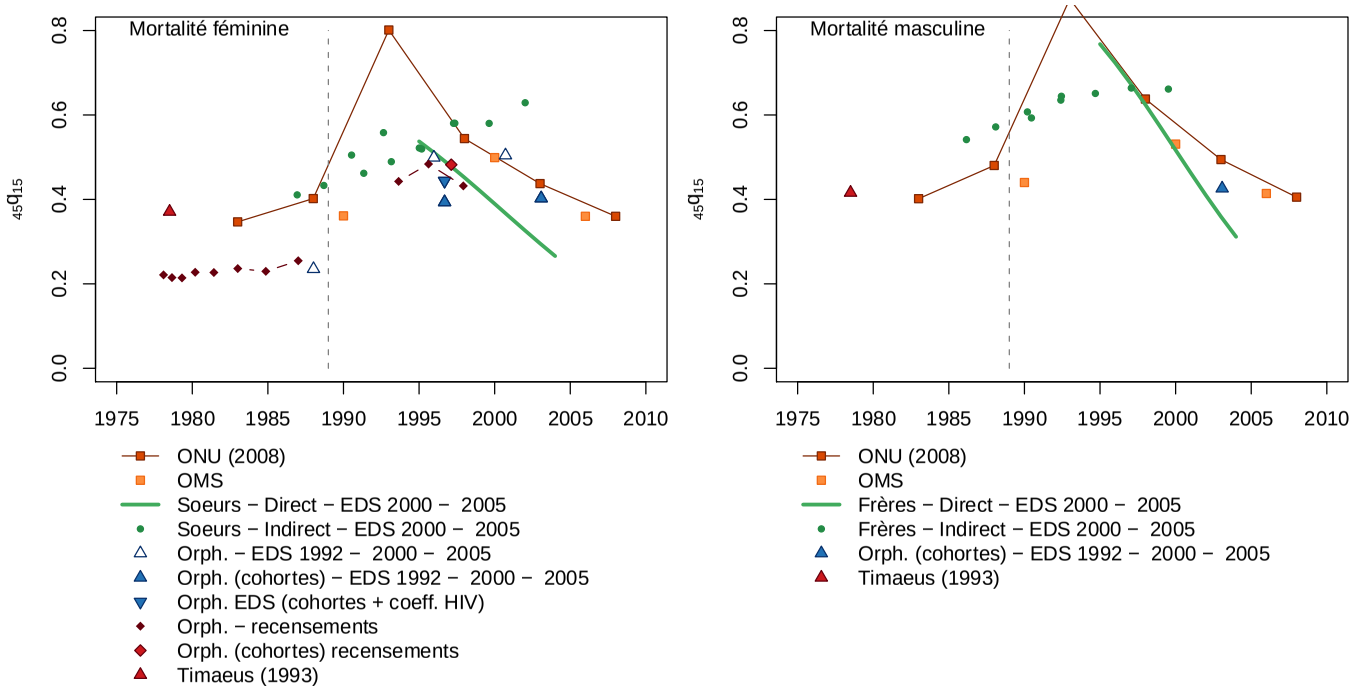

FIGURE 7.14 – *Rwanda : Estimations du quotient $_{45}q_{15}$ à partir de la survie des frères et soeurs des parents, estimations UN et OMS*

7.4 Afrique australe

Le VIH-sida n'a nulle part frappé plus durement les populations qu'en Afrique australe. A l'exception de l'*Angola* et du *Mozambique*[18], tous les pays de la sous-région ont dû faire face à des niveaux de prévalence d'au moins 12%. Pour huit pays de la région, nous disposons de données sur les fratries, ce qui nous a permis de dégager les tendances présentées dans le graphe 7.15. Dans certains cas, comme au Zimbabwe, l'impact du sida est tel que les risques de décès ont triplé sur l'espace de 15 ans.

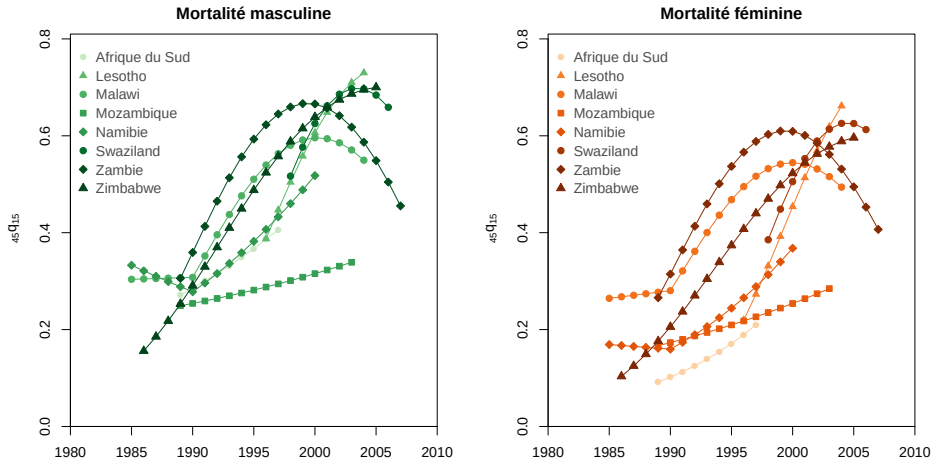

Figure 7.15 – *Quotients de mortalité de 15 à 60 ans calculés à partir de la survie des frères et sœurs et du modèle de Timaeus et Jasseh (2004)*

Quatre situations nationales méritent une attention particulière : l'Afrique du Sud (p. 333), le Malawi (p. 333), la Zambie (p. 333) et le Zimbabwe (p. 333).

L'Afrique du Sud est certainement l'un des pays africains les mieux fournis en matière de données démographiques, et celles collectées sur les parents ou les frères et sœurs ne jouent donc pas le même rôle que dans les cas où il s'agit quasiment de la seule information disponible (comme au Tchad, par exemple). Bradshaw et Timaeus (2003), ainsi que Dorrington et al. (2001; 2004) notamment, combinent de nombreuses sources d'informations pour reconstituer les tendances récentes. On se contentera de relever ici le fait que la mortalité des frères et sœurs est visiblement sous-estimée, tout particulièrement dans les années plus éloignées de l'enquête. En revanche, la méthode des orphelins appliquée aux recensements de 1996, 2001 et 2007 saisit bien l'ampleur de la reprise de mortalité.

Ce que nous observons au *Malawi* et en *Namibie* est interpellant, parce qu'il s'avère que dans les deux cas, les données sur la fratrie suggèrent que la hausse de la mortalité aurait démarré environ 5 à 10 ans plus tôt que ce que prévoient les Nations Unies. Sur les graphes relatifs à ces deux pays, on constate également que la méthode des orphelins semble induire une sous-estimation importante des risques de décès à mesure que progresse l'épidémie du sida, comme nous avons pu le montrer dans le chapitre 4. Ni l'utilisation des coefficients de Timaeus et Nunn (1997) ni le

18. Ainsi que de l'Ile Maurice, dont nous ne discutons pas ici.

recours aux cohortes synthétiques ne semblent procurer des estimations plus plausibles, que du contraire [19]. En comparant ces graphes à ceux établis pour le Zimbabwe, où les risques de décès continuent d'augmenter, on est amené à penser que l'une des sources de cette sous-estimation est liée au fait que la mortalité pourrait avoir amorcé une baisse après avoir augmenté rapidement. On se référera aux travaux de Feeney (2001) et de Reniers et al. (forthcoming) pour une comparaison plus systématique des différentes estimations disponibles au Zimbabwe.

19. Sauf dans le cas de la mortalité masculine en Zambie.

7.4.1 Afrique du Sud

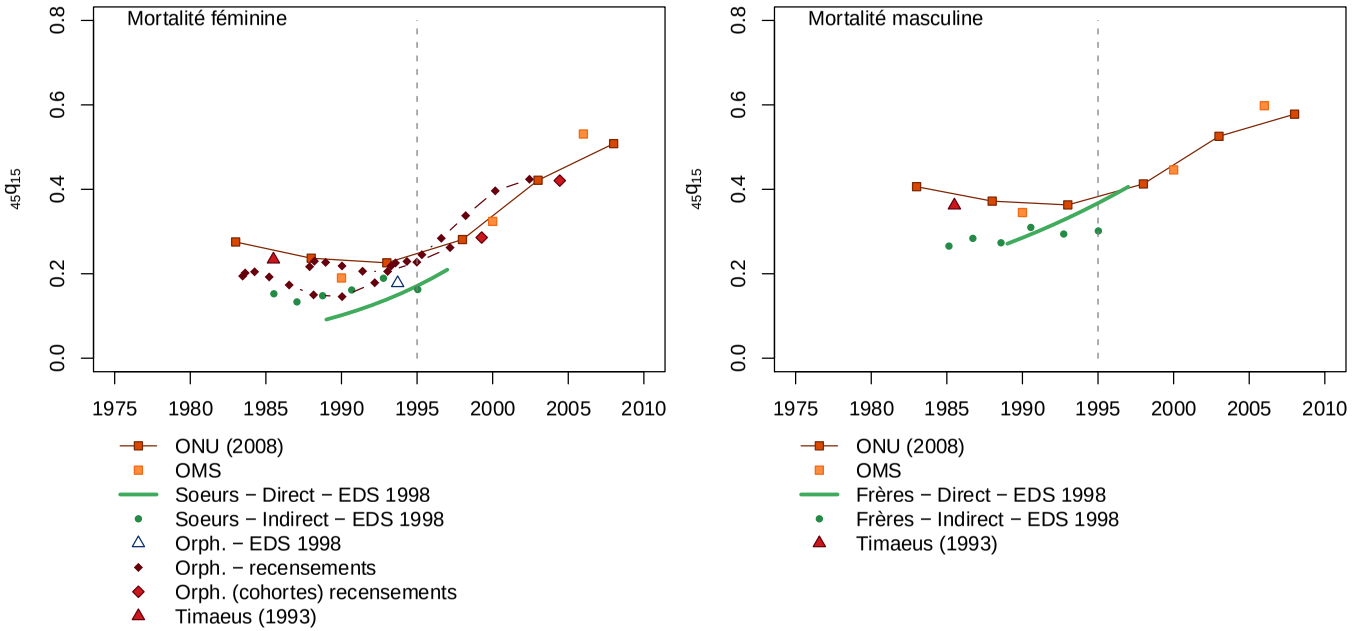

FIGURE 7.16 – *Afrique du Sud : Estimations du quotient $_{45}q_{15}$ à partir de la survie des frères et soeurs des parents, estimations UN et OMS*

7.4.2 Malawi

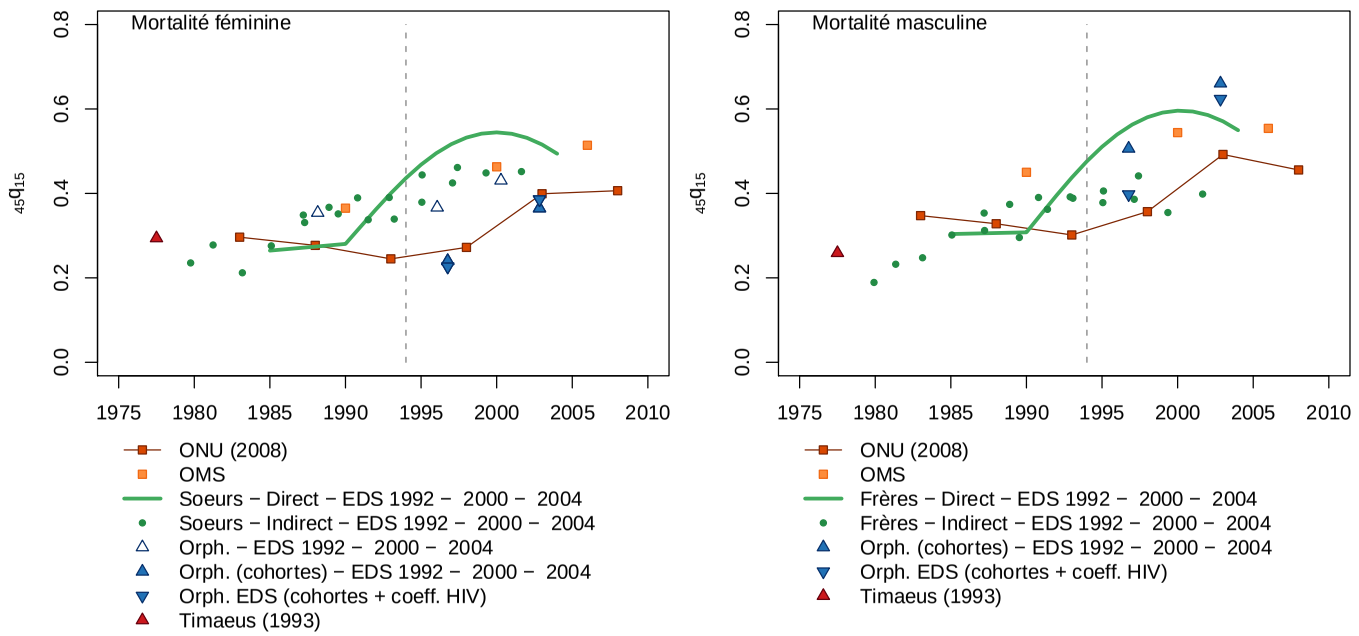

FIGURE 7.17 – *Malawi : Estimations du quotient ${}_{45}q_{15}$ à partir de la survie des frères et soeurs des parents, estimations UN et OMS*

7.4.3 Zambie

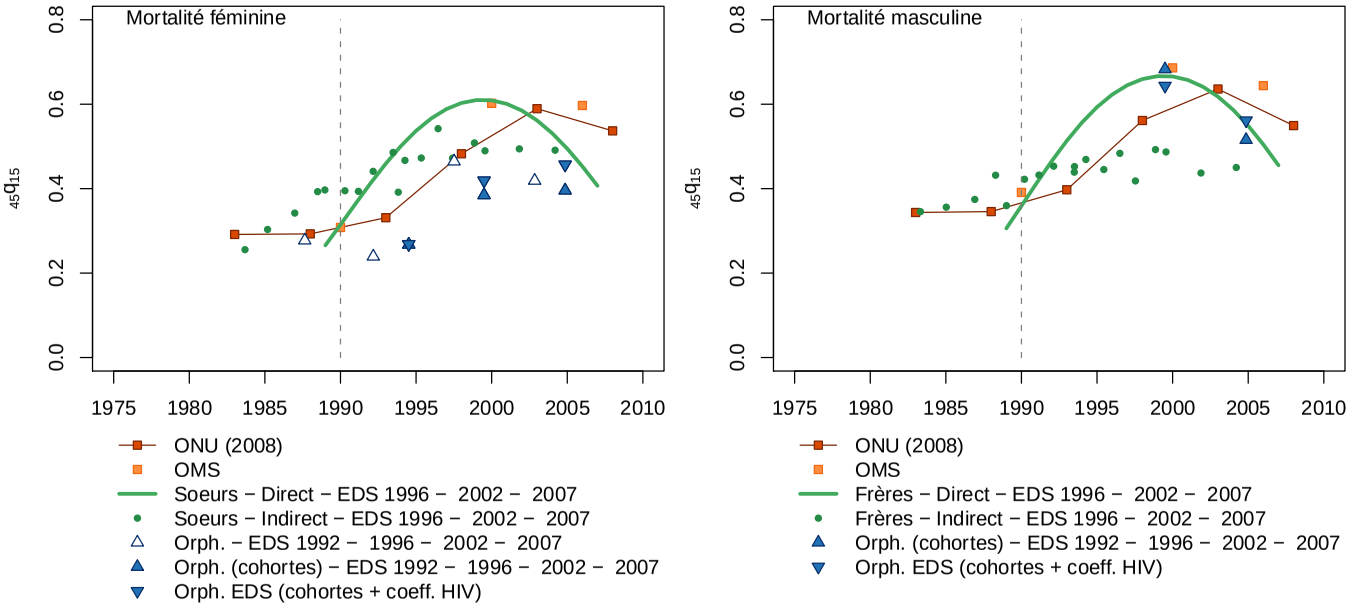

FIGURE 7.18 – *Zambie : Estimations du quotient $_{45}q_{15}$ à partir de la survie des frères et soeurs des parents, estimations UN et OMS*

7.4.4 Zimbabwe

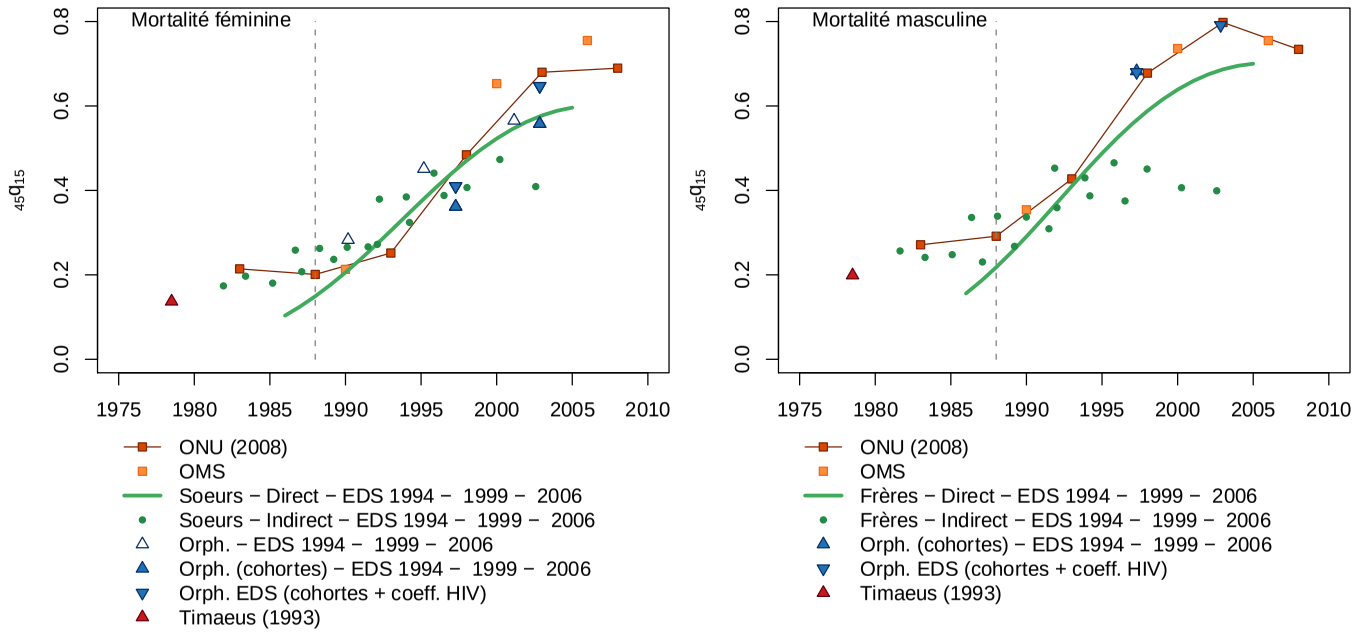

FIGURE 7.19 – *Zimbabwe : Estimations du quotient ${}_{45}q_{15}$ à partir de la survie des frères et soeurs des parents, estimations UN et OMS*

Conclusion

Il n'est pas certain que les données qui portent sur les parents ou les frères et sœurs puissent répondre à davantage de questions qu'elles n'en posent. La combinaison des différentes approches procure souvent un nuage de points relativement étendu au sein duquel il est difficile de distinguer ce qui reflète réellement les tendances passées de ce qui a trait aux erreurs de données et aux différents biais. Mais pour autant, certains des constats qui s'en dégagent peuvent posent question, et c'est déjà une information en soi. Cela invite à examiner de plus près les niveaux de mortalité adultes calculés classiquement à partir des enfants. C'est le cas du *Malawi* et de la *Namibie*, par exemple, ou de la *Guinée*, du *Gabon* et de la *République Centrafricaine*. A priori, saufs perturbations majeures (conflits et sida), on s'attend à ce que les niveaux de mortalité déduits des proches capturent plus ou moins la tendance de la mortalité, tout en étant inférieurs aux estimations issus des modèles. Lorsque ce n'est pas le cas, il faudrait idéalement pouvoir utiliser d'autres approches (méthodes de distribution des décès, etc.) pour chercher une explication à ces situations inattendues.

Conclusion

Plus de vingt ans après le début de la pandémie du VIH-sida en Afrique subsaharienne, de nombreux pays font face à des épidémies matures, et voient leurs courbes d'incidence et de prévalence se stabiliser ou amorcer une baisse. Par contre, en raison de la longue période d'incubation du VIH, la diminution du nombre d'infections ne se traduit pas directement par une baisse des risques de mortalité aux âges adultes. Ces derniers continuent d'ailleurs d'augmenter dans de nombreux pays, particulièrement en Afrique de l'Est et en Afrique australe, et les cas de baisse nette de la mortalité restent rares (Ouganda, Zambie, ...). Dans le même temps, certains pays ont connu une mortalité adulte remarquablement stable sur les dernières décennies, et d'autres ont même continué à enregistrer des progrès. On assiste donc à une diversification croissante des niveaux de mortalité adulte.

En dépit de cette grande diversité des situations, l'impression générale qui prévaut est toutefois celle d'une reprise de la mortalité, essentiellement en raison de la crise du VIH-sida. Sur les 20 pays pour lesquels nous pouvons estimer la mortalité adulte masculine à partir de la survie des frères pour l'année 1990, *aucun* ne connaissait des probabilités de décès entre 15 et 60 ans (${}_{45}q_{15}$) supérieures à 0.5. *Dix ans plus tard, c'était le cas d'un pays sur 3.* Cette reprise de mortalité est manifeste sur la carte 7.20 (p. 342), qui rassemble les différentes estimations basées sur les frères et soeurs pour les années 1995, 2000 et 2005.

De nombreuses incertitudes entourent encore l'estimation de ces mortalités adultes élevées, qui restent souvent appréhendées par le biais de la mortalité des enfants et de schémas de mortalité standards, auxquels on vient ajouter ensuite les décès causés par le VIH-sida. Comme l'indique la comparaison des estimations de l'OMS et des Nations Unies (p.313), les niveaux de mortalité estimés de la sorte sont encore largement déterminés par des options méthodologiques qui sont parfois difficiles à justifier. Vérifier la cohérence des estimations issues de la modélisation avec celles déduites des enquêtes et des recensements est une chose, intégrer dès le départ des niveaux de mortalité adulte au cours de la modélisation en est une autre. Et si la seconde option n'est pas exempte de risques, la première fait courir celui d'une sur- (ou sous-)estimation systématique, et parfois substantielle.

Par rapport à cette approche dominante basée sur la modélisation, nous avons consacré ce travail de thèse à la revalorisation des données sur la survie des proches, à l'identification de leur potentiel et de leurs faiblesses, ainsi qu'à l'examen des méthodes permettant de les exploiter.

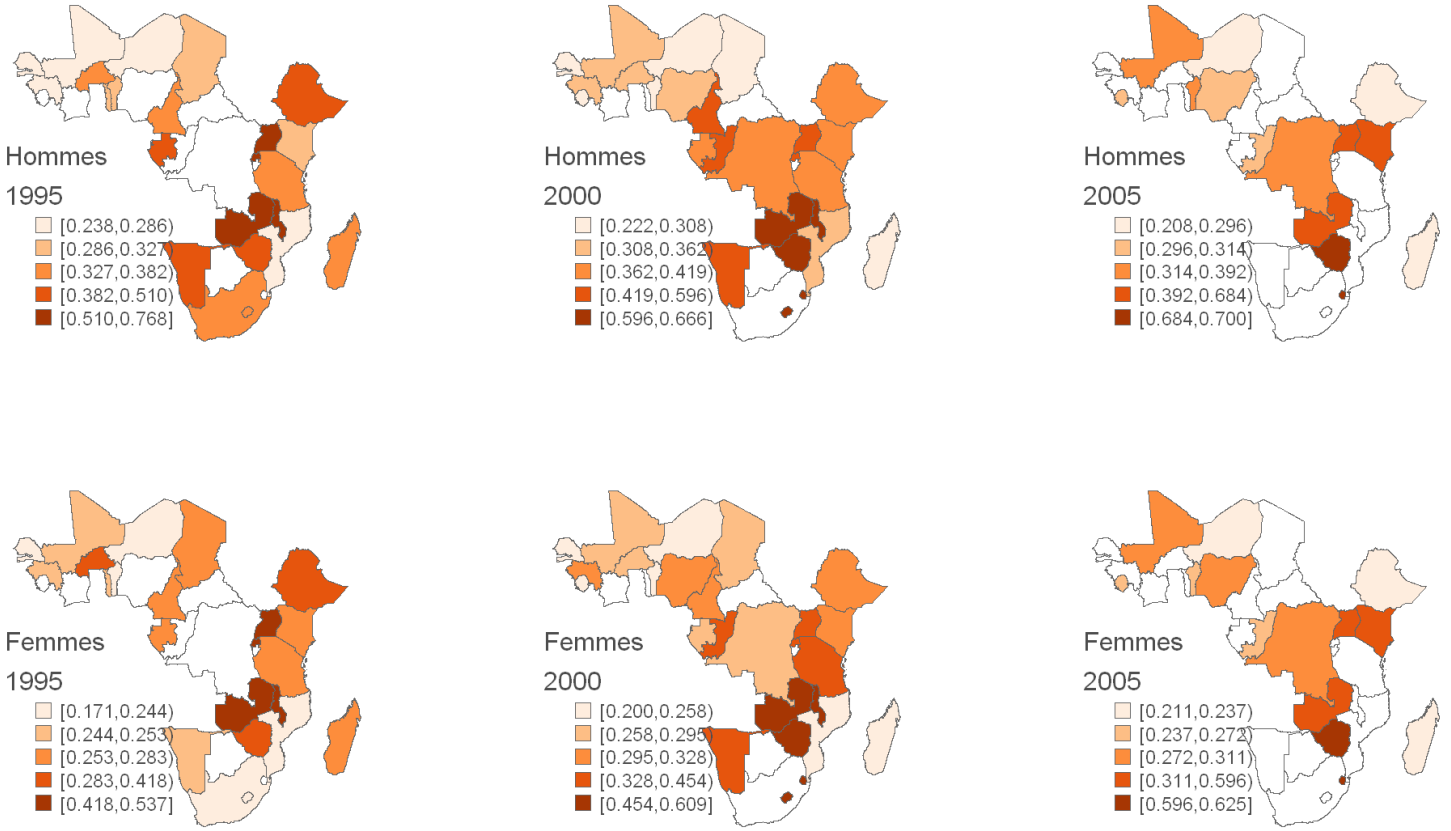

FIGURE 7.20 – *Valeurs estimées de la probabilité de décéder de 15 à 60 ans pour 1995, 2000 et 2005 à partir des données EDS relatives aux frères et soeurs (exploitation directe).*

Dès nos premières recherches documentaires, il est apparu clairement que la mortalité adulte est encore le parent pauvre de la démographie africaine. Face aux défis posés par la crise du VIH-sida et par la persistance d'une mortalité maternelle élevée, la *santé sexuelle et reproductive* s'est peu à peu imposée parmi les premières préoccupations des organisations internationales et des gouvernements, ainsi que des démographes. Mais cet intérêt ne s'est que rarement étendu à la mortalité *générale* au-delà de quinze ans. Au niveau des programmes de collecte, comme au niveau des programmes de recherche, l'accent reste souvent mis soit sur le VIH-sida, soit sur la santé des mères et des enfants.

Il est d'ailleurs assez emblématique de noter qu'aujourd'hui encore, les données sur la survie des frères et soeurs - qui constituent l'une des rares sources permettant d'analyser la mortalité adulte de façon directe - restent récoltées comme des sous-produits de la mesure de la santé reproductive. En effet, ces données sont toujours collectées dans les EDS au sein d'un module appelé "mortalité maternelle", et elles sont à l'origine destinées à l'estimation de celle-ci. Du côté des données sur la survie des parents, on peut faire les mêmes constats : le fait de limiter les questions sur la survie des parents aux moins de 15 ans réduit grandement leur utilité pour reconstituer des tendances. Etendre cette limite d'âge aux adultes permettrait de calculer de nouvelles proportions de parents survivants et d'obtenir des estimations de mortalité adulte moins entachées par le biais d'adoption ou les biais associés au VIH-sida.

Les choses changent toutefois peu à peu, et la littérature récente témoigne d'un regain d'intérêt pour l'estimation de la mortalité adulte. D'un côté, les méthodes de distribution des décès, qui permettent de corriger l'enregistrement incomplet des décès, ont fait l'objet de récents développements (Hill et al. 2009, Murray et al. 2010). Par ailleurs, les données basées sur la survie des proches ont également été exploitées de façon plus systématique (Gakidou et al. 2004, Timaeus et Jasseh 2004, Obermeyer et al. 2010, Rajaratnam et al. 2010, Reniers et al. forthcoming). Ces informations collectées sur la survie des proches constituent un patrimoine statistique important, à valoriser davantage notamment pour mieux reconstituer les tendances de la mortalité adulte. C'est ce que nous avons cherché à faire ici, en combinant les enquêtes démographiques et de santé (EDS), des données de recensements ainsi que des microsimulations.

Objectifs poursuivis

Cette recherche a poursuivi les objectifs suivants :

1. Il s'agissait d'abord de *fournir un aperçu aussi complet que possible des données disponibles sur la survie des proches à l'échelle nationale*. Ces données s'avèrent en fait *relativement nombreuses*, dans un contexte où les autres types de données restent rares (état civil, décès déclarés dans les recensements, etc.). En combinant les différents programmes de collecte, près d'une centaine d'enquêtes nationales ont été menées en Afrique subsaharienne depuis 1990, avec une série de questions sur la survie des frères et soeurs. Plus de 135 enquêtes nationales et

au moins 50 recensements ont également intégré des questions sur la survie des parents depuis 1965. Ces données sont souvent standardisées ou du moins comparables, et disponibles quelques années à peine après leur collecte. Elle sont également facilement accessibles, de sorte qu'il est possible de combiner les enquêtes démographiques et de santé (EDS), les enquêtes par grappes à indicateurs multiples de l'UNICEF (MICS), les enquêtes mondiales de la santé de l'OMS (WHS), et les données issues de certains recensements. Une analyse plus systématique de ces différentes séries de données sortait du cadre de cette recherche, mais elle en constitue certainement un prolongement intéressant.

2. Nous avons ensuite *revisité les différentes méthodes permettant d'exploiter ces données, en cherchant à traiter de façon unifiée la survie des parents et celle des frères et soeurs*. Deux types d'exploitation des données ont été présentés : l'exploitation *directe*, qui rapporte simplement les décès aux unités d'exposition, et le calcul *indirect*, qui consiste à convertir des indicateurs de structure (des proportions de survivants) en indicateurs de mouvements (des probabilités de survie). A ce jour, seules les données sur les frères et soeurs permettent ces deux types d'exploitation, mais on peut très bien imaginer collecter des données sur l'âge des parents à l'enquête ou sur leur âge au décès.

   Parce que de nombreux éléments viennent perturber les analyses menées sur des données empiriques - au premier rang desquels on trouve les problèmes de qualité des données - la microsimulation nous est rapidement apparue comme étant indispensable pour revisiter les différentes méthodes de calcul. Dans cet esprit, nous avons mis au point *une série de microsimulations représentatives de la diversité des dynamiques démographiques africaines*. Leur intérêt est que les liens de parenté sont conservés au cours des simulations, à l'échelle individuelle. Ceci constitue un outil très intéressant qui pourrait d'ailleurs être exploité à d'autres niveaux (structures de soutien au sein des réseaux familiaux, etc.). L'intérêt des microsimulations pour l'analyse de la mortalité adulte est que les différentes proportions de proches survivants peuvent être observées directement plutôt que d'être modélisées, ce qui autorise davantage de souplesse. Ceci nous a permis de mener différents types de tests sur les méthodes d'estimation, en examinant notamment les perturbations induites par le VIH-sida.

3. Enfin, le troisième objectif principal consistait à rassembler par pays les différentes estimations qu'il était possible de dégager de la survie des proches, afin de les comparer à celles fournies par l'OMS et l'ONU, qui sont essentiellement basées sur une combinaison de la mortalité des enfants avec des tables de mortalité ainsi que sur une modélisation de l'épidémie du VIH-sida.

Tableaux de synthèse des méthodes applicables aux données imparfaites

Avant de relever les principaux apports de cette thèse vis-à-vis de la littérature existante, les quelques pages suivantes présentent des tableaux de synthèse reprenant les différentes méthodes d'estimation de la mortalité adulte applicables aux données imparfaites. Ils en précisent le principe, les données nécessaires, les développements ainsi que leurs principales limites.

Les deux premiers tableaux (pp. 346 et 347) sont consacrés à des méthodes qui n'ont pas été analysées ici, mais qui ont été discutées dans le chapitre 1. Elles reposent surtout sur la collecte d'informations sur les décès survenus dans les ménages. Toutefois, comme cela a déjà été suggéré par Hill (1998), on pourrait imaginer appliquer les méthodes de distribution des décès (moyennant sans doute certains aménagements) aux données sur les frères et soeurs, afin d'estimer l'ampleur de la sous-déclaration des décès.

Les tableaux 7.6 et 7.7 reprennent ensuite les différentes options disponibles pour exploiter les données sur la survie des parents dans les contextes peu perturbés par le VIH-sida (p. 348) et là où le VIH-sida introduit des changements importants au niveau des schémas par âge et des tendances de la mortalité (p. 349). Nous revenons plus loin sur les principales conclusions auxquelles nous aboutissons à propos de ces approches.

Enfin, le tableau 7.8 présente simplement les méthodes disponibles pour exploiter les données sur les frères et soeurs.

TABLE 7.4 – ***Méthodes d'estimation de la mortalité adulte (1) : Méthodes de base***

	Description	Type	Source de données	Développements	Limitations
Calcul direct depuis l'enregistrement des décès	Rapport du nombre de décès sur l'exposition	Direct	Décès rapportés dans les ménages (recensements ou état civil) distribués par âge et sexe		Infaisable dans beaucoup de pays en raison de la faible qualité des données collectées[a]
Méthodes de survie intercensitaire	Estimation de l'attrition des cohortes durant la période intercensitaire par comparaison des effectifs distribués par groupes d'âge	Indirect	Distribution par âge de deux recensements	Recours aux taux de croissance intercensitaire pour travailler avec des intervalles non multiples de 5 ans[b].	Biais introduits en raison de variations dans la complétude de l'enregistrement de la population, des erreurs d'âge systématiques et de flux migratoires importants.

[a] L'analyse menée par Pison et al. (1997) sur les données du recensement sénégalais de 1988 procure toutefois des résultats satisfaisants. *Cf.* p. 325

[b] *Cf.* Preston et Bennett (1983)

TABLE 7.5 – *Méthodes d'estimation de la mortalité adulte (2) : Méthodes de distribution des décès*

	Description	Type	Source de données	Développements	Limitations
Generalized Growth Balance Method (GGB)[a]	Estimation de la complétude de l'enregistrement des décès via la comparaison du taux de croissance intercensitaires par âge et de la distribution par âge de la population	Indirect	Décès rapportés dans les ménages (recensements ou état civil) distribués par âge et sexe entre deux recensements dont sont issues deux distributions par âge		Population supposée fermée, complétude de l'enregistrement des décès et de la population supposés constants par âge. Méthode sensible aux erreurs d'âge[b]
Synthetic extinct generation method (SEG)[c]	Comparaison de la distribution par âge des décès et de la population	Indirect	Idem	Hill et al. (2009) proposent de combiner la SEG et la GGB (cfr. infra)	Mêmes hypothèses que la GGB, en plus d'une constance de la couverture de l'enregistrement dans le temps.
Méthode hybride SEG + GGB[d]	Ajuster, selon les résultats de la GGB, les effectifs de population selon la complétude de l'enregistrement du recensement 2 par rapport au recensement 1, puis appliquer la méthode SEG	Indirect	Idem	Murray et al. (2010) discutent des intervalles d'âges sur lesquels effectuer les calculer pour obtenir les résultats les moins biaisés.	Mêmes hypothèses que la GGB, donc sans l'hypothèse de la constance de la couverture de l'enregistrement dans le temps de la SEG.

[a] *Cf.* Hill (1987). Les deux méthodes procurent une estimation de la complétude de l'enregistrement des décès relative aux deux recensements. La CGB et la méthode hybride (SEB+GGB) procurent en outre une estimation de la complétude de l'enregistrement de la population lors du second recensement (par rapport au premier).

[b] Si des taux de migration par âge sont disponibles, il est possible d'intégrer la migration. Possibilité également d'effectuer les calculs sur certains âges uniquement.

[c] *Cf.* Bennett et Horiuchi (1984).

[d] *Cf.* Hill et al. (2009), Dorrington et al. (2006c), Murray et al. (2010). Voir également la note 13, p. 17, ou nous présentons brièvement les conclusions de Murray et al. (2010).

TABLE 7.6 – ***Méthodes d'estimation de la mortalité adulte (3.1) : Méthodes des orphelins - contextes peu perturbés par le VIH-sida***

	Description	Type	Source de données	Développements	Limitations
Méthode de base avec datation indirecte[a]	Conversion de la proportion de parents survivants en probabilité de survie, datation indirecte (p.143)	Indirect	Proportion de non-orphelins de mère et de père par âge	Possibilité de procéder à une datation directe s'il existe des données sur les années de décès des parents (p.148)	Problèmes liés à la datation des estimations lorsque la mortalité n'a pas évolué linéairement (p.183), sous-estimation due à l'effet d'adoption (pp.176, 278), légers biais de sélection (p. 163), inapplicable en cas de fortes perturbations dues au VIH-sida.
Calcul basée sur les cohortes synthétiques[b]	Calcul d'un facteur d'ajustement en cumulant les taux de croissance des proportions de non-orphelins	Indirect	Proportion de non-orphelins de mère et de père par âge observées dans deux opérations de collecte	Possibilité de limiter les biais d'adoption en partant des proportions d'orphelins à l'âge adulte[c] ou au mariage[d]	Nécessité de disposer d'estimations récentes et dans certains cas, relatives aux adultes. Inapplicable à partir des enfants en cas de fortes perturbations dues au VIH-sida (p. 193)

[a] Plusieurs coefficients sont disponibles - *cf.* section 3.2.1 (p.111) pour les coefficients de Timaeus (1992) et section 3.2.3 (p.122) pour les autres, qui procurent globalement les mêmes résultats.
[b] Deux méthodes disponibles selon que les comparaisons entre les deux enquêtes (ou recensements) sont menées par cohorte ou par groupe d'âge. *cf.* p.149.
[c] *Cf.* Timaeus (1991b) et section 3.4.1, p. 156.
[d] *Cf.* Timaeus (1991a), et section 3.4.2, p. 159

TABLE 7.7 – *Méthodes d'estimation de la mortalité adulte (3.2) : Méthodes des orphelins - contextes perturbés par le VIH-sida*

	Description	Type	Source de données	Développements	Limitations
Méthode de Timaeus et Nunn (1997)	Correction des biais de sélection selon la prévalence atteinte (p.201), conversion de la proportion de parents survivants en probabilité de survie à partir de coefficients adaptés au VIH-sida (p. 203)	Indirect	Proportion de non-orphelins de mère et de père par âge, prévalence au moment de la naissance des enfants		Coefficients provisoires estimés à partir de l'observatoire de Masaka, applicables à la mortalité féminine uniquement
Nouveaux coefficients estimés par simulation (p. 218)[a]	Application de la méthode des cohortes synthétiques avec nouveaux coefficients basés sur la prévalence atteinte	Indirect	Idem		Coefficients sousmis à titre provisoire, inadaptés aux contextes où la prévalence du sida est faible

[a] Ces coefficients sont calculés à partir des proportions d'orphelins estimés par la méthode de Grassly et Timaeus (2005), qui sont ensuite regressées sur les probabilités de survie estimées par les Nations Unies.

TABLE 7.8 – ***Méthodes d'estimation de la mortalité adulte (4) : Méthodes basées sur la survie des frères et soeurs***

	Description	Type	Source de données	Développements	Limitations
Calcul direct à partir de l'historique des fratries	Transformation du fichier d'enquête en fichier "frères et soeurs" puis en fichier "personnes-périodes", taux calculés en rapportant les décès aux unités d'exposition	Direct	Age à l'enquête, âges aux décès des frères et soeurs, durées écoulées depuis leur décès	Modèle de régression permettant de modéliser les tendances et les schémas par âge de la mortalité[a], possibilité de tenir compte de la sous-déclaration des décès dans le passé lointain	Biais introduits si la mortalité adulte varie avec la taille des fratries aux âges adultes[b], sensibilité aux erreurs d'âge, problèmes potentiels au niveau de la datation des événements, sous-déclaration importante des décès.
Méthode indirecte[c]	Conversion des proportions de frères et soeurs survivants en probabilités de survie	Indirect	Proportion de frères et soeurs survivants parmi ceux qui ont survécu à 15 ans (distribuée par âge des enquêtés).		Sensible aux erreurs sur l'âge des enquêtés, hypothèse de linéarité des tendances pour la datation.

[a] *Cf.* section 5.1.2, p. 229.

[b] Obermeyer et al. (2010) ont proposé un mécanisme de pondération basé sur les travaux de Gakidou et King (2006) mais nous montrons dans la section 5.2.4 qu'elle n'est pas nécessaire et qu'elle peut introduire une sur-estimation importante de la mortalité.

[c] La méthode la mieux adaptée est celle proposée par Timaeus et al. (2001) car elle ne concerne que les frères et soeurs ayant survécu jusqu'à 15 ans.

Principaux apports

Cette recherche apporte des éléments nouveaux sur trois plans : au niveau du modèle de microsimulations, au niveau de la méthode des orphelins et au niveau de l'estimation de la mortalité adulte à partir des frères et soeurs.

Au niveau des microsimulations :

- Nous proposons un modèle de microsimulation basé sur les estimations des Nations Unies. Ce modèle fournit des populations virtuelles qui reflètent les trajectoires démographiques de la plupart des pays subsahariens, au sein desquelles il est possible de retracer les liens de parenté.
- Ce modèle tient compte de nombreuses implications démographiques du VIH-sida, et modélise la survie des personnes infectées au moyen d'une série de probabilités de transition entre différents sous-groupes. Nous proposons également de modéliser la fécondité par parité au départ des schémas de fécondité par rang dégagés des enquêtes EDS.
- Ce modèle permet de s'approcher des conditions d'application réelles des méthodes directes et indirectes d'estimation de la mortalité adulte basées sur les proches. Les microsimulations s'avèrent notamment très utiles quand il s'agit de répliquer des situations d'enquête ou de tirer plusieurs centaines d'échantillons au sein des populations simulées.

Au niveau de la survie des parents :

- Nous observons que les proportions d'orphelins issues des microsimulations sont significativement *plus élevées* que celles dégagées des enquêtes EDS. Ce constat rejoint celui fait par Grassly et al. (2004) à partir des simulations au niveau macro détaillées dans le chapitre 4. Différentes recherches relatives au biais d'adoption laissent penser que les orphelins sont sensiblement sous-déclarés dans les enquêtes et les recensements. Pour cette raison, nous rappelons que les niveaux de mortalité qui seraient en-deça de ceux dégagés des orphelins devraient être considérés avec beaucoup de prudence.
- Nous montrons que dans les pays où les tendances ont été approximativement linéaires, la méthode des orphelins "de base" permet de procurer des niveaux de mortalité plausibles. En l'absence du VIH-sida, les biais de sélection restent limités - ils sont bien moindres que les problèmes rencontrés au niveau de la qualité des données.
- En revanche, nous montrons également que l'application "classique" de la méthode des orphelins dans les contextes perturbés par le sida induit une sous-estimation importante des risques de décès à partir des plus jeunes enquêtés, qui va en grandissant à mesure que l'épidémie avance en âge. L'un des problèmes principaux est posé par l'hypothèse de linéarité des tendances qui est à la base des méthodes de datation. La combinaison de deux opérations de collecte successives permet de contourner ce problème de la datation et de

fournir des estimations plus récentes, mais le recours à des coefficients adaptés au VIH-sida reste nécessaire.
– Parce que la littérature sur ce sujet reste particulièrement pauvre, à l'exception des recherches de Timaeus et Nunn (1997), nous avons ensuite estimé de nouveaux coefficients, permettant de recourir à la méthode des orphelins dans les contextes perturbés par le VIH-sida. L'application de ces coefficients aux micro-simulations et aux données d'enquête donne des résultats encourageants et il s'agit d'une piste de développements futurs.

Au niveau de la survie des frères et soeurs :
– La démonstration mathématique de Trussell et Rodriguez (1990) est corroborée par les microsimulations. Elle peut même être étendue aux situations des enquêtes, où seules les femmes adultes de 15 à 49 ans sont interrogées.
– Nous illustrons à l'aide de simulations l'importance d'interroger tous les individus éligibles et de ne pas s'en tenir à une déclaration par fratrie ou par ménage, sans quoi la mortalité risque d'être surestimée (surtout là où la mortalité reste faible et où les fratries sont de petite taille).
– Les microsimulations indiquent également qu'il est préférable de ne pas remonter trop loin dans le passé, car les erreurs aléatoires augmentent rapidement (sans parler encore de la sous-déclaration des décès).
– L'exploitation classique des données EDS sur les frères et soeurs nous semble procurer des résultats sans biais. La mortalité aux âges adultes ne paraît pas varier avec la taille des fratries, contrairement à ce qu'avaient montré Gakidou et King (2006).
– Par ailleurs, nous utilisons les microsimulations pour démontrer que la pondération suggérée par ces auteurs ne peut être appliquée comme telle aux enquêtes EDS. Une application trop rapide qui ne tiendrait pas compte des spécificités des enquêtes (et notamment des caractéristiques des enquêtées) procurerait selon nous des niveaux de mortalité tout à fait biaisés. Notamment, nous pensons que les niveaux de mortalité élevés, publiés par Obermeyer et al. (2010), et repris récemment par Rajaratnam et al. (2010), sont biaisés à la hausse.
– Nous apportons également quelques réserves à l'égard des indicateurs traditionnellement utilisés pour juger de la qualité des données sur les frères et soeurs. Par exemple, si dans les rapports d'enquêtes EDS, le nombre de frères et soeurs déclarés par âge est utilisé comme un indicateur de qualité, nous montrons qu'il apporte peu d'informations en soi, et qu'il est préférable de croiser l'évolution de la taille des fratries avec celle relative aux nombres d'enfants nés vivants. A propos de la composition par sexe des fratries, nous mettons en évidence une grande variation régionale, qui recoupe celle observée en matière de rapport de masculinité à la naissance, mais qui peut également s'expliquer par une plus grande fréquence de la co-résidence des soeurs âgées de 15 à 49 ans au sein des ménages enquêtés.
– En adoptant une démarche déjà appliquée par Timaeus et Jasseh

(2004), Obermeyer et al. (2010) et Reniers, Masquelier, et Gerland (forthcoming), nous montrons en outre comment la sous-déclaration des décès des frères et soeurs varie dans le temps. L'année qui précède directement l'enquête est caractérisée par une légère sous-déclaration des décès des frères et soeurs. Les décès ne semblent pas sous-déclarés quand ils surviennent 5 ans avant l'enquête, ce qui reflète une attirance sur 5, mais au-delà, la qualité des informations chute rapidement, et la sous-déclaration va dépasser les 20% à partir de 6 ans avant l'enquête. Les résultats indiquent aussi que dans l'ensemble, la sous-déclaration des décès est un peu plus importante pour les frères que pour les soeurs.

– Enfin, et de façon quelque peu annexe, nous proposons une procédure d'appariement permettant d'identifier les frères et soeurs adultes éligibles au questionnaire individuel EDS et vivant sous un même toit.

Perspectives en termes d'analyses

Au terme de cette recherche doctorale, nous pouvons identifier plusieurs pistes de développements.

1. Creuser davantage la question de la qualité des données sur les frères et soeurs retiendra notre attention pour nos recherches futures. Notamment, nous pouvons supposer que les enquêtés vont fournir des informations de meilleure qualité pour les frères et soeurs qui leur sont *proches* dans la fratrie - leur frère/soeur aîné(e) ou cadet(te) - par rapport aux frères et soeurs *éloignés* dans la fratrie [20]. Ceci pourrait être testé, afin de mieux situer où se situent les principaux problèmes de données (erreurs sur les âges et les âges aux décès, etc.). Dans le même esprit, comme différents enquêtés sont parfois interviewés dans le même ménage, il serait possible d'utiliser la procédure que nous avons développée pour apparier les fratries afin de confronter leurs déclarations respectives pour en examiner la cohérence. Il conviendrait aussi d'examiner les schémas par âge de la mortalité dégagés des estimations directes basées sur les frères et soeurs (Timaeus et Jasseh 2004).

2. Nous nous sommes concentrés ici sur les niveaux de la mortalité, sans accorder d'attention aux différentiels. Or, s'il est clair que l'Afrique subsaharienne fait aujourd'hui face à une crise sanitaire sans précédent, celle-ci n'affecte pas tous les groupes sociaux avec la même intensité. Et l'un des principaux intérêts des données sur la survie des proches est précisément qu'elles nous renseignent, indirectement, sur les *caractéristiques du cercle familial des personnes décédées*. En l'absence d'informations sur les personnes soumises au risque, il reste possible de recourir aux informations disponibles sur les individus interrogés, en supposant que certaines caractéristiques

20. Si l'enquêté est de rang 4, les déclarations qu'il va fournir sur les frères et soeurs de rang 3 et 5 sont susceptibles d'être de meilleure qualité que celles fournies sur les frères et soeurs de rang 1,2 et 6,7

sont partagées entre frères et soeurs ou transmises des parents aux enfants.

La première recherche basée sur cette approche a porté sur la mortalité maternelle, à partir d'une dizaine d'enquêtes EDS (Graham et al. 2004). Les auteurs ont qualifié l'approche de "technique familiale", et ils ont montré que les risques de décès maternels étaient étroitement associés aux caractéristiques des soeurs enquêtées. Cette approche a également été testée dans les contextes de conflits, notamment par de Walque (2005) et de Walque et Verwimp (2010). Nous l'avons nous-même appliquée pour mettre à jour de profondes inégalités ethniques de mortalité adulte à partir des données sur la fratrie (Masquelier 2008). Cette approche pourrait être étendue aux inégalités géographiques. On peut en effet faire l'hypothèse que les enfants vivent dans un environnement proche de celui dans lequel vivent (ou ont vécu) leurs parents - surtout lorsque ces enfants sont encore mineurs. Il en va de même pour les frères et soeurs, bien que cela nécessite plus de discussion.

3. Une troisième catégorie de perspectives de recherche consiste à reconnaître que *les risques de décès au sein des fratries ne sont pas indépendants*. En d'autres termes, les risques de décès de deux frères sont plus proches que les risques de décès de deux personnes prises au hasard (en raison notamment de caractéristiques familiales partagées). Or ceci peut amener à mettre en question les procédures d'estimation classique. A la répétition des fratries induites par la structure des questionnaires s'ajoute une éventuelle corrélation des risques de décès des différents frères et soeurs au sein d'une même fratrie. Ce "clustering" des décès est bien documenté pour la survie des enfants mais peu exploré dans le cas de la mortalité des adultes. N'est-il pas nécessaire de tenir compte de cette hétérogénéité des fratries en recourant à une analyse multi-niveaux ? Est-il possible, au-delà du niveau individuel, d'expliquer une fraction des risques de décès en faisant intervenir un niveau relatif à la fratrie ?

4. Une quatrième piste de développements futurs concerne deux types de causes de décès pour lesquelles les données d'enquête permettent de mener une analyse spécifique : la *mortalité maternelle* et la *mortalité violente*. L'analyse des décès maternels et de la fraction qu'ils représentent dans l'ensemble des décès est possible à partir des données portant sur les frères et soeurs. Deux récentes séries d'estimation ont font usage : celles distribuées par l'OMS en 2010 (WHO, UNICEF, UNFPA and The World Bank 2010) et celles publiées par Hogan et al. (2010). Il nous paraît opportun d'examiner plus en détail les biais de sélection spécifiques à ce type de données, notamment par le biais des simulations. La mortalité violente peut également être analysée à partir des données sur les frères et soeurs (de Walque 2005, de Walque et Verwimp 2010, Obermeyer et al. 2008, Iraq Family Health Survey Study Group 2008). Là encore, des biais de sélection

bien spécifiques sont susceptibles de caractériser ces données, et ceci pourrait faire l'objet de simulations également.

5. Enfin, il serait très utile de développer des schémas de mortalité par âge standards qui tiendraient compte de l'excès de mortalité associé au VIH-sida et que l'on pourrait faire évoluer dans le temps (au moyen d'un système logit, par exemple) en intégrant des informations sur la prévalence du VIH. Le recours aux données du réseau INDEPTH pourraient sans doute le permettre.

Perspectives en termes de collecte

A plusieurs reprises, nous avons rappelé que le premier frein à l'estimation précise des risques de décès adultes se situait au niveau de la collecte des données d'enquête. Une série de changements à apporter aux modes de collecte utilisés actuellement pourrait grandement améliorer nos capacités d'analyse.

1. *Collecter des informations sur l'année du décès des parents* : Chackiel et Orellana (1985) l'ont suggéré dès les années 1980, mais ceci n'a jamais été intégré aux grands programmes d'enquête ni aux recensements. Dater les décès des parents permettrait d'affiner la reconstitution des tendances, un enjeu important aujourd'hui en raison des reprises de mortalité encore mal appréhendées.
2. *Collecter davantage de données sur la survie des parents au mariage ou à la première naissance* : là encore, des développements méthodologiques existent, mais ils ont connu peu d'applications pratiques, en raison du manque de données.
3. *Dans les EDS et les MICS, étendre les questions sur la survie des parents à tous les membres du ménage*, et pas uniquement aux enfants de moins de 15 ans. Ceci permettrait (1) d'obtenir des estimations plus anciennes lorsqu'une seule opération de collecte est disponible, (2) d'accroître la fiabilité des données - car les déclarations qui concernent les enfants sont davantage sujettes à caution, en raison de l'"effet d'adoption", et (3) de réduire les biais de sélection induits par le VIH-sida.
4. *Identifier les soeurs et les frères qui sont interrogés dans le même ménage*, notamment pour permettre des tests de qualités des données plus poussés, ou pour corriger les déclarations.

Bibliographie

O. Aalen. Phase Type Distributions in Survival Analysis. *Scandinavian Journal of Statistics*, 22 (4) :447–463, 1995. (Cité page 424.)

A. Abbott. Sequence analysis : new methods for old ideas. *Annual Review of Sociology*, 21 :93–113, 1995. (Cité page 390.)

M. Ainsworth. Economic aspects of child fostering in Cote d'Ivoire. *Research in Population Economics*, 8 :25–62, 1996. (Cité page 279.)

M. Ainsworth et D. Filmer. Inequalities in children's schooling : AIDS, orphanhood, poverty, and gender. *World Development*, 34(6) :1099 – 1128, 2006. (Cité pages 164 et 195.)

R. M. Anderson, R. M. May, T. W. Ng, et J. T. Rowley. Age-Dependent Choice of Sexual Partners and the Transmission Dynamics of HIV in Sub-Saharan Africa . *Philosophical Transactions : Biological Sciences*, 336 (1277) :135–155, 1992. (Cité page 75.)

Ph. Antoine et Ph. Bocquier. Une méthode de recueil de la parenté dans les enquêtes démographiques : un compromis avec l'approche anthropologique. *Cahiers québécois de démographie*, Volume 21, numéro 1 :7–27, 1992. (Cité page 279.)

T. Araya, G. Reniers, A. Schaap, D. Kebede, A. Kumie, N. Nagelkerke, R. Coutinho, et E. Sanders. Lay diagnoses of causes of death for monitoring AIDS mortality in Addis Ababa, Ethiopia. *Tropical Medicine and International Health*, 9(1) :178–186, 2004. (Cité page 18.)

B. Auvert, G. Buonamico, E. Lagarde, et B. Williams. Sexual Behavior, Heterosexual Transmission, and the Spread of HIV in Sub-Saharan Africa : A Simulation Study. *Computers and Biomedical Research*, 33(1) :84 – 96, 2000. (Cité page 75.)

J-F. Bayart. *L'Etat en Afrique : La politique du ventre*. Fayard, 2006. (Cité page 1.)

K. Beegle, D. Filmer, A. Stokes, et L. Tiererova. Orphanhood and the Living Arrangements of Children in Sub-Saharan Africa. *World Development*, 38(12) :1727–1746, 2010. (Cité pages 195 et 287.)

K. Beegle et S. Krutikova. Adult mortality and children's transition into marriage. *Demographic Research*, 19(42) :1551–1574, 2008. (Cité page 159.)

N. Bennett et S. Horiuchi. Mortality estimation from registered deaths in less developed countries. *Demography*, 21(2) :217–234, 1984. (Cité pages 6 et 347.)

P. Bhat. Estimating Transition Probabilities of Age Misstatement. *Demography*, 27(1) :pp. 149–163, 1990. (Cité page 303.)

G. Bicego. Estimating adult mortality rates in the context of the AIDS epidemic in sub-Saharan Africa : analysis of DHS sibling histories. *Health Transition Review*, 2 :7–22, 1997. (Cité pages 226 et 295.)

G. Bicego, S. Curtis, H. Raggers, S. Kapiga, et S. Ngallaba. *Sumve survey on adult and childhood mortality, Tanzania, 1995 : in-depth study on estimating adult and childhood mortality in settings of high adult mortality*. Macro International, 1997. (Cité pages 165, 166 et 303.)

G. Bicego, S. Rutstein, et K. Johnson. Dimensions of the emerging orphan crisis in sub-Saharan Africa. *Social Science & Medicine*, 56(6) :1235 – 1247, 2003. (Cité pages 87, 164, 182, 287, 443, 444, 445 et 446.)

J. Blacker. The Estimation of Adult Mortality in Africa from Data on Orphanhood. *Population Studies*, 31(1) :107–128, 1977. (Cité pages 20, 130 et 145.)

J. Blacker. Experience in the use of special mortality questions in multipurpose surveys : the single-round approach. Dans United Nations (ed), *Data bases for mortality measurement*, Chapitre IX, pages 79–89. United Nations, New York, 1984. (Cité pages 20, 22, 170, 171, 173, 182, 278, 281 et 288.)

J. Blacker et J. Mukiza-Gapere. The indirect measurement of adult mortality in Africa : results and prospects. Dans International Union for the Scientific Study of Population (ed), *African Population Conference, Dakar, Senegal, November 7-12, 1988*, pages 23–38. IUSSP, 1988. (Cité pages 21, 130 et 282.)

J. Bongaarts. The formal demography of families and households : An overview. *IUSSP Newsletter*, 17 :27–42, 1983. (Cité page 44.)

J. Bongaarts. The Fertility Impact of Changes in the Timing of Childbearing in the Developing World. *Population Studies*, 53(3) :277–289, 1999. (Cité page 137.)

H. Booth. Transforming Gompertz's Function for Fertility Analysis : The Development of a Standard for the Relational Gompertz Function. *Population Studies*, 38(3) :495–506, 1984. (Cité pages 113, 114 et 272.)

D. Bradshaw et I. Timaeus. Levels and Trends of Adult Mortality. Dans D. Jamison, R. Feachem, M. Makgoba, E. Bos, F. Baingana, K. Hofman, et K. Rogo (eds), *Disease and Mortality in Sub-Saharan Africa*, Chapitre 4, pages 31–42. World Bank, 2003. (Cité pages 11, 312, 318 et 334.)

P. Braitstein, M. Brinkhof, F. Dabis, et M. et al. Schechter. Mortality of HIV-1-infected patients in the first year of antiretroviral therapy : comparison between low-income and high-income countries. *Lancet*, 367(9513) :817–824, Mar 2006. (Cité page 92.)

W. Brass. On the scale of mortality. Dans W. Brass (ed), *Biological aspects of demography*, pages 69–110. Taylor et Francis, London, 1971. (Cité page 112.)

W. Brass. Perspectives in population prediction : illustrated by the statistics of England and Wales. *Journal of the Royal Statistical Society*, 137 : 532–583, 1974. (Cité page 113.)

W. Brass et E. Bamgboye. The time location of reports of survivorship : estimates for maternal and paternal orphanhood and the ever-widowed. Centre for Population Studies Research Paper, 81-1, 1981. (Cité pages 48, 143, 144, 145, 147, 148, 184, 316 et 430.)

W. Brass, A. Coale, P. Demeny, D. Heisel, F. Lorimer, A. Romaniuk, et E. van de Walle. *The Demography of tropical Africa*. Princeton University Press, 1968. (Cité pages 58 et 414.)

W. Brass et K. Hill. Estimating adult mortality from orphanhood. Dans International Union for the Scientific Study of Population (ed), *Proceedings of the International Population Conference : Liège*, volume 3, pages 111–123, 1973. (Cité pages 6, 101, 110, 111, 119, 122, 123, 125, 126, 127, 129 et 132.)

M. Brockerhoff et P. Hewett. Inequality of child mortality among ethnic groups in sub-Saharan Africa. *Bulletin of the World Health Organization*, 78 :30–41, 2000. (Cité page 171.)

T. Brown, N. Grassly, G. Garnett, et K. Stanecki. Improving projections at the country level : the UNAIDS Estimation and Projection Package 2005. *Sexually Transmitted Infections*, 82 :iii34–iii40, 2006. (Cité page 80.)

T. Brown, J. Salomon, L. Alkema, A. Raftery, et E. Gouws. Progress and challenges in modelling country-level HIV/AIDS epidemics : the UNAIDS Estimation and Projection Package 2007. *Sexually Transmitted Infections*, 84 (S1) :i5–i–10, 2008. (Cité page 66.)

W. J. Browne, S. V. Subramanian, K. Jones, et H. Goldstein. Variance Partitioning in Multilevel Logistic Models That Exhibit Overdispersion. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 168(3) : 599–613, 2005. (Cité page 228.)

S. Brunel. *L'Afrique - Un continent en réserve de développement*. Bréal, 2003. (Cité page 1.)

J. Caldwell et A. Igun. An Experiment with Census-Type Age Enumeration in Nigeria. *Population Studies*, 25(2) :pp. 287–302, 1971. (Cité page 303.)

A. Cameron et K. Trivedi. *Regression Analysis of Count Data*. Cambridge University Press, 1998. (Cité page 226.)

P. Cantrelle, I. Diop, M. Garenne, M. Gueye, et A. Sadio. The Profile of Mortality and Its Determinants in Senegal, 1960-1980. Dans United Nations (ed), *Determinants of Mortality Change and Differentials in Developing Countries : The Five-Country Case Study Project*. Department of International Economic and Social Affairs. Population Division andWorld Health

Organization. Division of Health Statistics. New York, NY, United Nations., 1986. (Cité page 321.)

A. Case, C. Paxson, et J. Ableidinger. Orphans in Africa : Parental Death, Poverty, and School Enrollment. *Demography*, 41(3) :483–508, 2004. (Cité pages 164, 195 et 287.)

G. Cauthen, A. Pio, et H. Dam. Annual risk of tuberculous infection (first ed. 1988). *Bulletin of the World Health Organisation*, 80 (6) :503–511, 2002. (Cité page 2.)

Centre of African studies. *African historical demography : proceedings of a seminar held in the Centre of African Studies, University of Edinburgh, 24th and 25th April 1981*. University of Edinburgh, 1981. (Cité page 58.)

J. Chackiel et H. Orellana. Adult female mortality trends from retrospective questions about maternal orphanhood included in census and surveys. Dans *International Population Conference - Florence*, volume 4. IUSSP, Liège, 1985. (Cité pages 100, 148, 149 et 355.)

S. Clark. Demographic Impacts of the HIV Epidemic and Consequences of Population-Wide Treatment of HIV for the Elderly : Results from Microsimulation. Dans Panel on Policy Research and Data Needs to Meet the Challenge of Aging in Africa, B. Cohen, et J. Menken (eds), *Aging in Sub-Saharan Africa : Recommendations for Furthering Research*, Chapitre 3, pages 92–116. National Research Council, 2006. (Cité pages 56 et 75.)

S. Clark, M. Collinson, K. Kahn, K. Drullinger, et S. Tollman. Returning home to die : Circular labour migration and mortality in South Africa. *Scandinavian Journal of Public Health*, 35 (3) supp 69 :35–44, 2007. (Cité page 2.)

S. Clark et J. Eaton. Demographic Consequences of HIV Epidemics and Effects of Different Male Circumcision Intervention Designs : Suggestive Findings from Microsimulation. Working Paper no. 85, Center for Statistics and the Social Sciences, University of Washington, 2008. (Cité page 75.)

A. Coale. A New Method for Calculating Lotka's r–the Intrinsic Rate of Growth in a Stable Population. *Population Studies*, 11(1) :92–94, 1957. (Cité pages 60 et 124.)

A. Coale. Estimates of Fertility and Mortality in Tropical Africa. *Population Index*, 32(2) :pp. 173–181, 1966. (Cité page 58.)

A. Coale. Age Patterns of Marriage. *Population Studies*, 25 (2) :193–214, 1971. (Cité page 412.)

A. Coale, P. Demeny, et B. Vaughn. *Regional model life tables and stable populations (2nd ed.)*. New York : Academic Press, 1983. (Cité pages 182 et 230.)

A. Coale et E. Kisker. Defects in data on old age mortality in the United States : New procedures for calculating approximately accurate mortality schedules and life tables at the highest ages. *Asian and Pacific Population Forum*, 4 :1–31, 1990. (Cité page 68.)

A. Coale et D. McNeil. The Distribution by Age of the Frequency of First Marriage in Female Cohort. *Journal of American Statistical Association*, 67 :743–749, 1972. (Cité pages 75, 159, 412, 413 et 417.)

A. Coale et J. Trussell. Model fertility schedules : variations in the structure of childbearing in human populations. *Population index*, 40 (2) :185–258, 1974. (Cité pages 113, 124, 125 et 148.)

J. Coussy et J. Vallin. *Crise et population en Afrique. Crises économiques, programmes d'ajustement et dynamiques démographiques*. CEPED, 1995. (Cité page 1.)

A. Crampin, S. Floyd, J. Glynn, N. Madise, A. Nyondo, M. Khondowe, C. Njoka, H. Kanyongoloka, B. Ngwira, B. Zaba, et P. Fine. The long-term impact of HIV and orphanhood on the mortality and physical well-being of children in rural Malawi. *AIDS*, 17(3) :389–397, 2003. (Cité pages 164 et 165.)

I. Danel, W. Graham, P. Stupp, et P. Castillo. Applying the sisterhood method for estimating maternal mortality to health facility-based sample : a comparison with results from a household-based sample. *International Journal of Epidemiology*, 25 (5) :1017–1022, 1996. (Cité pages 32 et 271.)

P. David, S. Kawar, et W. Graham. Estimating maternal mortality in Djibouti : an application of the sisterhood method. *International Journal of Epidemiology*, 20 (2) :551–557, 1991. (Cité pages 32 et 271.)

D. de Walque. Selective Mortality during the Khmer Rouge Period in Cambodia. *Population and Development Review*, 31(2) :351–368, 2005. (Cité pages 260, 309 et 354.)

D. de Walque et P. Verwimp. The Demographic and Socio-economic Distribution of Excess Mortality during the 1994 Genocide in Rwanda. *Journal of African Economies*, 19(2) :141–162, 2010. (Cité pages 238, 245, 250, 260 et 354.)

A. Diop. *La famille wolof*. Karthala, 1985. (Cité page 278.)

R. Dorrington, D. Bourne, D. Bradshaw, R. Laubscher, et I. Timaeus. The impact of HIV/AIDS on adult mortality in South Africa. Technical Report, Burden of Disease Research Unit Medical Research Council, 2001. (Cité pages 14, 15, 16 et 334.)

R. Dorrington, Johnson L., Bradshaw D., et Daniel T. *The Demographic Impact of HIV/AIDS in South Africa. National and Provincial Indicators for 2006*. Centre for Actuarial Research, South African Medical Research Concil and Actuarial Society of South Africa, 2006a. (Cité pages 75 et 94.)

R. Dorrington, T. Moultrie, et T. Daniel. *The Demographic Impact of HIV/AIDS in Botswana*. UNDP and NACA, Gaborone, Botswana, 2006b. (Cité page 67.)

R. Dorrington, T. Moultrie, et I. Timaeus. *Estimation of mortality using the South African Census 2001 data*. CARe Monograph 11, 2004. (Cité pages 286 et 334.)

R. Dorrington, I. Timaeus, et S. Gregson. Adult mortality in Southern Africa using deaths reported by households : some methodological issues and results. Paper presented at the Population Association of America, 2006 Annual Meeting, Los Angeles, California, March 30 - April 1, 2006c. (Cité pages 8, 17 et 347.)

G. Duthé et G. Pison. Adult mortality in a rural area of Senegal : Non-communicable diseases have a large impact in Mlomp. *Demographic Research*, 19(37) :1419–1434, 2008. (Cité page 3.)

C. Dye, S. Schelle, D. Polin, V. Pathania, et M. Raviglione. Global Burden of Tuberculosis : Estimated Incidence, Prevalence and Mortality by Country. *Journal of the American Medical Association*, 282 :677–686, 1999. (Cité page 2.)

D. Evans et E. Miguel. Orphans and Schooling in Africa : A Longitudinal Analysis. *Demography*, Volume 44, Number 1 :pp. 35–57, 2007. (Cité page 195.)

D. Ewbank. Age misreporting and age-selective underenumeration : Sources, patterns and consequences for demographic analysis. Committee on Population and Demography, Report No. 4, 1981. (Cité pages 43, 288 et 289.)

FAO. *L'état de l'insécurité alimentaire dans le monde 2008*. FAO, 2008. (Cité page 3.)

J. Faraway. *Extending the linear model with R*. Tyalor & Francis, 2006. (Cité page 408.)

G. Feeney. The impact of HIV/AIDS on adult mortality in Zimbabwe. *Population and Development Review*, 27, 4 :771–780, 2001. (Cité pages 14, 15, 17, 235, 260 et 335.)

F. Font, M. Alonso Gonzalez, R. Nathan, F. Lwilla, J. Kimario, M. Tanner, et P. Alonso. Maternal mortality in a rural district of southeastern Tanzania : an application of the sisterhood method. *International Journal of Epidemiology*, 29 (1) :107–112, 2000. (Cité pages 32 et 271.)

K. Ford et V. Hosegood. AIDS Mortality and the Mobility of Children in Kwazulu Natal, South Africa. *Demography*, 42(4) :757–768, 2005. (Cité page 163.)

N. Gaffga, R. Tauxe, et E. Mintz. Cholera : A New Homeland in Africa ? *American Journal of Tropical Medicine and Hygiene*, 77(4) :705–713, 2007. (Cité page 2.)

E. Gakidou, M. Hogan, et A. Lopez. Adult mortality : time for a reappraisal. *International Journal of Epidemiology*, 33, 4 :710–717, 2004. (Cité pages 8, 9, 297 et 343.)

E. Gakidou et G. King. Death by survey : estimating adult mortality without selection bias from sibling survival data. *Demography*, 43 (3) : 569–585, 2006. (Cité pages 245, 246, 247, 248, 249, 250, 251, 254, 255, 257, 258, 260, 261, 262, 264, 266, 267, 350 et 352.)

M. Garenne. Sex ratios at birth in African populations : a review of survey data. *Human Biology*, 74 (6) :889–900, 2002. (Cité page 297.)

M. Garenne. Fertility Changes in Sub-Saharan Africa. DHS Comparative Reports No. 18, Calverton, Maryland, USA : Macro International Inc., 2008. (Cité page 58.)

M. Garenne et F. Friedberg. Accuracy of Indirect Estimates of Maternal Mortality : A Simulation Model. *Studies in Family Planning*, 28(2) :pp. 132–142, 1997. (Cité pages 32, 271 et 439.)

M. Garenne, M. Madison, D. Tarantola, B. Zanou, J. Aka, et R. Dogore. *Conséquences démographiques du sida en Abidjan : 1986-1992*. Les études du CEPED, 10, 1995. (Cité page 15.)

P. Ghys, N. Walker, W. McFarland, R. Miller, et G. Garnett. Improved data, methods and tools for the 2007 HIV and AIDS estimates and projections. *Sexually Transmitted Infections*, 84 (S1) :i1–i4, 2008. (Cité page 66.)

P. Ghys, B. Zaba, et M. Prins. Survival and mortality of people infected with HIV in low and middle income countries : results from the extended ALPHA network. *AIDS*, 21 Suppl 6 :S1–S4, Nov 2007. (Cité page 88.)

M. Gibril. *Appréciation des erreurs d'observation dans les recensements : une étude de cas de la Gambie*. OECD, 1979. (Cité page 288.)

N. Goldman. Estimating the Intrinsic Rate of Increase of a Population from the Average Numbers of Younger and Older Sisters. *Demography*, 15(4) :pp. 499–507, 1978. (Cité pages 272 et 298.)

L. Goodman, N. Keyfitz, et T. Pullum. Family formation and the frequency of various kinship relationships. *Theorical Population Biology*, 5(3) :1–27, 1974. (Cité pages 103 et 105.)

E Goody. *Parenthood and Social Reproduction : Fostering and Occupational Roles in West Africa*. Cambridge University Press, 1982. (Cité page 279.)

E. Gouws, V. Mishra, et T. Fowler. Comparison of adult HIV prevalence from national population-based surveys and antenatal clinic surveillance in countries with generalised epidemics : implications for calibrating surveillance data. *Sexually Transmitted Infections*, 84 (S1) :i17–i23, 2008. (Cité page 80.)

W. Graham, S. Ahmed, C. Stanton, C. Abou-Zahr, et O. Campbell. Measuring maternal mortality : An overview of opportunities and options for developing countries. *BMC Medicine*, 6(1) :12, 2008. (Cité page 30.)

W. Graham, W. Brass, et R. Snow. Estimating maternal mortality : the sisterhood method. *Studies in Family Planning*, 20(3) :125–135, 1989. (Cité pages 19, 32, 101, 267, 270, 271, 272, 439 et 440.)

W. Graham, D. Phila, A. Fitzmaurice, J. Bell, et J. Cairns. The familial technique for linking maternal death with poverty. *The Lancet*, Volume 363, Issue 9402, 3 :23–27, 2004. (Cité page 354.)

N. Grassly, J. Lewis, M. Mahy, N. Walker, et I. Timaeus. Comparison of household-survey estimates with projections of mortality and orphan numbers in sub-Saharan Africa in the era of HIV/AIDS. *Population Studies*, 58 :2 :207–217, 2004. (Cité pages 180, 182, 183, 281 et 351.)

N. Grassly et I. Timaeus. Methods to estimate the number of orphans as a result of AIDS and other causes in sub-Saharan Africa. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, 39 :635–75, 2005. (Cité pages 11, 44, 87, 95, 195, 197, 207, 208, 209, 211, 212, 213, 214, 215, 216, 222 et 349.)

P. Green et B. Silverman. *Nonparametric regression and generalised linear models*. Champan and Hall, 1994. (Cité page 408.)

S. Gregson, G. Garnett, et R. Anderson. Assessing the Potential Impact of the HIV-1 Epidemic on Orphanhood and the Demographic Structure of Populations in sub-Saharan Africa. *Population Studies*, 48(3) :438–458, 1994. (Cité pages 75 et 195.)

J. Hajnal. Age at marriage and proportion marrying. *Population Studies*, 7, 2 :111–136, 1953. (Cité pages 414 et 417.)

T. Hallett, S. Gregson, J. Lewis, B. Lopman, et G. Garnett. Behaviour change in generalised HIV epidemics : impact of reducing cross-generational sex and delaying age at sexual debut. *Sexually Transmitted Infections*, 83(suppl 1) :i50–i54, 2007. (Cité pages 83 et 423.)

T. Hallett, B. Zaba, J. Todd, B. Lopman, W. Mwita, S. Biraro, S. Gregson, et J. Boerma. Estimating Incidence from Prevalence in Generalised HIV Epidemics : Methods and Validation. *PLoS Med*, 5(4) :e80, 2008. (Cité pages 85, 90, 91 et 92.)

E. Hammel, D. Hutchinson, K. Wachter, R. Lundry, et R. Deuel. The SOCSIM demographic-sociological microsimulation program operating manual. Research series, University of California, Berkeley, Institute of International Studies, ; no. 27, 1976. (Cité page 50.)

V. Hertrich. Nuptialité et rapports de genre en Afrique. Tendances de l'entrée en union, 1950-99. Dans T Locoh (ed), *Genre et société en Afrique*, pages 281–307. Paris, Ined, Les Cahiers de l'Ined, n° 160, 2007. (Cité pages 412, 414 et 415.)

P. Heuveline. HIV and population dynamics : a general model and maximum-likelihood standards for East Africa. *Demography*, 40(2) :217–245, 2003. (Cité pages 7, 9 et 75.)

K. Hill. Estimating Adult Mortality Levels from Information on Widowhood. *Population Studies*, 31(1) :75–84, 1977. (Cité page 6.)

K. Hill. A proposal for the use of information on residence of siblings to estimate emigration by age. IUSSP Papers, No. 18, pp. 19-34, 1981. (Cité page 267.)

K. Hill. An evaluation of indirect methods for estimating mortality. Dans J. Vallin, J. Pollard, et H. Heligman (eds), *Methodologies for the collection and analysis of mortality data*. Ordina, IUSSP, Liège, 1984. (Cité pages 29, 282 et 288.)

K. Hill. Estimating Census and Death Registration Completeness. *Asian and Pacific Population Forum*, 1(3) :8–13, 1987. (Cité pages 6 et 347.)

K Hill. The measurement of adult mortality : An assessment of data availability, data quality and estimation methods. Dans Chamie J. et Cliquet R. (eds), *Health and Mortality Issues of Global Concern*, pages 72–83. Brussels : Population and Family Study Center, CBGS, and New York : Population Division, United Nations Secretariat, 1998. (Cité page 345.)

K. Hill. Indirect Estimation Methods. Dans G. Caselli, G. Wunsch, et J. Vallin (eds), *Demography : Analysis and Synthesis : A Treatise in Population*, volume 1, Chapitre 127. Academic Press, 2005. (Cité page 20.)

K. Hill et A. Amouzou. Trends in Child Mortality 1960 to 2000. Dans D. Jamison, E. Bos, F. Baingana, et R. Feachem (eds), *Disease and mortality in sub-Saharan Africa*, volume 1, Chapitre 3, pages 15–25. World Bank, 2006. (Cité page 5.)

K. Hill, C. Stanton, et N. Gupta. Measuring Maternal Mortality from a Census : Guidelines for Potential Users. MEASURE Evaluation Manual Series, No. 4, Carolina Population Center, University of North Carolina at Chapel Hill, 2001. (Cité page 31.)

K. Hill et J. Trussell. Further developments in indirect mortality estimation. *Population Studies*, 31(2) :313–334, 1977. (Cité pages 6, 19, 100, 101, 123, 124, 126, 128, 267, 269, 270, 271, 290, 297 et 308.)

K. Hill, D. You, et Y. Choi. Death distribution methods for estimating adult mortality : Sensitivity analysis with simulated data error. *Demographic Research*, 21 :235–254, 2009. (Cité pages 8, 16, 17, 343 et 347.)

M. Hogan, K. Foreman, M. Naghavi, S. Ahn, M. Wang, S. Makela, A. Lopez, R. Lozano, et Ch. Murray. Maternal mortality for 181 countries, 1980-2008 : a systematic analysis of progress towards Millennium Development Goal 5. *The Lancet*, 375(9726) :1609 – 1623, 2010. (Cité pages 30, 261, 266 et 354.)

INDEPTH. *Population and health in developing countries, Population, health and survival at INDEPTH sites*, volume 1. International Development Research Centre, 2002. (Cité page 18.)

INDEPTH. *INDEPTH model life tables for Sub-Saharan Africa*. Ashgate, 2004. (Cité pages 18, 19, 119, 120, 121 et 274.)

INSEE-SEDES. *Enquête démographique au Tchad 1964*. 1966. (Cité page 131.)

Iraq Family Health Survey Study Group. Violence-Related Mortality in Iraq from 2002 to 2006. *The New England Journal of Medicine*, 358(5) : 484–493, 2008. (Cité pages 245, 260, 266 et 354.)

U. Isiugo-Abanihe. Child Fosterage in West Africa. *Population and Development Review*, 11(1) :pp. 53–73, 1985. (Cité page 279.)

D. Jamison, E. Bos, F. Baingana, R. Feachem, M. Makgoba, K. Hofman, et H. Rogo (eds). *Disease and Mortality in Sub-Saharan Africa*. World Bank, 2006. (Cité page 2.)

A. Jasilioniene, D. Jdanov, T. Sobotka, E. Andreev, K. Zeman, V. Shkolnikov, J. Goldstein, D. Philipov, et G. Rodrigues. Methods protocol for the Human Fertility Database. Working paper, Max Planck Intitute for Demographic Research, 2009. (Cité page 411.)

L. Jiang. Changing Kinship Structure and its Implications for Old-Age Support in Urban and Rural China. *Population Studies*, 49(1) :127–145, 1995. (Cité page 50.)

N. Keyfitz. A life table that agrees with the data. *Journal of the American Statistical Association*, 63 (324) :305–312, 1966. (Cité page 68.)

N. Keyfitz et H. Caswell. The demographic theory of kinship. Dans N. Keyfitz et H. Caswell (eds), *Applied Mathematical Demography (3rd ed.)*, Chapitre 15, pages 370–398. Springer, 2005. (Cité pages 53, 103, 108, 268, 270, 299 et 301.)

D. Kirk et B. Pillet. Fertility Levels, Trends, and Differentials in Sub-Saharan Africa in the 1980s and 1990s. *Studies in Family Planning*, 29 (1) :1–22, 1998. (Cité page 139.)

E. Korenromp, B. Williams, E. Gouws, C. Dye, et R. Snow. Measurement of trends in childhood malaria mortality in Africa : an assessment of progress toward targets based on verbal autopsy. *THE LANCET Infectious Diseases*, 3 :349–359, 2003. (Cité page 3.)

A. Kostaki et V. Panousis. Expanding an abridged life table. *Demographic Research*, 5(1) :1–22, 2001. (Cité page 68.)

H. Le Bras. Parents, grands-parents bisaieux. *Population*, 28(1) :9–38, 1973. (Cité page 103.)

M. Lech et A. Zwane. Survey on maternal mortality in Swaziland using the sisterhood method. *Paediatric and Perinatal Epidemiology*, 16(2) :101–107, 2002. (Cité pages 32 et 271.)

P. Leclerc, A. Matthews, et M. Garenne. Fitting the HIV Epidemic in Zambia : A Two-Sex Micro-Simulation Model. *PLoS ONE*, 4(5) :e5439, 05 2009. (Cité page 75.)

R. Lee et L. Carter. Modeling and Forecasting U.S. Mortality. *Journal of the American Statistical Association*, 87 (419) :659–671, 1992. (Cité pages 54 et 406.)

R. Lesthaeghe, G. Kaufmann, et D. Meekers. The Nuptiality Regimes in Sub-Saharan Africa. Dans R. Lesthaeghe (ed), *Reproduction and Social Organization in Sub-Saharan Africa*, Chapitre 6, pages 238–251. Berkeley : University of California Press, 1989. (Cité page 415.)

J. Lewis, C. Ronsmans, A. Ezeh, et S. Gregson. The population impact of HIV on fertility in sub-Saharan Africa. *AIDS*, 18 (S2) :35–43, 2004. (Cité pages 2, 80, 94, 95 et 198.)

I. Longini, W. Scott, R. Byers, J. Ward, W. Darrow, G. Lemp, et H. Hethcote. Statistical analysis of the stages of HIV infection using a Markov model. *Statistics in Medicine*, Volume 8 Issue 7 :Pages 831 – 843, 1989. (Cité page 424.)

D. Loomis, D. Richardson, et L. Elliot. Poisson regression analysis of ungrouped data. *Occupational and environmental medicine*, 62 :625–629, 2005. (Cité page 226.)

A. Lotka. Orphanhood in Relation to Demographic Factors. *Metron*, 2 : 37–109, 1931. (Cité pages 103 et 104.)

C. Makinson. Estimates of adult mortality in Burundi. *Journal of Biosocial Science*, 25(02) :169–186, 1993. (Cité pages 20, 23 et 149.)

C. Malaker. Estimation of Adult Mortality in India : 1971-1981. *Demography India*, 15 :126–136, 1986. (Cité page 21.)

M. Marston, J. Todd, J. Glynn, K. Nelson, R. Rangsin, T. Lutalo, M. Urassa, S. Biraro, L. Van der Paal, P. Sonnenberg, et B. Zaba. Estimating 'net' HIV-related mortality and the importance of background mortality rates. *AIDS*, 21 :–, 2007. (Cité page 89.)

M. Marston, B. Zaba, J. Salomon, H. Brahmbhatt, et D. Bagenda. Estimating the Net Effect of HIV on Child Mortality in African Populations Affected by Generalized HIV Epidemics. *Journal of Acquired Immune Deficiency Syndromes*, 38 (2) :219–227, 2005. (Cité pages 94 et 208.)

C Mason. Socsim Oversimplified. Note technique distribuée à l'occasion de la Summer School "Microsimulation using Socsim" organisée au Max Planck, Rostock en juin 2010 (note disponible en ligne : http ://lab.demog.berkeley.edu/socsim/, consultée le 28 juin 2010), 2010. (Cité page 51.)

B. Masquelier. Minorités ethniques et mortalité adulte en Afrique Subsaharienne. Dans *Démographie et Cultures (Actes du colloque de Québec, 2008)*, pages 835–852. Paris, INED, 2008. (Cité pages 171 et 354.)

C. Mathers, D. Fat, M. Inoue, C. Rao, et A. Lopez. Counting the dead and what they died from : an assessment of the global statuts of cause of death data. *Bulletin of the World Health Organization*, 83 :171–177, 2005. (Cité pages 14 et 15.)

G. Mbaruku, F. Vork, D. Vyagusa, R. Mwakipiti, et J. Roosmalen. Estimates of Maternal Mortality in Western Tanzania by the Sisterhood Method. *African Journal of Reproductive Health*, 7(3) :pp. 84–91, 2003. (Cité pages 32 et 271.)

P. McCullagh et J. A. Nelder. *Generalized Linear Models*. London, Chapman & Hall/CRC, 1989. (Cité pages 227 et 228.)

C. K. McDaniel et E. A. Hammel. A Kin-Based Measure of r and an Evaluation of its Effectiveness. *Demography*, 21(1) :41–51, 1984. (Cité pages 50 et 272.)

R. Monasch et J. Boerma. Orphanhood and childcare patterns in sub-Saharan Africa : an analysis of national surveys from 40 countries. *AIDS*, 18(2) :55–65, 2004. (Cité pages 2, 24, 163, 164, 195 et 284.)

K. Moser, V. Shkolnikov, et D. Leon. World mortality 1950-2000 : divergence replaces convergence from the late 1980s. *Bulletin of the World Health Organization*, 83(3) :202–209, 2005. (Cité page 3.)

T. Mulder et P. Johnson. Analysis of demographic models used to incorporate HIV/AIDS-related mortality. Paper presented at the XXV International Population Conference, IUSSP, Tours, France, July 18-23, 2005. (Cité page 67.)

M. Murphy. Tracing Very Long-Term Kinship Networks Using SOCSIM. *Demographic research*, 10 (7) :171–196, 2004. (Cité page 50.)

M. Murphy. The role of assortative mating on population growth in contemporary developed societies. Dans F. Billari, T. Fent, A. Prskawetz, et J. Scheffran (eds), *Agent-based computational modelling : applications in demography, social, economic and environmental sciences*, Chapitre 4, pages 61–84. Physica Verlag, Heidelberg, Germany, 2006. (Cité pages 50, 401 et 402.)

Ch. Murray, B. Ferguson, A. Lopez, M. Guillot, J. Salomon, et O. Ahmad. Modified logit life table system : principles, empirical validation, and application. *Population Studies*, 57(2) :165–182, 2003. (Cité page 311.)

Ch. Murray, Th. Laakso, K. Shibuya, K. Hill, et A. Lopez. Can we achieve Millennium Development Goal 4 ? New analysis of country trends and forecasts of under-5 mortality to 2015. *LANCET*, 370(9592) :1040–1054, 2007. (Cité page 234.)

Ch. Murray, J.l Rajaratnam, J. Marcus, Th. Laakso, et A. Lopez. What Can We Conclude from Death Registration ? Improved Methods for Evaluating Completeness. *PLoS Med*, 7(4) :e1000262, 04 2010. (Cité pages 8, 16, 17, 343 et 347.)

J. Nakiyingi, M. Bracher, J. Whitworth, A. Ruberantwari, J. Busingye, S. Mbulaiteye, et B. Zaba. Child survival in relation to mother's HIV infection and survival : evidence from a Ugandan cohort study. *AIDS*, 17(12) :1827–34, 2003. (Cité pages 98, 164, 165 et 195.)

J. Ng'weshemi, M. Urassa, R. Isingo, G. Mwaluko, J. Ngalula, T. Boerma, M. Marston, et B. Zaba. HIV Impact on Mother and Child Mortality in Rural Tanzania. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, 33(3) :393–404, 2003. (Cité pages 98, 164, 195 et 200.)

V. Notkola, I. Timaeus, et H. Siiskonen. Impact on mortality of the AIDS epidemic in northern Namibia assessed using parish registers. *AIDS*, 18(7) :1061–1065, 2004. (Cité page 18.)

Z. Obermeyer, C. Murray, et E. Gakidou. Fifty years of violent war deaths from Vietnam to Bosnia : analysis of data from the world health survey programme. *BMJ*, 336 :1482–1486, 2008. (Cité pages 37, 245, 260, 261, 309 et 354.)

Z. Obermeyer, J. Rajaratnam, C. Park, E. Gakidou, M. Hogan, A. Lopez, et Ch. Murray. Measuring Adult Mortality Using Sibling Survival : A New Analytical Method and New Results for 44 Countries, 1974-2006. *PLoS Medicine*, 7(4) :e1000260, 04 2010. (Cité pages 8, 233, 245, 250, 252, 253, 254, 255, 260, 261, 262, 263, 264, 265, 266, 267, 305, 307, 343, 350, 352 et 353.)

C. O'Donoghue. Dynamic Microsimulation : A Methodological Survey. *Brazilian Electronic Journal of Economics*, 4(2), 2001. (Cité page 50.)

A. Omran. The epidemiologic transition : a theory of the epidemiology of population change. *Milbank Memorial Fund Quarterly*, 49 (4) :p. 509–538, 1971. (Cité page 3.)

K. Orroth, E. Freeman, R. Bakker, A. Buvé, J. Glynn, M. Boily, R. White, J. Habbema, et R. Hayes. Understanding the differences between contrasting HIV epidemics in east and west Africa : results from a simulation model of the Four Cities Study. *Sex Transm Infect*, 83 Suppl 1 : 5–16, 2007. (Cité page 75.)

M. Over, R. Ellis, J. Huber, et O. Solon. The consequences of adult ill-health. Dans Feachem R., T. Kjellstrom, Ch. Murray, C. Over, et M. Phillips (eds), *The health of adults in the developing world*, pages 161–207. Oxford University Press of the World Bank, 1992. (Cité page 166.)

H. J. Page et G. Wunsch. Parental Survival Data : Some Results of the Application of Ledermann's Model Life Tables. *Population Studies*, 30 (1) :59–76, 1976. (Cité page 119.)

W. Paget et I. Timaeus. A Relational Gompertz Model of Male Fertility : Development and Assessment. *Population Studies*, Vol. 48, No. 2. :333–340, 1994. (Cité pages 113 et 132.)

A. Palloni et L. Heligman. Re-estimation of structural parameters to obtain estimates of mortality in developing countries. *Population Bulletin of the United Nations*, 18 :10–33, 1985. (Cité pages 48, 101, 125, 126 et 127.)

A. Palloni, M. Massagli, et J. Marcotte. Estimating Adult Mortality with Maternal Orphanhood Data : Analysis of Sensitivity of the Techniques. *Population Studies*, 38(2) :255–279, 1984. (Cité pages 100, 147, 148, 163, 167, 168, 169, 170, 174, 175, 184, 280 et 430.)

S. Pennec. La place des familles à quatre générations en France. *Population (French Edition)*, 51(1) :31–59, 1996. (Cité pages 47 et 53.)

S. Pennec et B. Bacon. Modelling fertility and mortality in the APPSIM dynamic microsimulation model. Online Working Paper - WP7, National Centre for Social and Economic Modelling, University of Canberra, 2007. (Cité pages 50 et 405.)

G. Pison. Why child mortality in sub-saharan africa has ceased declining since the early 1990s ? The example of Senegal, a country where the HIV epidemic has remained at a low level. Dans J. Caraël, M. et Glynn (ed), *HIV, Resurgent infections and population change in Africa*, Chapitre 3, pages 41–65. Springer, IUSSP, 2007. (Cité page 3.)

G. Pison, K. Hill, B. Cohen, et K. Foote. La mortalité. Dans *Les changements démographiques au Sénégal*, Chapitre 5. PUF, INED, National Academy of Sciences, 1997. (Cité pages 321 et 346.)

G. Pison et A. Langaney. Age Patterns of Mortality in Eastern Senegal : Comparison of Micro and Survey Approaches. Dans V. Caldwell J., Hill A. et Hull (ed), *Micro-approaches to demographic research*, pages 297–317. Kegan Paul International, London, 1988. (Cité page 281.)

E. Preble. Impact of HIV/AIDS on African children. *Social Science & Medicine*, 31(6) :671–680, 1990. (Cité page 195.)

S. Preston et N. Bennett. A census-based method for estimating adult mortality. *Population Studies*, 37(1) :91–104, 1983. (Cité pages 5, 18 et 346.)

S. Preston et A. Coale. Age Structure, Growth, Attrition, and Accession : A New Synthesis. *Population Index*, 48(2) :217–259, 1982. (Cité page 153.)

S. Preston, P. Heuveline, et M. Guillot. *Demography : Measuring and Modeling Population Processes*. Blackwell, 2001. (Cité pages 60, 106, 152 et 414.)

Th. Pullum. An Assessment of Age and Date Reporting in the DHS Surveys, 1985-2003. DHS Methodological Reports No. 5. Calverton, Maryland : Macro International Inc., 2006. (Cité pages 43 et 288.)

Th. Pullum. An Assessment of the Quality of Data on Health and Nutrition in the DHS Surveys, 1993-2003. DHS Methodological Reports No. 6., Calverton, Maryland : Macro International Inc., 2008. (Cité page 231.)

J. Rajaratnam, J. Marcus, A. Levin-Rector, A. Chalupka, H. Wang, L. Dwyer, M. Costa, A. Lopez, et Ch. Murray. Worldwide mortality in men and women aged 15-59 years from 1970 to 2010 : a systematic analysis. *The Lancet*, 375(9727) :1704 – 1720, 2010. (Cité pages 8, 261, 266, 343 et 352.)

S. Renata. Child fostering in Africa : When labor and schooling motives may coexist. *Journal of Development Economics*, 88(1) :157 – 170, 2009. (Cité page 279.)

G. Reniers. Divorce and Remarriage in Rural Malawi. *Demographic Research*, S1(6) :175–206, 2003. (Cité page 421.)

G. Reniers et J. Eaton. Refusal bias in HIV prevalence estimates from nationally representative seroprevalence surveys. *AIDS*, 23(5) :621–629, 2009. (Cité page 81.)

G. Reniers, B. Masquelier, et P. Gerland. Adult Mortality in Africa. Dans R. Rogers et E. Crimmins (eds), *International Handbook of Adult Mortality*. Springer, forthcoming. (Cité pages 3, 4, 8, 66, 183, 214, 226, 229, 260, 305, 322, 335, 343 et 353.)

R. Retherford. The Brass fertility polynomial. *Asian Pacific Census Forum*, 5(4) :15–19, 22, 1979. (Cité page 431.)

L. Robertson, S. Gregson, C. Madanhire, N. Walker, P. Mushati, G. Garnett, et C. Nyamukapa. Discrepancies between UN models and DHS survey estimates of maternal orphan prevalence : insights from analyses of survey data from Zimbabwe. *Sexually Transmitted Infections*, 84 S(1) :57–62, 2008. (Cité pages 281 et 282.)

G. Rodriguez et J. Trussell. Maximum likelihood estimation of the parameters of Coale's model nuptiality schedule from survey data. World Fertility Survey, Technical Bulletins, No 7, 1980. (Cité pages 417, 419 et 420.)

S Ruggles. *Prolonged connections : the rise of the extended family in nineteenth-century England and America*. University of Wisconsin Press, 1987. (Cité pages 44, 49, 50 et 53.)

S. Ruggles. Confessions of a microsimulator. *Historical Methods*, 26 :161–169, 1993. (Cité page 49.)

N. Rutenberg et J. Sullivan. Direct and indirect estimates of maternal mortality from the sisterhood method. Proceedings of the Demographic and Health Surveys World Conference, August 5-7 1991, Washington, D.C., Vol. III. IRD/MACRO International Inc., Columbia , Maryland., 1991. (Cité page 33.)

S. Rutstein. Finding the Missing Maternal and Paternal Orphans. Rapport technique, Report submitted to the Joint United Nations Programme on HIV/AIDS (UNAIDS), Macro International Inc. Calverton, Maryland, USA, 2008. (Cité pages 282, 286 et 444.)

S. Rutstein et G. Rojas. *Guide to DHS Statistics*. ORC Macro, Calverton, Maryland, USA, 2006. (Cité page 225.)

G. Salem, E. Cadot, et F. Fournet. Villes africaines et santé : de la nouvelle jeunesse des vieilles endémies à l'émergence de nouvelles épidémies. *Afrique contemporaine*, 195 :60–70, 2000. (Cité page 3.)

B. Schoumaker. Une approche personnes-périodes pour l'analyse des histoires génésiques. *Population (French Edition)*, 59(5) :783–796, 2004. (Cité pages 72 et 226.)

B. Schoumaker. Stalls and Reversals in Fertility Transition in Sub-Saharan Africa. Real or Spurious ? Paper presented at the IUSSP Seminar on Human Fertility in Africa. Trends in the Last Decade and Prospect for Change, Cape Coast, Ghana, 2008. (Cité page 102.)

B. Schwartländer, K. Stanecki, T. Brown, P. Way, R. Monasch, J. Chin, D. Tarantola, et N. Walkera. Country-specific estimates and models of HIV and AIDS : methods and limitations. *AIDS*, 13 (17) :2445–2458, 1999. (Cité page 80.)

J. Siegel, D. Swanson, et H. Shryock. *The Methods and Materials of Demography (Second Edition)*. Emerald Group Publishing, 2004. (Cité pages 68 et 408.)

J. Smith. The Computer Simulation of Kin Sets and Kin Counts. Dans J. Bongaarts, T. Burch, et K. Wachter (eds), *Family Demography : Methods and Their Applications*. Oxford : Clarendon Press, 1987. (Cité page 50.)

L. Smith, R. Hyndman, et S. Wood. Spline interpolation for demographic variables : the monotonicity problem. *Journal of Population Research*, 21(1) :95–98, 2004. (Cité pages 68 et 70.)

C. Stanton, J. Hobcraft, K. Hill, N. Kodjogbe, W. Mapeta, F. Munene, M. Naghavi, V. Rabeza, B. Sisouphantong, et O. Campbell. Every Death Counts : Measurement of Maternal Mortality via a Census. *Bulletin of the World Health Organization*, 79 (7) :657–664, 2001. (Cité pages 19 et 31.)

C. Stanton, A. Noureddine, et K. Hill. DHS Maternal Mortality Indicators : an assessment of data quality and implications for Data Use. DHS Analytical Reports No.4, Calverton, Maryland : Macro International Inc., 1997. (Cité pages 33, 294, 295 et 298.)

C. Stanton, A. Noureddine, et K. Hill. DHS Maternal Mortality Indicators : an assessment of data quality and implications for Data Use. *Studies in Family planning*, 31,2 :111–123, 2000. (Cité pages 9, 290, 291, 297 et 298.)

J. Stover. Projecting the demographic consequences of adult HIV prevalence trends : the Spectrum Projection Package. *Sexually Transmitted Infections*, 80(suppl 1) :i14–i18, 2004. (Cité page 89.)

J. Stover. AIM : A Computer Program for Making HIV/AIDS Projections and Examining the Demographic and Social Impacts of AIDS. Futures Group International, Health Policy Initiative, Task Order 1., 2009. (Cité pages 82, 83, 86, 88, 89 et 94.)

J. Stover, P. Johnson, B. Zaba, M. Zwahlen, F. Dabis, et R. Ekpini. The Spectrum projection package : improvements in estimating mortality, ART needs, PMTCT impact and uncertainty bounds. *Sexually Transmitted Infections*, 84(Suppl 1) :i24–i30, Août 2008. (Cité pages 67, 75 et 92.)

J. Stover, N. Walker, N. Grassly, et M. Marston. Projecting the demographic impact of AIDS and the number of people in need of treatment : updates to the Spectrum projection package. *Sexually Transmitted Infections*, 82 (suppl 3) :iii45–iii50, 2006. (Cité page 94.)

D. Tabutin. La fécondité et la mortalité dans les recensements africains des 25 dernières années. *Population*, 39(2) :295–312, 1984. (Cité page 14.)

D. Tabutin. Les relations entre pauvreté et fécondité dans les pays du Sud et en Afrique sub-saharienne. Dans B Ferry (ed), *L'Afrique face à ses défis démographiques*, pages 253–288. AFD-CEPED-Karthala, 2007. (Cité pages 175 et 176.)

D. Tabutin, C. Gourbin, et G. Beninguisse. Surmortalité et santé des petites filles en Afrique. Tendances des années 1970 aux années 1990. Dans T. Locoh (ed), *Genre et sociétés en Afrique : implications pour le développement*, Chapitre 5, pages 137–369. INED, 2007. (Cité page 296.)

D. Tabutin et B. Schoumaker. La démographie de l'Afrique au sud du Sahara des années 1950 aux années 2000. *Population*, 3-4 :521–622, 2004. (Cité page 412.)

C. Tichit et N. Robette. *Appréhender la composition familiale des ménages dans les Enquêtes démographiques et de santé*. Collections du CEPED. 2006. (Cité page 39.)

I. Timaeus. An Assessment of Methods for Estimating Adult Mortality from Two Sets of Data on Maternal Orphanhood. *Demography*, 23, 3 : 435–450, 1986. (Cité pages 11, 111, 145, 150, 152, 154, 280 et 282.)

I. Timaeus. Estimation of Fertility and Mortality from WFS Household Surveys. Dans J. Cleland et C. Scott (eds), *The World Fertility Survey : An assessment*, Chapitre 4, pages 93–129. Oxford University Press, 1987. (Cité page 20.)

I. Timaeus. Estimation of Adult Mortality from Orphanhood Before and Since Marriage. *Population Studies*, 45(3) :455–472, 1991a. (Cité pages 11, 23, 159, 282 et 348.)

I. Timaeus. Estimation of Mortality from Orphanhood in Adulthood. *Demography*, 28 (2) :213–227, 1991b. (Cité pages 11, 111, 152, 156, 157, 282, 348 et 430.)

I. Timaeus. New estimates of adult mortality from DHS data on the timing of orphanhood relative to marriage. Proceedings of the Demographic and Health Surveys World Conference, Washington D.C., 1991c. Vol. II. Columbia, Md. : Institute for Resource Development/Macro International, Inc. (Cité page 321.)

I. Timaeus. Estimation of adult mortality from paternal orphanhood : a reassessment and a new approach. *Population Bulletin of the United Nations*, 33 :47–63, 1992. (Cité pages 11, 101, 111, 113, 114, 115, 116, 117, 122, 123, 125, 126, 127, 129, 130, 132, 133, 136, 144, 145, 183, 316, 318 et 348.)

I. Timaeus. Adult mortality. Dans K. Foote, K. Hill, et L. Martin (eds), *Demographic change in Sub-Saharan Africa*, Chapitre 6. Panel on the Population Dynamics of Sub-Saharan Africa, Committee on Population, Commission on Behavioral and Social Sciences and Education, National Research Council, 1993. (Cité pages 17, 18, 318, 320, 321, 329 et 330.)

I. Timaeus. Mortality in Sub-Saharan Africa. Dans J. Chamie et R. Cliquet (eds), *Health and mortality, issues of global concern, Proceedings of the symposium on health and mortality*, pages 110–131. New York : United Nations and Population and Family Study Centre, 1999. (Cité pages 11 et 14.)

I. Timaeus et M. Jasseh. Adult mortality in sub-saharan Africa : evidence from demographic and health survey. *Demography*, 41, 4 :757–772, 2004. (Cité pages 11, 226, 229, 230, 232, 233, 234, 235, 258, 260, 264, 275, 304, 305, 307, 315, 316, 317, 320, 322, 326, 329, 334, 343, 352 et 353.)

I. Timaeus et A. Nunn. Measurement of adult mortality in populations affected by AIDS : an assessment of the orphanhood method. *Health transition Review*, 7(2) :23–43, 1997. (Cité pages 11, 18, 21, 100, 111, 197, 200, 201, 202, 203, 204, 205, 206, 217, 219, 220, 221, 318, 334, 349, 352, 435 et 437.)

I. Timaeus, B. Zaba, et M. Ali. Estimation of adult mortality from data on adult siblings. Dans B. Zaba et J. Blacker (eds), *Brass tacks : essays in medical demography*, pages 43–66. Athlone Press, Londres, 2001. (Cité pages 11, 101, 235, 267, 268, 271, 272, 273, 275, 316, 350 et 441.)

J. Todd, J. Glynn, M. Marston, T. Lutalo, S. Biraro, W. Mwita, V. Suriyanon, R. Rangsin, K. Nelson, P. Sonnenberg, D. Fitzgerald, E. Karita, et B. Zaba. 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 :S55–63, 2007. (Cité pages 85, 87 et 88.)

J. Trussell et G. Rodriguez. A note on the sisterhood estimator of maternal mortality. *Studies in Family Planning*, 21 (6) :344–346, 1990. (Cité pages 39, 234, 235, 237, 238, 239, 241, 248 et 352.)

UN. *Demographic Yearbook 1965*. Dept. of International Economic and Social Affairs, United Nations, 1966. (Cité pages 130 et 301.)

UN. *Demographic Yearbook 1969*. Dept. of International Economic and Social Affairs, United Nations, 1970. (Cité page 130.)

UN. *Demographic Yearbook 1975*. Dept. of International Economic and Social Affairs, United Nations, 1976. (Cité page 130.)

UN. *Demographic Yearbook 1979 - Historical supplement*. Dept. of International Economic and Social Affairs, United Nations, 1979. (Cité page 414.)

UN. *Model Life Tables for Developing Countries*. New York : United Nations, 1982. (Cité pages 147 et 230.)

UN. *Demographic Yearbook 1981*. Dept. of International Economic and Social Affairs, United Nations, 1983a. (Cité page 130.)

UN. *Manual X : Indirect Techniques for Demographic Estimation*. Population Division, Dept. of International Economic and Social Affairs, United Nations, New York, 1983b. (Cité pages 21, 101 et 428.)

UN. *First marriage : patterns and determinants*. Population Division, Dept. of International Economic and Social Affairs, United Nations, New York, ST/ESA/SER.R/76, 1988. (Cité pages 412 et 414.)

UN. *Patterns of First Marriage : Timing and Prevalence*. Numéro ST/ESA/SER.R/111. Population Division, Dept. of International Economic and Social Affairs, United Nations, New York, 1990. (Cité pages 412 et 414.)

UN. *Demographic Yearbook 1997 - Historical supplement*. Dept. of International Economic and Social Affairs, United Nations, 1997. (Cité pages 14, 137, 144, 301 et 414.)

UN. World Marriage Patterns 2000. Population Division, Dept. of International Economic and Social Affairs, United Nations, New York, 2000. (Cité pages 412 et 414.)

UN. *Demographic Yearbook 1999 - Natality Statistics (CD)*. Dept. of International Economic and Social Affairs, United Nations, 2002a. (Cité page 71.)

UN. *Methods for Estimating Adult Mortality*. ESA/P/WP.175, Population Division, Dept. of International Economic and Social Affairs, United Nations, New York, 2002b. (Cité page 143.)

UN. *World Fertility Report 2003*. Population Division, Dept. of International Economic and Social Affairs, United Nations, ESA/P/WP.189, New York, 2004. (Cité pages 412 et 414.)

UN. *World Population Prospects : The 2004 Revision*, volume III : Analytical Report. Population Division, Dept. of International Economic and Social Affairs, United Nations, New York, 2005. (Cité pages 7, 66, 67 et 95.)

UN. *World Mortality Report 2005*. Population Division, Dept. of International Economic and Social Affairs, United Nations, New York, 2006. (Cité page 14.)

UN. *Demographic Yearbook 2004*. Dept. of International Economic and Social Affairs, United Nations, 2007. (Cité page 14.)

UN. *Demographic Yearbook 2006*. Dept. of International Economic and Social Affairs, United Nations, 2008a. (Cité pages 14, 15 et 16.)

UN. *Principles and Recommendations for Population and Housing Censuses*. Dept. of International Economic and Social Affairs, United Nations, New York, 2008b. (Cité page 31.)

UN. Report on the United Nations Workshop on the Improvement of Civil Registration and Vital Statistics in SADC Region, 1-5 December 2008, Blantyre, Malawi, 2008c. (Cité page 15.)

UN. *Demographic Yearbook 2007*. Dept. of International Economic and Social Affairs, United Nations, 2009a. (Cité pages 14, 15, 16 et 137.)

UN. *World Marriage Data 2008*. Population Division, Dept. of International Economic and Social Affairs, United Nations, New York, 2009b. (Cité pages 412 et 414.)

UN. *World Population Prospects : The 2008 Revision*. Population Division, Dept. of International Economic and Social Affairs, United Nations, New York, 2009c. (Cité pages 3, 5, 8, 64, 86, 93, 94, 139, 211, 217 et 296.)

UN. World Population Prospects : The 2008 Revision. CD-ROM Edition - Extended Dataset in Excel and ASCII formats. (United Nations publication, ST/ESA/SER.A/283), 2009d. (Cité pages 64, 69, 70, 211, 225 et 312.)

UNAIDS. Improved methods and assumptions for estimation of the HIV/AIDS epidemic and its impact : Recommendations of the UNAIDS Reference Group on Estimates, Modelling and Projections : The UNAIDS Reference Group on Estimates, Modelling and Projections. *AIDS*, 16(9) :1–14, 2002a. (Cité pages 94 et 195.)

UNAIDS. *2008 Report on the global AIDS epidemic*. UNAIDS, 2008. (Cité pages 1 et 233.)

UNAIDS. *AIDS epidemic update : 2009*. UNAIDS, 2009. (Cité pages 78, 79, 320 et 329.)

USAID UNAIDS, UNICEF. *Children on the Brink 2002 : A Joint Report on Orphan Estimates and Program Strategies*. UNAIDS, UNICEF, USAID, Washington DC, 2002b. (Cité page 183.)

UNICEF. *Africa's Orphaned and Vulnerable Generations : Children Affected by AIDS*. UNICEF, UNAIDS and PEPFAR, 2006. (Cité pages 87, 195, 213, 214 et 282.)

M. Urassa, J. Boerma, R.l Isingo, J. Ngalula, J. Ng'weshemi, G. Mwaluko, et B. Zaba. The impact of HIV/AIDS on mortality and household mobility in rural Tanzania. *AIDS*, 15(15) :2017–23, 2001. (Cité pages 17 et 163.)

E Van de Walle. L'âge au mariage : tendances récentes. Dans K. Foote, K. Hill, et L. Martin (eds), *Changements démographiques en Afrique subsaharienne*, Chapitre 4, pages 119–154. Ined, coll. "Travaux et documents", 1996. (Cité page 414.)

C. Van der Ploeg, C. Van Vliet, S. De Vlas, J. Ndinya-Achola, L. Fransen, G. Van Oortmarssen, et J. Habbema. STDSIM : a Microsimulation Model for Decision Support in STD Control. *Interfaces*, 28(3) :84–100, 1998. (Cité page 75.)

E. Van Imhoff et W. Post. Méthodes de micro-simulation pour des projections de population. *Population*, 4 :889–932, 1997. (Cité pages 45, 46 et 48.)

C. Vandermeersch. Child Fostering under Six in Senegal in 1992-1993. *Population (English Edition)*, 57(4/5) :pp. 659–685, 2002. (Cité page 279.)

K. Wachter. The Sisters' Riddle and the Importance of Variance when Guessing Demographic Rates from Kin Counts. *Demography*, 17(1) :103–114, 1980. (Cité pages 50 et 272.)

K. Wachter. Kinship Resources for the Elderly. *Philosophical Transactions : Biological Sciences*, 352(1363) :1811–1817, 1997. (Cité page 50.)

K. Wachter, D. Blackwell, et E. Hammel. Testing the validity of kinship microsimulation. *Journal of Mathematical and Computer Modeling*, 26 :89–104, 1997. (Cité page 50.)

K. Wachter, J. Knodel, et M. Vanlandingham. AIDS and the elderly of Tahiland : projecting familial impacts. *Demography*, 39(1) :25–41, 2002. (Cité pages 50, 56 et 76.)

K. Wachter, J. Knodel, et M. VanLandingham. Parental bereavement : heterogeneous impacts of AIDS in Thailand. *Journal of Econometrics*, 112 (1) :193 – 206, 2003. (Cité pages 56 et 76.)

N. Walker, K. Stanecki, T. Brown, J. Stover, S. Lazzari, J. Garcia-Calleja, B. Schwartlander, et P. Ghys. Methods and procedures for estimating HIV/AIDS and its impact : the UNAIDS/WHO estimates for the end of 2001. *AIDS*, 17(15) :2215–25, 2003. (Cité page 80.)

G. Walraven, R. Mkanje, J. van Roosmalen, P. van Dongen, et W. Dolman. Assessment of maternal mortality in Tanzania. *British Journal of Obstetrics and Gynaecology*, 101 (5) :414–417, 1994. (Cité pages 32 et 271.)

D. Waltisperger et F. Meslé. Crise économique et mortalité : Le cas d'Antananarivo 1976-2000. *Population (French Edition)*, 60(3) :pp. 243–275, 2005. ISSN 00324663. (Cité page 329.)

D. Waltisperger et T. Rabetsitonta. Un bilan de trente ans de mesures directes de la mortalite adulte en Afrique. African Population Conference, Dakar, 1988, Tome 2. Liège : International Union for the Scientific Study of Population. Annexes. Unpublished tables, Centre Français sur la Population et le Developpment, Paris, 1988. (Cité pages 329 et 330.)

C. Weinberg et B. Gladen. The Beta-Geometric Distribution Applied to Comparative Fecundability Studies. *Biometrics*, 42(3) :547–560, 1986. (Cité page 401.)

WHO. *HIV Transmission Through Breastfeeding : a review of available evidence (2007 Update)*. WHO Library Cataloguing-in-Publication Data, 2007a. (Cité page 94.)

WHO. *Global Tuberculosis Control : Surveillance, Planning and Financing*. WHO, 2008a. (Cité page 2.)

WHO. *The global burden of disease : 2004 update*. World Health Organization, 2008b. (Cité page 2.)

UNICEF WHO, UNAIDS. *Towards universal access : scaling up priority HIV/AIDS interventions in the health sector. Progress report, April 2007*. WHO, UNAIDS, UNICEF, 2007b. (Cité pages 90, 92 et 93.)

WHO, UNICEF, UNFPA and The World Bank. *Maternal Mortality in 2005. Estimates developed by WHO, UNICEF, UNFPA and the World Bank*. WHO Library Cataloguing-in-Publication Data, 2008. (Cité page 30.)

WHO, UNICEF, UNFPA and The World Bank. *Trends in Maternal Mortality : 1990 to 2008. Estimates developed by WHO, UNICEF, UNFPA and the World Bank*. WHO Library Cataloguing-in-Publication Data, 2010. (Cité pages 29, 30 et 354.)

C. Wilson. On the scale of global demographic convergence, 1950-2000. *Population and Development Review*, 27(1) :155–171, 2001. (Cité page 3.)

D. Wolf. The Role of Microsimulation in Longitudinal Data Analysis. *Canadian Studies in Population*, 28(2) :313–339, 2001. (Cité page 44.)

B. Zaba et S. Gregson. Measuring the impact of HIV on fertility in Africa. *AIDS*, 12 (S1) :41–50, 1998. (Cité page 94.)

B. Zaba, M. Marston, A. Crampin, R. Isingo, S. Biraro, T. Bernighausen, B. Lopman, T. Lutalo, J. Glynn, et J. Todd. Age-specific mortality patterns in HIV-infected individuals : a comparative analysis of African community study data. *AIDS*, 21 :S87–S96, 2007. (Cité pages 85 et 87.)

B. Zaba, J. Whitworth, M. Marston, J. Nakiyingi, A. Ruberantwari, M. Urassa, R.i Issingo, G. Mwaluko, S. Floyd, A. Nyondo, et A. Crampin. HIV and Mortality of Mothers and Children : Evidence From Cohort Studies in Uganda, Tanzania, and Malawi. *Epidemiology*, 16(3) :–, 2005. (Cité pages 2, 98, 165, 195 et 208.)

Basia Zaba. The Indirect Estimation of Migration : A Critical Review. *International Migration Review*, 21(4) :pp. 1395–1445, 1987. (Cité page 268.)

E. Zagheni. The Effect of the HIV-AIDS Epidemic on Orphanhood Probabilities and Kinship Structure in Zimbabwe. Results Based on Formal Demography and Microsimulation. Paper Presented at the Population Association of America, 2010 Annual Meeting, Dallas, Texas, April, 2010. (Cité pages 57, 76 et 209.)

Z. Zhao. Computer microsimulation and historical study of social structure : A comparative review of SOCSIM and CAMSIM. *Revista de la Asociacion de Demografia Historica*, XXIV (2) :59–88, 2006. (Cité page 50.)

F. Zimmerman. Cinderella Goes to School : The Effects of Child Fostering on School Enrollment in South Africa. *J. Human Resources*, 38(3) :557–590, 2003. (Cité page 279.)

H. Zlotnik et K. Hill. The use of hypothetical cohorts in estimating demographic parameters under conditions of changing fertility and mortality. *Demography*, 18(1) :103–122, 1981. (Cité page 150.)

A Annexes

Sommaire

A.1 Annexes au chapitre 1 380
A.1.1 Classification UNECA des pays africains 380
A.1.2 Enquêtes nationales disposant d'informations sur la survie des parents 381
A.1.3 Enquêtes nationales disposant d'informations sur la survie des frères et sœurs 385
A.1.4 Notes sur l'appariement des frères et sœurs adultes dans les fichiers EDS 388
A.2 Annexes au chapitre 2 394
A.2.1 Comparaison des différentes options disponibles pour modéliser la fécondité en micro-simulation 394
A.2.2 Notes sur le modèle de Lee et Carter (1992) 406
A.2.3 Comparaison des taux de croissance équivalents aux populations stables (paramètres de 1950-1955) avec les taux de croissance estimés en 1950-1955 par les Nations Unies 407
A.2.4 Notes sur les fonctions splines 408
A.2.5 Distribution des décès infantiles utilisée dans les microsimulations 409
A.2.6 Notes sur le calcul des taux de fécondité par rang à partir des rapports de taux 410
A.2.7 Notes sur l'intégration de la nuptialité dans les microsimulations africaines 412
A.2.8 Notes sur les distributions "phase-type" 424
A.3 Annexes au chapitre 3 426
A.3.1 Calcul des probabilités ${}_np_{25}$ à partir de la survie des mères : 6 recensements est-africains 426
A.3.2 Coefficients utilisés pour les méthodes indirectes 428
A.3.3 Modèle polynomial de Brass 431
A.3.4 Notes sur l'estimation indirecte de l'âge moyen à la naissance des enfants des pères 432
A.4 Annexes au chapitre 4 435
A.4.1 Formules de Timaeus et Nunn (1997) 435
A.5 Annexes au chapitre 5 438
A.5.1 Démonstration de Trussell et Rodriguez (1990) 438
A.5.2 Méthode des sœurs développée par Graham et al. (1989) . 439

A.5.3 Coefficients utilisés pour dater les estimations indirectes basées sur la survie des frères et sœurs 441
A.6 Annexes au chapitre 6 442
A.6.1 Données manquantes sur la survie des parents 442
A.6.2 Notes sur le traitement des réponses manquantes relatives à la survie des parents 443
A.6.3 Données manquantes sur les frères et sœurs 447
A.7 Annexes au chapitre 7 457
A.7.1 Bénin 457
A.7.2 Burkina Faso 458
A.7.3 Centrafique 459
A.7.4 Comores 460
A.7.5 Congo 461
A.7.6 Congo RDC 462
A.7.7 Côte d'Ivoire 463
A.7.8 Ethiopie 464
A.7.9 Gabon 465
A.7.10 Ghana 466
A.7.11 Guinée 467
A.7.12 Kenya 468
A.7.13 Lesotho 469
A.7.14 Liberia 470
A.7.15 Mozambique 471
A.7.16 Namibie 472
A.7.17 Nigeria 473
A.7.18 Sierra Leone 474
A.7.19 Swaziland 475
A.7.20 Tanzanie 476
A.7.21 Togo 477

A.1 Annexes au chapitre 1

A.1.1 Classification UNECA des pays africains

Source : Commission économique pour l'Afrique (UNECA).

Les codes entre parenthèses renvoient aux codes ISO utilisés dans les graphes pour repérer les pays.

Les pays notés par une astérisque sont non autonomes (et non membres de l'UNECA). Ils sont repris ici car ils sont interviennent dans les moyennes régionales présentées dans ce document.

L'**Afrique du Nord** (n=8) n'est pas étudiée ici. Selon la classification UNECA, elle reprend l'Algérie, l'Égypte, la Libye, la Mauritanie, le Maroc, le Soudan, la Tunisie et le Sahara Occidental*.

Afrique de l'Ouest (n=16) :
- Bénin (BEN)
- Burkina-Faso (BFA)
- Cap Vert (CPV)
- Côte d'Ivoire (CIV)
- Gambie (GMB)
- Ghana (GHA)
- Guinée (GIN)
- Guinée Bissau (GNB)
- Liberia (LBR)
- Mali (MLI)
- Niger (NER)
- Nigeria (NGA)
- Saint Hélène* (SHN)
- Sénégal (SEN)
- Sierra Leone (SLE)
- Togo (TGO)

Afrique centrale (n=7) :
- Cameroun (CMR)
- Congo (COG)
- Guinée équatoriale (GNQ)
- Gabon (GAB)
- République Centrafricaine (CAF)
- Sao Tome et Principe (STP)
- Tchad (TCD)

Afrique de l'Est (n=15) :
- Burundi (BDI)
- Comores (COM)
- Congo RDC (COD)
- Djibouti (DJI)
- Érythrée (ERI)
- Éthiopie (ETH)
- Kenya (KEN)
- Madagascar (MDG)
- Mayotte* (MYT)
- Réunion* (REU)
- Rwanda (RWA)
- Seychelles (SYC)
- Somalie (SOM)
- Tanzanie (TZA)
- Ouganda (UGA)

Afrique australe (n=11) :
- Angola (AGO)
- Botswana (BWA)
- Lesotho (LSO)
- Malawi (MWI)
- Maurice (MUS)
- Mozambique (MOZ)
- Namibie (NAM)
- Afrique du Sud (ZAF)
- Swaziland (SWZ)
- Zambie (ZMB)
- Zimbabwe (ZWE)

A.1.2 Enquêtes nationales disposant d'informations sur la survie des parents

Enquête		Type	Année
Afrique du Sud	LSMS		1993
	DHS	STAND	1998
Angola	MICS	MICS1	1996
	MICS	MICS2	2001
Bénin	DHS	STAND	1996
	DHS	STAND	2001
	DHS	STAND	2006
Botswana	MICS	MICS2	2000
Burkina Faso	DHS	STAND	1993
	DHS	STAND	2003
	MICS	MICS3	2006
Burundi	DHS	STAND	1987
	MICS	MICS2	2000
	MICS	MICS3	2005
Cameroun	DHS	STAND	1991
	DHS	STAND	1998
	MICS	MICS2	2000
	DHS	STAND	2004
	MICS	MICS3	2006
Centrafique	DHS	STAND	1995
	MICS	MICS2	2000
	MICS	MICS3	2006
Comores	DHS	STAND	1996
	MICS	MICS2	2000
Congo	DHS	STAND	2005

Enquête		Type	Année
Congo RDC	MICS	MICS2	2001
	DHS	STAND	2007
Côte d'Ivoire	LSMS		1985
	LSMS		1986
	LSMS		1987
	LSMS		1988
	DHS	STAND	1994
	MICS	MICS2	2000
	DHS	AIS	2005
	MICS	MICS3	2006
Ethiopie	DHS	STAND	2000
	DHS	STAND	2005
Gabon	DHS	STAND	2000
Gambie	MICS	MICS2	2001
	MICS	MICS3	2006
Ghana	LSMS		1988
	DHS	STAND	1988
	LSMS		1989
	LSMS		1992
	DHS	STAND	1993
	MICS	MICS1	1995
	LSMS		1998
	DHS	STAND	1998
	DHS	STAND	2003
	MICS	MICS3	2006
	DHS	STAND	2008
Guinée	DHS	STAND	1999
	DHS	STAND	2005
Guinée-Bissau	MICS	MICS2	2000
	MICS	MICS3	2006
Guinée Equ.	MICS	MICS2	2000
Kenya	DHS	STAND	1989
	DHS	STAND	1993
	DHS	STAND	1998
	MICS	MICS2	2000
	DHS	STAND	2003

Enquête		Type	Année
Lesotho	MICS	MICS2	2000
	DHS	STAND	2004
Liberia	DHS	STAND	1986
	MICS	MICS1	1995
	DHS	STAND	2007
Madagascar	DHS	STAND	1992
	DHS	STAND	1997
	MICS	MICS2	2000
	DHS	STAND	2004
	DHS	STAND	2009
Malawi	DHS	STAND	1992
	MICS	MICS1	1995
	DHS	STAND	2000
	DHS	STAND	2004
	LSMS		2004
	MICS	MICS3	2006
Mali	DHS	STAND	1987
	DHS	STAND	1996
	DHS	STAND	2001
	DHS	STAND	2006
Mozambique	MICS	MICS1	1995
	DHS	STAND	1997
	DHS	STAND	2003
	MICS	MICS3	2008
Namibie	DHS	STAND	1992
	DHS	STAND	2000
	DHS	STAND	2007
Niger	DHS	STAND	1992
	MICS	MICS1	1995
	DHS	STAND	1998
	MICS	MICS2	2000
	DHS	STAND	2006
Nigeria	DHS	STAND	1999
	DHS	STAND	2003
	MICS	MICS3	2007
	DHS	STAND	2008
Ouganda	DHS	STAND	1988
	DHS	STAND	1995
	DHS	STAND	2001
	DHS	STAND	2006

	Enquête	Type	Année
Rwanda	DHS	STAND	1992
	DHS	STAND	2000
	MICS	MICS2	2000
	DHS	STAND	2005
Sao Tome	MICS	MICS2	2000
	MICS	MICS3	2006
Sénégal	DHS	STAND	1986
	DHS	STAND	1993
	MICS	MICS1	1995
	DHS	STAND	1999
	MICS	MICS2	2000
	DHS	STAND	2005
Sierra Leone	MICS	MICS2	2000
	MICS	MICS3	2005
	DHS	STAND	2008
Somalie	MICS	MICS2	1999
	MICS	MICS3	2006
Swaziland	MICS	MICS2	2000
	DHS	STAND	2006
Tanzanie	DHS	STAND	1992
	DHS	STAND	1996
	DHS	STAND	1999
	DHS	AIS	2003
	DHS	STAND	2004
	DHS	AIS	2007
Tchad	DHS	STAND	1997
	MICS	MICS2	2000
	DHS	STAND	2004
Togo	DHS	STAND	1988
	DHS	STAND	1998
	MICS	MICS2	2000
	MICS	MICS3	2006
Zambie	DHS	STAND	1992
	MICS	MICS1	1995
	DHS	STAND	1996
	MICS	MICS2	1999
	DHS	STAND	2002
	DHS	STAND	2007

Enquête		Type	Année
Zimbabwe	DHS	STAND	1988
	DHS	STAND	1994
	DHS	STAND	1999
	DHS	STAND	2006

A.1.3 Enquêtes nationales disposant d'informations sur la survie des frères et sœurs

Les enquêtes dont les fichiers de données ne sont pas distribués dans le domaine public (via internet sans autorisation des offices nationaux) ne sont pas repris dans ce tableau.

Toutes ces enquêtes ne permettent pas un calcul direct des taux de mortalité (*cf.* p.36)

Enquête		Type	Année	Enquêtés
Afrique du Sud	DHS	STAND	1998	F
Angola	MICS	MICS1	1996	F
Bénin	DHS	STAND	1996	F
	DHS	STAND	2006	F
Burkina Faso	DHS	STAND	1999	F
	DHS	STAND	2003	F
Burundi	MICS	MICS2	2000	F
	MICS	MICS3	2005	F
Cameroun	DHS	STAND	1998	F
	DHS	STAND	2004	F+M
	MICS	MICS3	2006	F
Centrafique	DHS	STAND	1995	F
	MICS	MICS3	2006	F
Comores	MICS	MICS2	2000	F
Congo	DHS	STAND	2005	F+M
Congo RDC	DHS	STAND	2007	F
Côte d'Ivoire	DHS	STAND	1994	F
	DHS	AIS	2005	F
	MICS	MICS3	2006	F
Ethiopie	DHS	STAND	2000	F
	DHS	STAND	2005	F

Enquête		Type	Année	Enquêtés
Gabon	DHS	STAND	2000	F
Gambie	MICS	MICS2	2001	F
Ghana	MICS	MICS3	2006	F
Guinée	DHS	STAND	1999	F
	DHS	STAND	2005	F
Guinée-Bissau	MICS	MICS2	2000	F
	MICS	MICS3	2006	F
Kenya	DHS	STAND	1998	F
	DHS	STAND	2003	F
	DHS	STAND	2009	F
Lesotho	DHS	STAND	2004	F
Liberia	DHS	STAND	2007	F+M
Madagascar	DHS	STAND	1992	F
	DHS	STAND	1997	F
	MICS	MICS2	2000	F
	DHS	STAND	2004	F
	DHS	STAND	2009	F
Malawi	DHS	STAND	1992	F+M
	DHS	STAND	2000	F
	DHS	STAND	2004	F
	MICS	MICS3	2006	F
Mali	DHS	STAND	1996	F
	DHS	STAND	2001	F
	DHS	STAND	2006	F
Mozambique	DHS	STAND	1997	F
	DHS	STAND	2003	F
	MICS	MICS3	2008	F
Namibie	DHS	STAND	1992	F
	DHS	STAND	2000	F
	DHS	STAND	2007	F+M
Niger	DHS	STAND	1992	F
	DHS	STAND	2006	F

Enquête		Type	Année	Enquêtés
Nigeria	DHS	STAND	1999	F+M
	MICS	MICS3	2007	F
	DHS	STAND	2008	F+M
Ouganda	DHS	STAND	1995	F+M
	DHS	STAND	2001	F
	DHS	STAND	2006	F
Rwanda	DHS	STAND	2000	F
	DHS	STAND	2005	F
Sénégal	DHS	STAND	1993	F
	DHS	STAND	2005	F
Sierra Leone	MICS	MICS2	2000	F
	MICS	MICS3	2005	F
	DHS	STAND	2008	F
Somalie	MICS	MICS3	2006	F
Swaziland	MICS	MICS2	2000	F
	DHS	STAND	2006	F
Tanzanie	DHS	STAND	1996	F+M
	DHS	STAND	2004	F
Tchad	DHS	STAND	1997	F
	DHS	STAND	2004	F
Togo	DHS	STAND	1998	F
Zambie	DHS	STAND	1996	F
	DHS	STAND	2002	F
	DHS	STAND	2007	F+M
Zimbabwe	DHS	STAND	1994	F+M
	DHS	STAND	1999	F
	DHS	STAND	2006	F+M

A.1.4 Notes sur l'appariement des frères et sœurs adultes dans les fichiers EDS

Nous présentons ici la technique mise au point pour identifier les fratries dupliquées au sein des fichiers EDS.

Chaque individu éligible au questionnaire individuel est mis en vis-à-vis de tous les autres individus éligibles du même ménage pour former des paires. Ensuite, l'appariement repose sur trois étapes : (1) un appariement partiel déterministe, (2) le calcul d'une métrique de distance entre les déclarations, (3) le classement des paires de déclarations comme se référant ou non à une même fratrie.

Appariement partiel déterministe

1. Toutes les paires d'observations constituées sont d'abord examinées sous l'angle du *lien de parenté qui associe chacun au chef de ménage*. En effet, dans le questionnaire ménage, chaque individu est associé au chef de ménage comme étant son fils, sa femme, un parent, etc. C'est la principale (et quasiment l'unique) information disponible sur les liens de parenté dans les enquêtes EDS.
   On considère ici toutes les combinaisons possibles. Celles-ci peuvent être soit :

   - Des combinaisons qui renvoient à une relation de fratrie *probable* : deux sœurs du chef de ménage sont nécessairement sœurs l'une de l'autre, tout comme deux enfants du chef de ménage…
   - Des combinaisons qui renvoient à une relation *indéterminée* : deux petites-filles du chef de ménage peuvent être sœurs l'une de l'autre, mais également cousines…
   - Des combinaisons dont il est *improbable* qu'elles renvoient à une relation de fratrie : la femme du chef de ménage ne peut pas être la sœur des enfants du chef de ménage, ni la sœur du chef de ménage…

   Prenons deux individus enquêtés. Mis ensemble, ils forment une paire, et l'on connaît pour chacun d'eux la relation qu'ils entretiennent avec le chef de ménage. A partir de là, un score de 2 est attribué à cette paire en cas de relation de fratrie *probable*, de 1 en cas de relation de fratrie *indéterminée*, et de 0 en cas de relation de fratrie *improbable* (Table A.1).[1] A ce stade, aucune distinction n'est faite entre les enfants nés de même *mère* et les enfants nés de même *père*.

2. Dans le questionnaire ménage, les individus de moins de 15 ans identifient leurs mères en procurant le numéro de la ligne à laquelle

1. Ces "distances" ont été retenues en autorisant une certaine marge d'erreur dans les déclarations ; en considérant par exemple que les individus déclarés comme non-apparentés puissent potentiellement être des frères ou des sœurs d'un des membres du ménage…

TABLE A.1 – *Matrice de distance entre les individus du ménage (relation de fratrie)*

	Head	Wife	Son/D.	Son/D. in law
Head	0	0	0	1
Wife	0	0	0	1
Son-Daughter	0	0	1	1
Son-d. in law	1	1	1	1
Grand-Child	0	0	0	0
Parent	0	1	0	0
Parent-in-law	1	0	0	0
Brother Sister	**2**	1	0	0
Co-Spouse	0	1	0	1
Other relative	1	1	1	1
Fost./Ad. child	0	1	1	1
Not related/DK	1	1	1	1
Other	1	1	1	1

	Grd- Child	Parent	P. in-law	Brtr/ Sistr.
Head	0	0	1	**2**
Wife	0	1	0	1
Son-Daughter	0	0	0	0
Son-d. in law	0	0	0	0
Grand-Child	1	0	0	0
Parent	0	0	1	0
Parent-in-law	0	1	0	0
Brother Sister	0	0	0	**2**
Co-Spouse	0	1	0	0
Other relative	1	1	1	1
Fost./Ad. child	0	0	0	0
Not related/DK	1	1	1	1
Other	1	1	1	1

	Co-spouse	Adpt fstr child	Not rltd/DK	Other
Head	0	0	1	1
Wife	1	1	1	1
Son-Daughter	0	1	1	1
Son-d. in law	1	1	1	1
Grand-Child	0	0	1	1
Parent	1	0	1	1
Parent-in-law	0	0	1	1
Brother Sister	0	0	1	1
Co-Spouse	1	1	1	1
Other relative	1	1	1	1
Fost./Ad. child	1	1	1	1
Not related/DK	1	1	1	1
Other	1	1	1	1

elles sont reprises. Dans une vingtaine d'enquête, la plage d'éligibilité pour cette question est étendue à 17 ans (Benin 2006, Zimbabwe 2006, etc.) Deux personnes procurant le même numéro de ligne sont donc frères et sœurs. Cette étape concerne uniquement les individus éligibles âgés de 15 à 18 ans, pour un nombre limité d'enquêtes.

3. Nous examinons ensuite le numéro de ligne du conjoint fourni par les femmes éligibles [2]. Deux conjoints ne peuvent pas être frères et sœurs.

Comparaison des déclarations faites sur les fratries

Si deux individus éligibles sont frères ou sœurs l'un de l'autre, il est possible d'établir des recoupements entre la description qu'ils font respectivement de leur fratrie. Pour que deux enquêtés soient frères et sœurs l'un de l'autre, il faut également qu'ils se soient mutuellement mentionnées dans leurs fratries respectives.

1. Tout d'abord, les déclarations faites sur les différents frères et sœurs par chaque individu éligible peuvent servir à établir si les fratries déclarées sont *compatibles avec la présence d'un frère ou d'une sœur dans le ménage*. En d'autres termes, pour chaque paire d'observations, nous cherchons, dans la fratrie déclarée par le premier individu, la présence d'un frère de même âge et de même sexe que le second individu qui forme la paire. Comme les imprécisions sur les âges sont fréquentes, des erreurs sont autorisées jusqu'à 2.5 ans.
2. Il s'agit ensuite de comparer entre elles les déclarations faites sur les frères et sœurs pour établir si elles sont suffisamment similaires que pour pouvoir être considérées comme *se référant à une même fratrie*. A nouveau, il faut tenir compte de la présence de déclarations approximatives dans les fichiers [3]. Le score de ressemblance/distance entre les déclarations doit donc être construit en autorisant une certaine marge d'erreurs et d'imprécisions.
   Pour ce faire, nous nous inspirons de l'*analyse des séquences*, une méthode de plus en plus utilisée en sciences sociales (Abbott 1995). L'idée consiste ici simplement à restructurer l'information fournie par les enquêtés en différentes séquences : les séquences des *naissances* par sexe et années, d'une part, et les séquences des *décès* par sexe et années d'autre part. Ensuite, nous calculons les distances qui séparent les différentes séquences déclarées par deux individus formant une paire. Cela permet de comparer les différentes fratries déclarées en retenant un maximum de l'information qu'elles contiennent.
   Par exemple, imaginons une fratrie fictive détaillée dans le tableau A.2.

2. Cette information n'est disponible que pour les enquêtes postérieures à 1993

3. D'une part, les enquêtés peuvent fournir des informations légèrement différentes sur une même fratrie (par exemple l'un d'entre eux peut omettre un frère , inclure un cousin, oublier un décès, etc.). D'autre part, les procédures d'imputation qui sont menées après la collecte (notamment sur les dates de naissance et de décès) peuvent avoir généré des informations légèrement différentes pour une même fratrie.

Table A.2 – *Exemple de fratrie déclarée dans le fichier individuel*

	Sexe	Statut	Année de naissance	Année de décès
Enquêtée	F	Survivant	1959	
1	H	Survivant	1962	
2	F	Survivant	1963	
3	H	Survivant	1965	
4	H	Survivant	1968	
5	H	Survivant	1971	
6	F	Décédé	1978	1981
7	H	Décédé	1980	1982
8	H	Survivant	1983	
9	F	Survivant	1985	

Nous pouvons l'exprimer en quatre séquences, après avoir intégré le répondant dans la fratrie (en fonction de son rang de naissance). Si chaque élément des séquences correspond à une année entre 1950 et 1980, cela donne ceci :

– une *séquence des naissances masculines* par année (obtenue en enchaînant le nombre de naissances masculines survenues dans la fratrie) :

000000000000011122233344444444455566666666666666666

– une *séquence des naissances féminines* par année :

000000000111122222222222222333333344444444444444

– une *séquence des décès masculins* par année :

000000000000000000000000000000001111111111111111

– une *séquence des décès féminins* par année :

000000000000000000000000000000001111111111111111

Pour chaque paire d'individus éligibles, les différentes séquences sont ensuite comparées en prenant la *somme des différences des valeurs* pour chaque année [4].
Plus la distance est faible entre deux séquences, plus le score est bas, plus la probabilité qu'il s'agisse de la même fratrie est élevée.

L'objectif consiste ensuite à *convertir ces différents indicateurs de concordance en une probabilité unique* que la paire de déclarations renvoie bien à une même fratrie. Comment procéder et quel seuil retenir pour déterminer si, à partir de la distance calculée, les déclarations sont suffisamment

4. Comme un accord parfait entre les déclarations est plus difficile à obtenir lorsque les séquences sont plus longues (c'est-à-dire pour les fratries plus anciennes), cette distance est ensuite rapportée au nombre maximal de caractères observés dans les deux séquences.

proches ? Quel poids accorder aux erreurs faites sur les sœurs par rapport aux frères, par exemple ? Ces questions sont traitées au moyen d'un appariement 'probabiliste'.

Appariement 'probabiliste' des fratries dupliquées

Les résultats de l'appariement réalisé sur base des étapes précédentes peuvent être utilisés comme *standard*. En prenant comme exemple l'enquête EDS menée au Kenya en 1998, on peut regrouper les paires dans différentes catégories (cfr. tableau A.3). A ce stade, on peut clairement isoler deux groupes :

- d'une part, pour une majorité de paires d'observations (plus de 40% des paires d'individus éligibles vivant sous le même toit), on peut être quasiment certain qu'elles ne correspondent pas à une même fratrie (*false matched pairs*) ; les déclarations ne sont pas compatibles avec la présence d'un frère ou d'une sœur de même âge dans le ménage, et les relations entretenues avec le chef de ménage sont incompatibles.
- d'autre part, pour une petite minorité de paires (moins de 2%), on sait déjà qu'il s'agit très certainement d'une même fratrie (*true matched pairs*) ; les déclarations sont toutes deux compatibles avec le présence d'un frère ou d'une sœur de même âge *et* les relations avec le chef de ménage correspondent à des relations de fratries.

TABLE A.3 – *Appariement selon la relation avec le chef de ménage et la compatibilité des déclarations (Exemple de l'enquête EDS Kenya 1998)*

		Compatibilité des déclarations avec la présence d'un frère de même âge		
		Aucune	Une des deux	Les deux
Appariement	Improbable	**40.8**	5.8	1.0
	Indéterminé	15.9	10.8	23.4
	Probable	0.3	0.3	**1.8**

En ne retenant temporairement que ces deux groupes, on peut utiliser le degré d'accord (ou de proximité) observé entre les déclarations qui font vraisemblablement référence à deux fratries différentes (les *false matched pairs*) et le degré d'accord observé entre les déclarations se référant à une même fratrie (les *true matched pairs*) pour calculer une probabilité qu'une paire donnée fasse ou non référence à une fratrie unique. L'idée étant bien entendu de décider s'il faut apparier ou non les paires indéterminées. Ce faisant, on tient compte de la possibilité que les déclarations portant sur une même fratrie soient partiellement entachées d'erreurs.

Une régression logistique peut être utilisée pour mettre en relation les différents indicateurs de discordance et l'appartenance à l'un des deux groupes (*true/false matched pair*). Ensuite, en retenant les coefficients de cette régression, un *score de propension* peut être recalculé pour toutes les paires d'observations. Ce score, qui varie de 0 à 1, renvoie à la probabilité que deux déclarations fassent référence à une même fratrie, compte tenu

des légères différences qui les séparent. Une très large majorité des paires obtiennent des scores très faibles (inférieurs à 1%). Par contre les paires de fratries qui ont été identifiées comme telles de façon déterministe présentent en moyenne des valeurs avoisonnant 75 à 80%. En retenant un seuil - prudent - de 80%, on peut donc classer toutes les paires dont le score de propension est supérieur à 0.8 comme étant des déclarations se réfèrent à une même fratrie [5].

Il reste ensuite à utiliser les règles de transitivité pour regrouper les différents frères et sœurs au sein de *fratries uniques* sur l'ensemble du fichier. En d'autres termes, si *A est frère de B*, que *B est frère de C*, et que *B est frère de C*, A, B et C devraient pouvoir être regroupés dans une même fratrie.

Les résultats de cet appariement sur les différentes enquêtes EDS sont détaillés dans l'encadré 3, page 38.

5. Sauf si les étapes antérieures ont classé cette paire parmi les relations de fratrie improbables.

A.2 Annexes au chapitre 2

A.2.1 Comparaison des différentes options disponibles pour modéliser la fécondité en micro-simulation

Cette section détaille les différentes options disponibles dans SOCSIM pour modéliser la fécondité, et précise les choix que nous avons retenus. Afin d'examiner ces différentes options, nous avons comparé cinq séries de simulations, à partir des mêmes taux de fécondité et de mortalité, mais **en modifiant la façon avec laquelle la fécondité est introduite dans les simulations**.

Les taux de fécondité de départ sont issus de l'enquête EDS menée au Congo-RDC en 2007[6]. Les taux ont été calculés en biographique, pour les cinq ans précédant l'enquête, à partir d'un fichier restructuré en personnes-années. Ils sont estimés par âge, puis par âge et par rang. Ces différents taux sont présentés dans le graphe A.1. Les lignes de couleur renvoient aux taux de fécondité par rang, et les carrés correspondent aux taux toutes parités confondues[7].

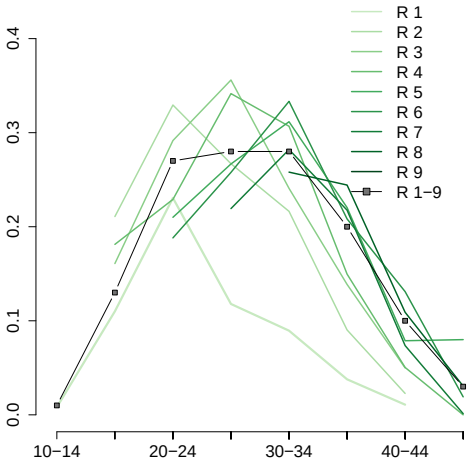

Figure A.1 – *Taux de fécondité par rang calculés à partir de l'enquête EDS menée au Congo RDC en 2007*

En maintenant ces taux constants, nous pouvons simuler une population stable, soumise à un schéma unique et invariable de mortalité. Nous retenons le modèle Ouest des tables de Princeton, en fixant le niveau d'espérance de vie à 47.5 ans (soit le niveau estimé par les Nations Unies pour 2005-2010 au Congo RDC)[8]. Pour modéliser la fécondité, cinq configurations différentes sont retenues :

6. Cette enquête est retenue car la fécondité a peu évolué sur les cinquante dernières années, et on peut donc maintenir des taux constants dans les simulations. Selon les Nations Unies, l'ISF s'élevait à 5.98 en 1950-1955, et à 6.07 en 2000-2005.

7. Les parités supérieures à 7 sont regroupées. Les groupes d'âge et de parité où il y a moins de 100 personnes-années ne sont pas repris dans le graphe.

8. Nous cherchons à nous approcher de la démographie de la RDC pour disposer d'un exemple, tout en étant bien conscient que nous faisons l'hypothèse de la stabilité de la population.

1. selon que les taux de fécondité sont variables par âge uniquement ou par âge et par parité,
2. selon qu'une durée minimale entre naissances de 12 mois est introduite ou non,
3. selon qu'une hétérogénéité de la fécondité est introduite ou non.

Le tableau A.4 détaille les paramètres de ces cinq configurations. Les trois premières sont basées sur les taux par âge (ASFR), les deux autres sur les taux par âge et parité (PASFR). Un intervalle minimal de 12 mois entre chaque naissance est retenu pour la deuxième configuration uniquement. Une hétérogénéité des taux de fécondité est introduite dans les configurations 3 et 5. Pour chacune de ces cinq configurations, 20 simulations différentes sont menées sur l'espace de 100 ans, et les populations finales sont rassemblées dans une seule population. En moyenne, plus de 330 000 naissances sont générées pour chacune des cinq configurations, ce qui permet de limiter la variation aléatoire autour des taux recalculés.

	DHS	SIM(1)	SIM(2)	SIM(3)	SIM(4)	SIM(5)
	a	*b*	*c*	*d*	*e*	*f*
Taux	DHS	ASFR	ASFR	ASFR	PASFR	PASFR
Int. min.		0	12	0	0	0
Hétérog.		0	0	1	0	1
ISF(15-49)	6.42	6.44	6.48	6.5	6.31	6.32
10-14	0.01	0.01	0.01	0.01	0.01	0.01
15-19	0.13	0.13	0.13	0.13	0.13	0.14
20-24	0.27	0.27	0.28	0.28	0.27	0.26
25-29	0.28	0.28	0.28	0.28	0.28	0.26
30-34	0.28	0.28	0.28	0.28	0.27	0.26
35-39	0.2	0.2	0.2	0.21	0.2	0.21
40-44	0.1	0.1	0.09	0.1	0.09	0.1
45-49	0.03	0.03	0.03	0.03	0.03	0.03
n		348318	357361	347672	304750	299639

TABLE A.4 – *Comparaison des différentes configurations (ISF et taux de fécondité par âge)*

A partir des populations finales, nous recalculons en biographique des taux de fécondité par âge, ainsi que des taux par âge et par rang. Nous présenterons ensuite différents indices, comme le nombre moyen d'enfants par femme, la distribution des parités atteintes, etc. Examinons d'abord les ISF et les taux par âge qui correspondent aux différentes options (ils sont repris dans le tableau A.4). Dans les deux premières configurations, les taux par âge issus des simulations sont **quasiment identiques à ceux introduits** (colonnes *b* et *c* à comparer avec la colonne *a*). Par contre, les taux recalculés **diffèrent davantage pour les autres simulations**, ce qui se traduit par quelques variations autour de l'ISF. Ces variations peuvent s'expliquer en examinant ce qui se passe lorsque les taux sont recalculés à la fois par âge et par rang. Ceci est présenté dans les figures A.2 à A.6, reprises dans les pages suivantes.

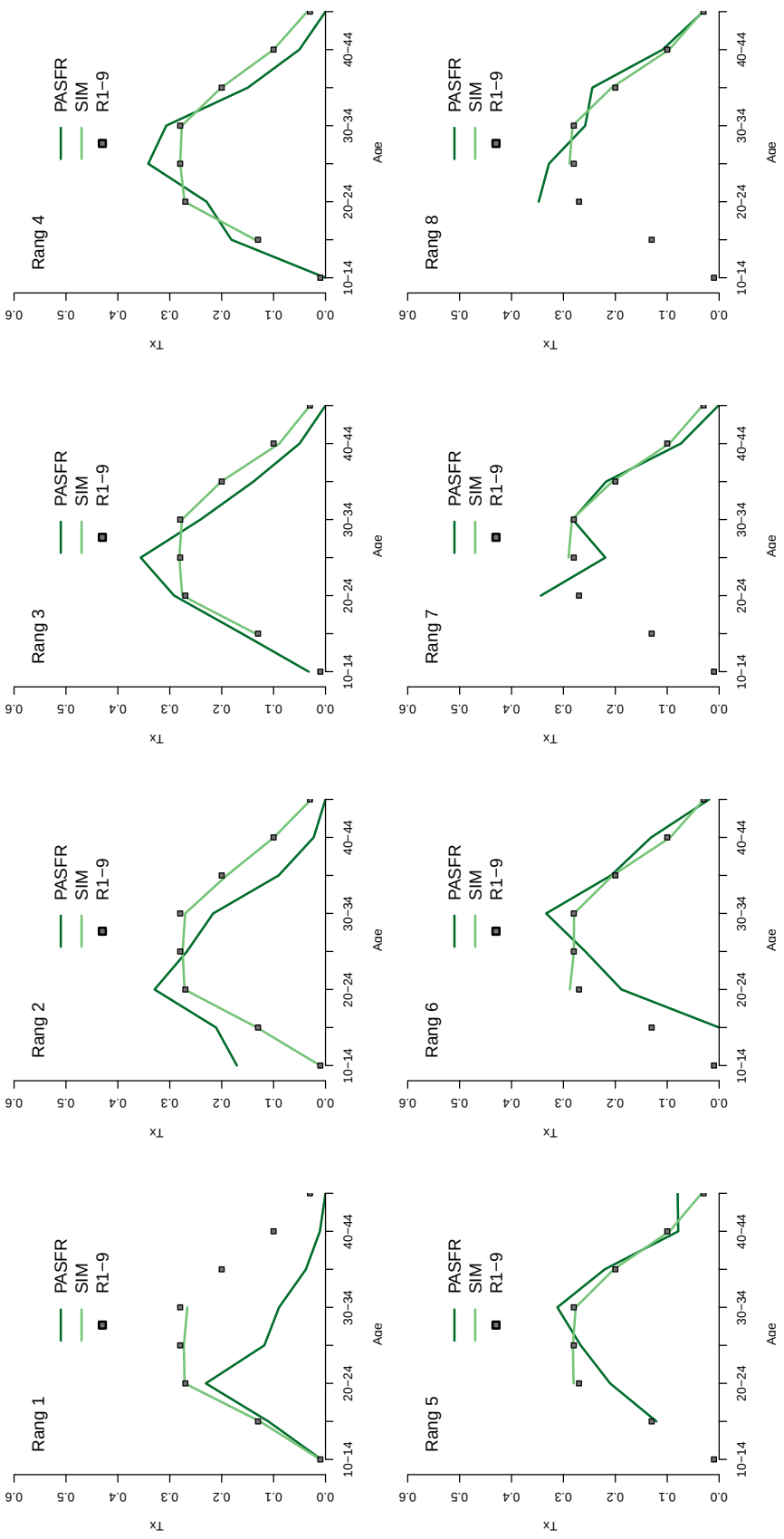

FIGURE A.2 – *Configuration 1 : Taux de fécondité par âge, pas d'intervalle minimal entre naissances, pas d'hétérogénéité de la fécondité*

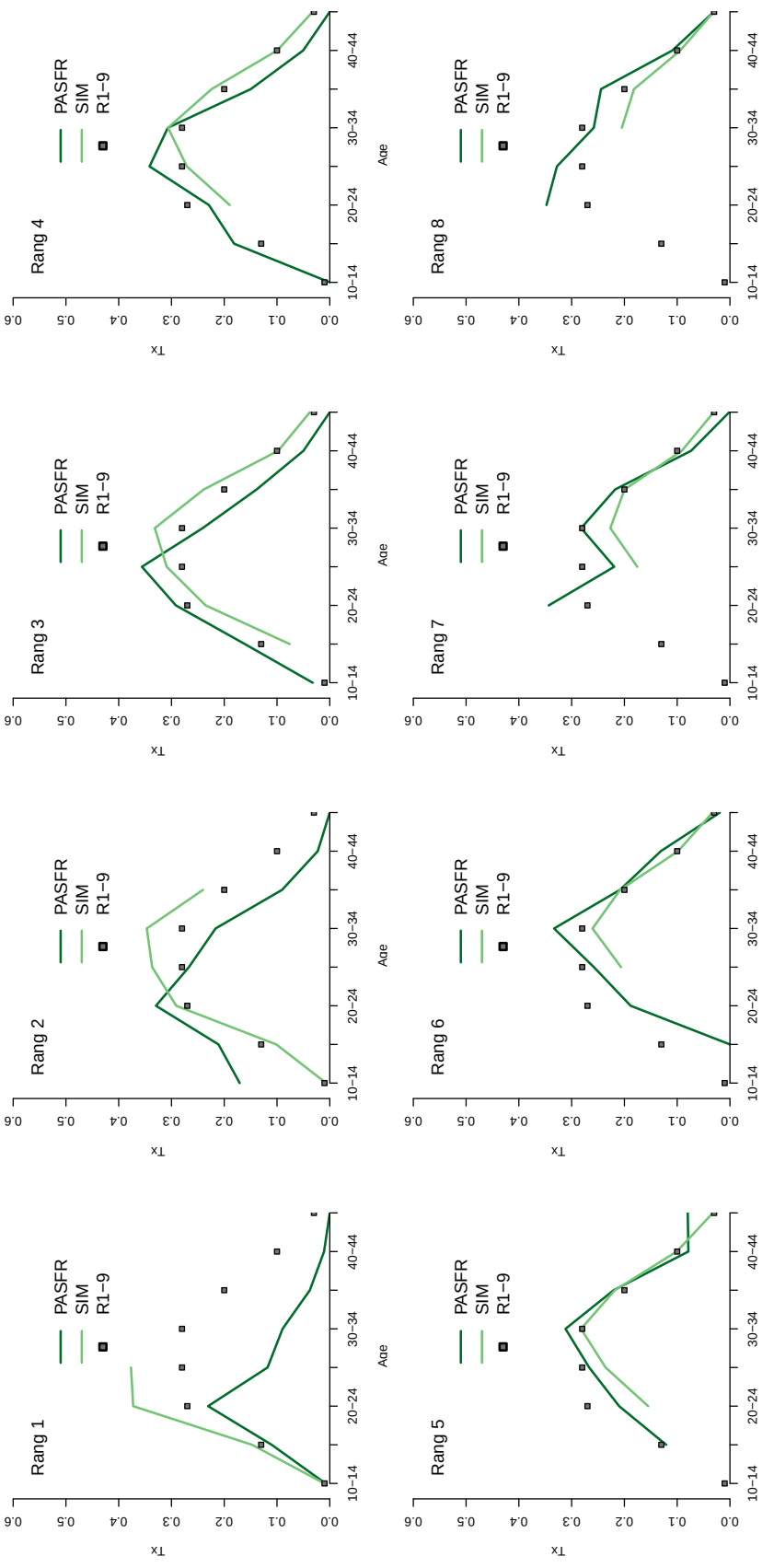

FIGURE A.3 – *Configuration 2 : Taux de fécondité par âge,* ***intervalle minimal entre naissances de 12 mois****, pas d'hétérogénéité de la fécondité*

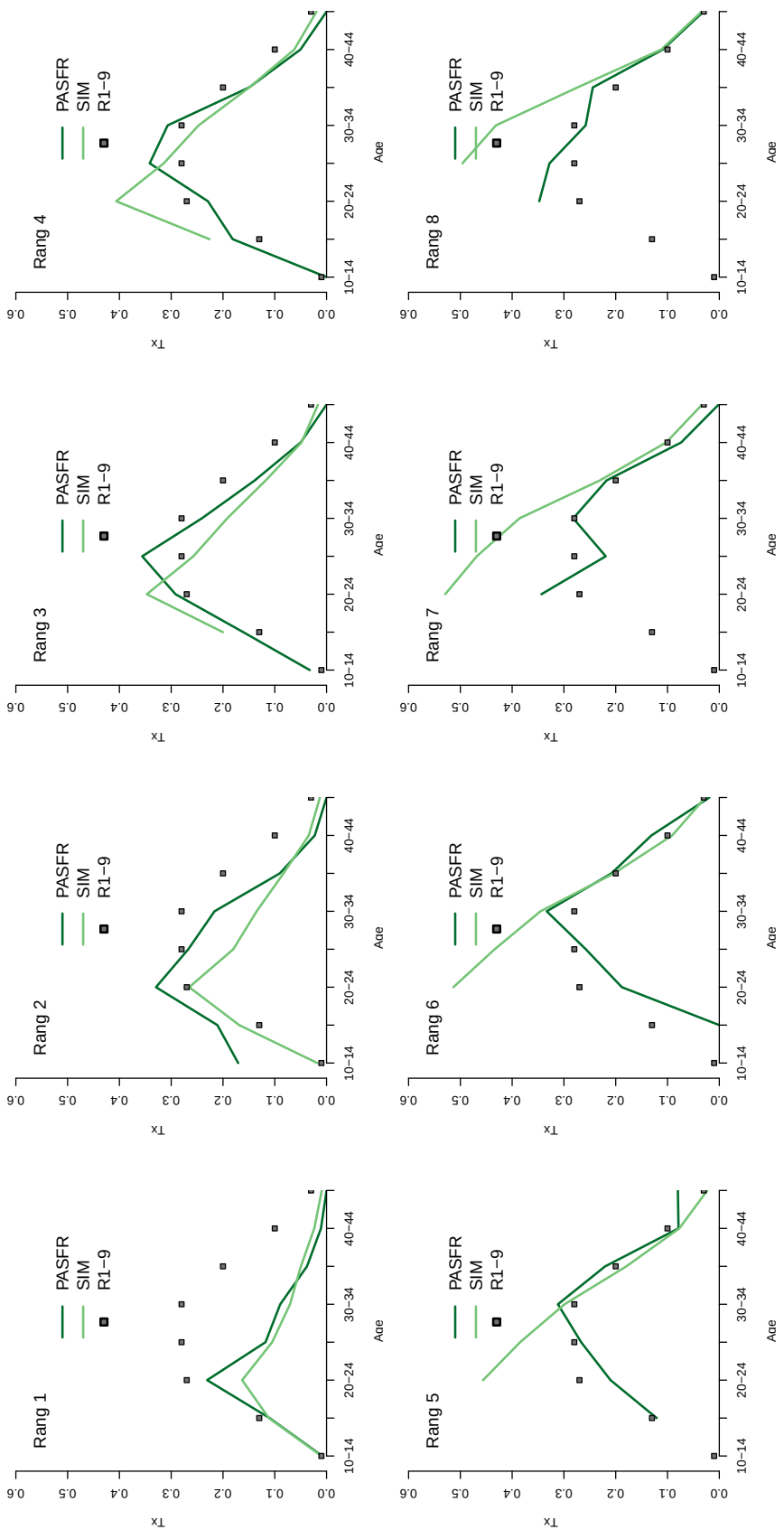

FIGURE A.4 – *Configuration 3 : Taux de fécondité par âge, pas d'intervalle minimal entre naissances,* ***hétérogénéité de la fécondité***

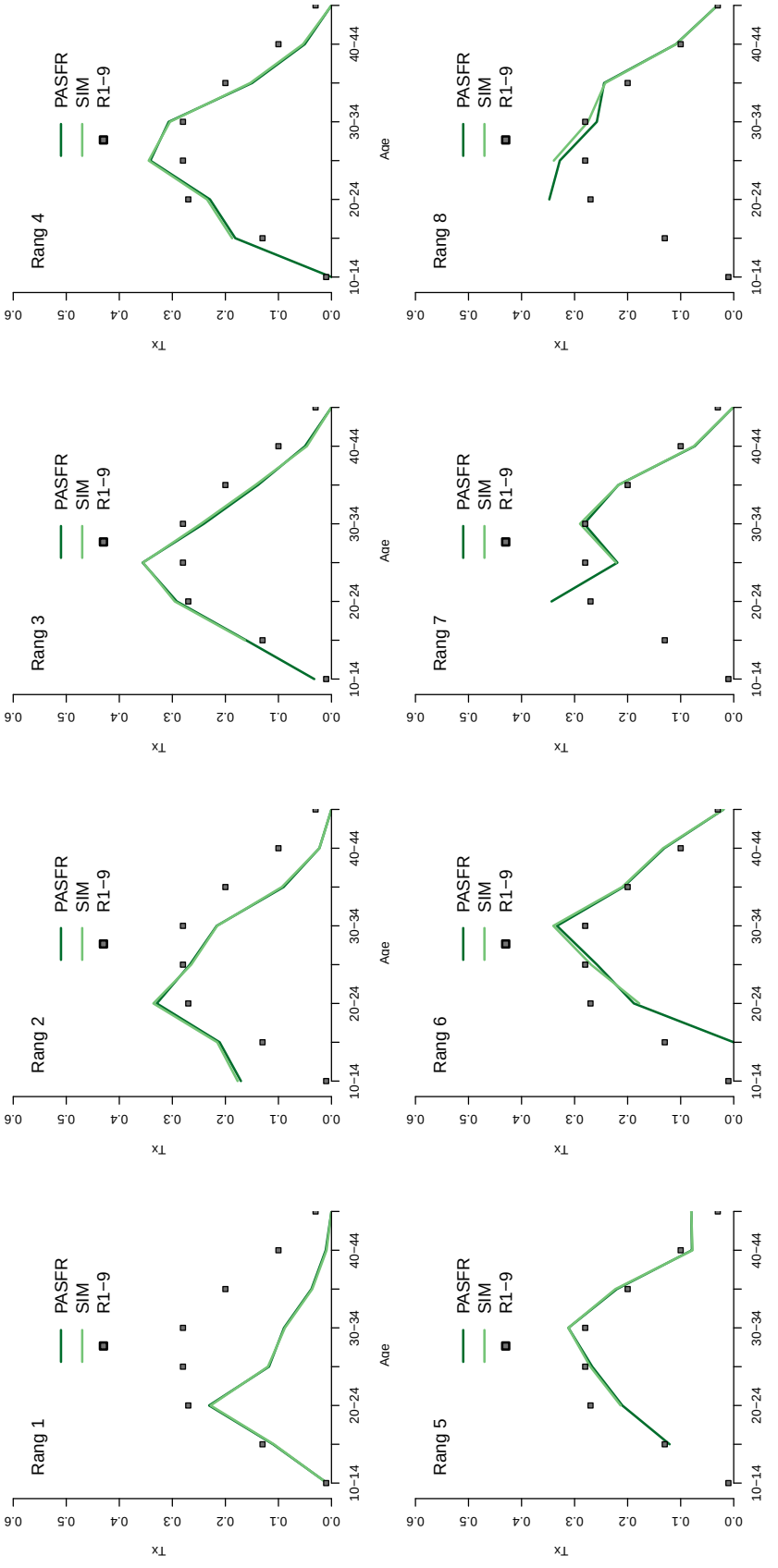

FIGURE A.5 – *Configuration 4 : Taux de fécondité par âge* ***et par parité****, pas d'intervalle minimal entre naissances, pas d'hétérogénéité de la fécondité*

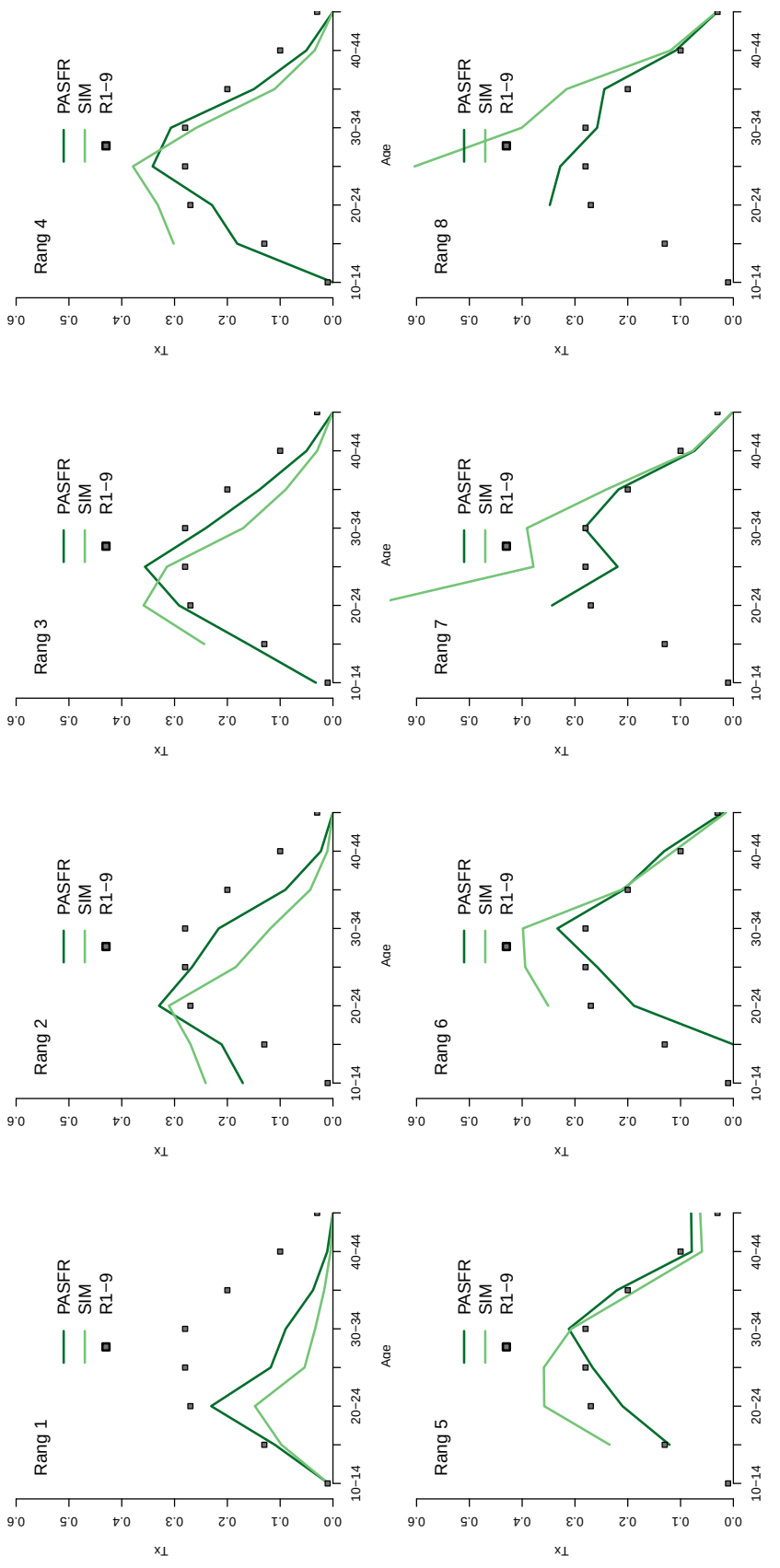

FIGURE A.6 – *Configuration 5 : Taux de fécondité par âge* ***et par parité****, pas d'intervalle minimal entre naissances,* ***hétérogénéité de la fécondité***

1. **Configuration 1 :** Dans la première série de simulations, les taux *introduits* sont simplement des taux par âge, sans intervalle intergénésiques minimal. Toutes les femmes sont soumises aux mêmes taux. Les taux de fécondité *recalculés à partir des simulations* (présentés dans le graphe A.2) sont donc simplement ces taux par âge, quelque soit la parité atteinte. On verra plus loin ce qu'il en résulte au niveau du nombre d'enfants nés vivants.

2. **Configuration 2 :** Dans le deuxième cas de figure, les taux *introduits* le sont toujours par âge uniquement, toutes les femmes sont soumises aux mêmes taux, mais un *intervalle intergénésique de 12 mois* au minimum est requis entre chaque naissance, afin de tenir compte des durées de gestation et d'aménorrhée postpartum. Si une grossesse survient dans les 12 mois, elle est simplement ignorée. En conséquence, SOCSIM doit légèrement ajuster les taux de fécondité à la hausse, au cours de la simulation, pour récupérer le nombre de naissances attendu. Après avoir réestimés les taux à partir des populations simulées avec cette option, on constate qu'ils correspondent bien à ceux introduits (tableau A.4). Toutefois, si l'on recalcule ces taux par âge et parité, **on observe une distribution par âge et parité assez variable, alors que la fécondité introduite ne varie que par âge**. En d'autres termes, imposer un intervalle minimal semble avoir des effets variables d'une parité à l'autre. On fait référence ici aux graphes A.3.

   Par exemple, la fécondité de rang 1 est sensiblement relevée, tandis que la fécondité pour les rangs supérieurs à 5 est réduite. Ce qui est en cause, c'est vraisemblablement l'ajustement par SOCSIM des taux de fécondité (pour compenser l'absence de naissance durant l'intervalle intergénésique). Cet ajustement semble introduire certaines distorsions sur les taux par parité. Notamment, la fécondité de rang 1 ne devrait pas être affectée par l'introduction d'un intervalle minimal, puisque ces femmes n'ont pas encore connu de grossesse. En conséquence, pour la suite, on préférera ne pas forcer d'intervalle intergénésique minimal. Ceci aura pour effet de sous-estimer notamment les écarts d'âges entre frères et sœurs.

3. **Configuration 3 :** En macro-simulation, toutes les femmes sont soumises aux mêmes risques de grossesse au même moment. Cela n'est pas nécessairement le cas en micro-simulation. Il est possible d'introduire une **hétérogénéité** de la fécondité, ce que nous faisons avec la troisième série de simulations. Ceci permet d'augmenter la variance des parités et des tailles de fratrie. Dans SOCSIM, un nombre aléatoire, alloué à chaque femme au moment de sa naissance, est utilisé comme un facteur multiplicatif des taux ; il va diminuer ou rehausser les taux de fécondité sans en modifier la structure par âge. Ce facteur multiplicatif est distribué comme une approximation cubique de la fonction béta, qui est souvent utilisée pour modéliser l'hétérogénéité de la fécondabilité (Weinberg et Gladen 1986, Murphy 2006). La distribution retenue dans le programme a une moyenne de 1, une

variance de 0.416 et des valeurs allant de 0 à 2.4. Elle est présentée dans le graphe A.7.

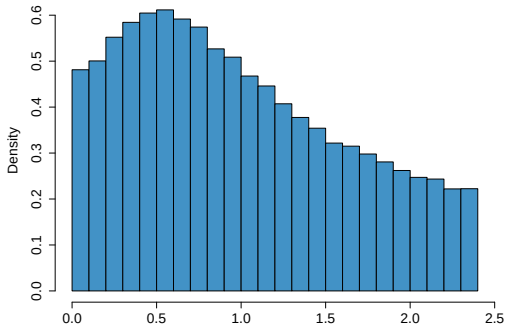

FIGURE A.7 – *Distribution du facteur multiplicatif de la fécondité utilisé dans* SOCSIM *pour accroître la variance des tailles de fratries*

Comme la moyenne de ces facteurs multiplicatifs vaut 1, une hétérogénéité est introduite sans que la fécondité moyenne de la population ne soit modifiée. On observe effectivement dans le tableau A.4 que les taux recalculés en biographique à partir des simulations sont quasiment identiques à ceux introduits. L'ISF n'est que légèrement supérieur à celui des taux introduits. Par contre, lorsque l'on recalcule les taux par âge et par rang à partir des populations finales, on constate que **ces facteurs multiplicatifs vont diminuer la fécondité pour les premiers et seconds rangs, et la rehausser pour les parités élevées**. On a donc là l'effet inverse de celui observé avec l'intervalle intergénésique minimal. Ceci s'explique facilement. Imaginons que ces facteurs multiplicatifs reflètent la fécondabilité. Les femmes plus "fécondes" (celles qui ont des facteurs multiplicatifs plus élevés) vont avancer plus vite dans les parités. Dès lors, comme la "fécondabilité" est fixée une fois pour toutes, on trouvera, chez les femmes de parités élevées, majoritairement des femmes plus fécondes. A l'inverse, à 30-34 ans, les femmes qui n'ont toujours pas eu d'enfants sont pour une bonne part des femmes dont la "fécondabilité" est moins élevée. Ceci explique pourquoi les schémas par âge et par rang de la fécondité soient si différents (graphes A.4), alors qu'à la base, les taux introduits varient uniquement par âge [9].

4. **Configuration 4 :** Le quatrième cas de figure est celui où la fécondité est directement introduite par âge et par rang. C'est sans doute la situation idéale, mais aussi la plus exigente en termes de données empiriques. Comme on l'observe sur les graphes A.5, les taux recalculés par âge et rang sont bien ceux introduits. Par contre, lorsque l'on combine les naissances de tous rangs, cette série de simulations procure un ISF moins élevé que la première série, où les taux sont

9. Notons que comme les liens de parentés sont enregistrés par SOCSIM, les facteurs multiplicatifs des taux peuvent être partiellement transmis aux enfants. Ceci introduit une *transmission des comportements reproductifs* d'une génération à l'autre, mais comme nous l'avons précisé, la calibration des simulations est rendue plus délicate (Murphy 2006).

introduits par âge uniquement (tableau A.4, p. 395). La différence se chiffre à 0.14 enfant par femme au total (entre la première et la quatrième configuration). A nouveau, cet écart s'explique facilement. Dans chaque groupe d'âge, le taux de fécondité *tous rangs* n'est autre que la moyenne des *taux par rang*, moyenne pondérée par la répartition par parité atteinte des femmes de ce groupe d'âge. En d'autres termes, les taux par rang et la répartition des femmes par parité interagissent pour produire des taux tous rangs. Or, la répartition des femmes par parité qui est observée dans l'enquête de 2007 n'est pas nécessairement celle des simulations, puisque ces dernières sont menées sur des populations stables. Si la fécondité a baissé, même légèrement, on trouvera davantage de femmes à parité élevées dans les groupes d'âge élévés que ne le laisse supposer l'application des taux de fécondité de 2002-2007 durant toute une vie reproductive. Pour cette raison, comme la fécondité des femmes de parité élevée est plus forte dans ces groupes d'âges, la structure par parité "stable" génère moins de naissances. Nous revenons sur ce problème dans la section 2.3.3, page 69.

5. **Configuration 5 :** Le dernier cas de figure est celui où la fécondité est introduite par âge et par rang, et où, de plus, une hétérogénéité de la fécondité est introduite. On peut alors faire les mêmes constats que pour la troisième série de simulations. En deux mots, comme les femmes plus "fécondes" avancent plus rapidement vers les parités élevées, la "fécondabilité" moyenne s'élève avec la parité atteinte, et la fécondité y est donc nettement supérieure à celle qui a été introduite.

En résumé, les séries de microsimulations 1, 2, 3 et 5 distribuent les naissances par âge et par rang de façon irréaliste [10] - même si les taux par âge recalculés sont proches de ceux introduits.

	DHS	SIM(1)	SIM(2)	SIM(3)	SIM(4)	SIM(5)
Taux	DHS	ASFR	ASFR	ASFR	PASFR	PASFR
Int. min.		0	12	0	0	0
Hétérog.		0	0	1	0	1
MACB	26.61	26.88	26.8	26.88	26.73	26.86
CEB(45-49)	6.62	6.43	6.45	6.5	6.34	6.33
VAR_CEB(45-49)	9.53	6.61	3.94	24.45	10.34	28.81

TABLE A.5 – *Comparaison des différentes configurations (Nombre d'enfants nés vivants et âge moyen à la naissance des enfants)*

Ceci se répercute sur la **distribution du nombre d'enfants nés vivants**, et donc également sur la distribution des tailles de fratries. Le tableau A.5 reprend, pour les femmes de 45-49 ans, le nombre moyen d'enfants nés vivants observé dans les différentes populations simulées. A titre indicatif, ces valeurs sont comparées à celles observés dans l'enquête congolaise de 2007. Toutes les séries de simulations procurent un nombre

10. En d'autres mots, ils vont générer des écarts d'âges entre frères et sœurs ou des distributions d'enfants nés vivants qui ne correspondent pas à ceux que l'on peut observer dans la réalité.

moyen d'enfants nés vivants qui avoisine 6.4 enfants, ce qui est légèrement en dessous de la valeur observée dans l'enquête (6.62)[11]. Par contre, la *variance* du nombre d'enfants nés vivant diffère sensiblement d'une série de simulations à l'autre. Dans l'enquête, cette variance s'établit à 9.53 enfants par femmes. Seule la quatrième série de simulations, basée sur les taux par âge et parité, parvient à une valeur comparable. Dans les autres cas, la variance est soit trop faible (1-2) soit sensiblement trop élevée (3 et 5).

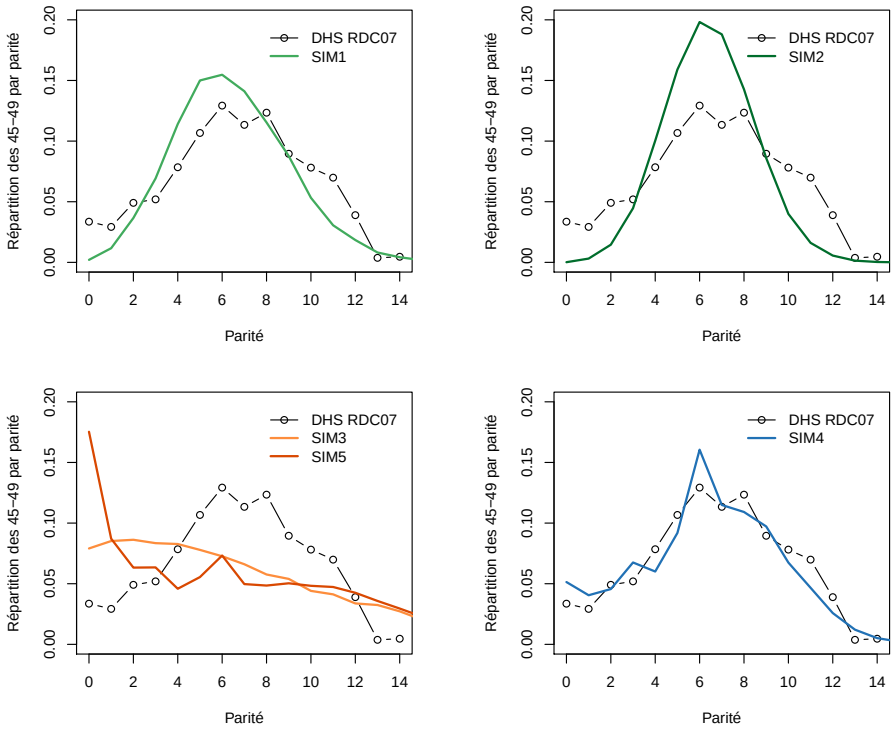

FIGURE A.8 – *Distribution du nombre d'enfants nés vivants (femmes de 45–49 ans) selon la série de simulation, et distribution observée dans l'enquête EDS Congo-RDC 2007*

Le graphe A.8 compare la distribution des femmes de 45-49 ans selon leur parité, telle qu'elle a été observée dans l'enquête EDS de 2007, avec ces mêmes distributions dégagées des populations simulées (en fonction des configurations retenues). Dans le cas où une hétérogénéité des taux de fécondité est introduite (séries 3 et 5 en bas à gauche), la répartition du nombre d'enfants nés vivants est implausible ; il y a beaucoup trop de femmes sans enfants et trop de femmes avec énormément d'enfants. La stratégie adoptée par SOCSIM pour augmenter la variance du nombre d'enfants nés vivants et des tailles de fratrie ne semble donc pas tout à fait appropriée, en tous cas pas pour les populations à forte fécondité comme au Congo RDC. Dans les séries 1 et 2, basées sur les taux par âge, c'est l'inverse : ces simulations sous-estiment le nombre de femmes qui restent sans enfants, ainsi que celles qui atteignent des parités élevées. La quatrième série cadre beaucoup mieux avec la distribution observée dans

11. A nouveau, cette différence peut s'expliquer par une baisse de la fécondité en RDC, alors que les simulations sont menées sur une population stable.

l'enquête EDS.

Une dernière comparaison entre les différentes options est fournie dans le graphe A.9. Celui-ci présente l'âge moyen à la naissance des enfants, calculé *par parité atteinte* (comme le font Pennec et Bacon (2007)). La quatrième série de simulations procure des âges moyens à la naissances identiques à ceux observés dans l'enquête. Par contre, pour les autres séries de simulations, les âges moyens sont soient trop élevés (1 et 2), soit trop bas (3 et 5).

En définitive, il apparaît essentiel de tenir compte de la variation de la fécondité par parité, surtout si les microsimulations doivent apporter des informations qui vont au-delà du nombre "moyen" d'enfants nés vivants, ou du nombre de frères et sœurs.

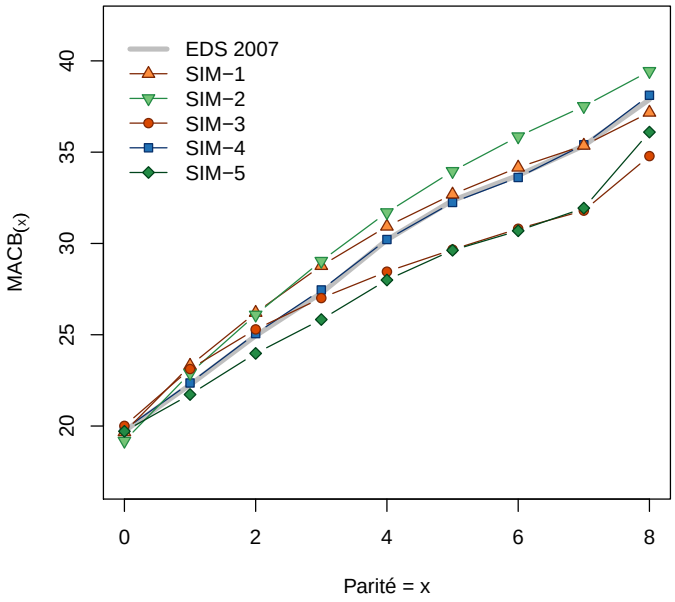

FIGURE A.9 – *Ages moyens à la naissance des enfants selon la parité atteinte*

A.2.2 Notes sur le modèle de Lee et Carter (1992)

Afin de procéder à des projections, Lee et Carter (1992) ajustent le logarithme des taux de mortalité par âge au moyen de l'expression suivante :

$$ln(m_{xt}) = a_x + K(T) \times b_x + \epsilon_{x,t}$$

Le paramètre $K(T)$ est l'indicateur de tendance (non observée) qui traduit l'évolution de l'intensité de la mortalité (tous âges) au temps t. Le paramètre a_x renvoie au calendrier de la mortalité par âge. Le paramètre b_x reflète la modification dans le temps de ce calendrier par âge. En d'autres termes, la distribution par âge des valeurs de b_x définit quels sont les taux qui vont augmenter/décliner rapidement et ceux qui vont augmenter/décliner moins rapidement.

Les auteurs ajustent ce modèle sur les taux de mortalité des Etats-Unis de 1933 à 1987. Ils montrent que le paramètre $K(T)$ a baissé linéairement durant le 20e siècle, à un taux constant et en conservant une variabilité relativement constante. Cet indicateur de tendance globale connaît donc des variations qui sont plus facilement appréhendées que les variations que connaît l'espérance de vie. Celle-ci n'a pas connu, sur les données américaines, de hausse linéaire, et sa variabilité a baissé sur l'intervalle. Cela rend le modèle assez attractif pour mener des projections. [12]

12. Dans la pratique, nous utilisons ce modèle pour mener les micro-simulations parce qu'il facilite le travail : un seul fichier reprenant les trois paramètres $K(t)$, a_x et b_x doit être fourni à SOCSIM, plutôt que de spécifier tous les taux par âge et année.

A.2.3 Comparaison des taux de croissance équivalents aux populations stables (paramètres de 1950-1955) avec les taux de croissance estimés en 1950-1955 par les Nations Unies

	Pays	r (stable)	r (UN)	r (diff)
1	Bénin	1.49	0.96	0.53
2	Afrique du Sud	2.81	2.30	0.51
3	Guinée	2.08	1.68	0.40
4	Kenya	3.15	2.76	0.39
5	Madagascar	2.48	2.15	0.33
6	Erythrée	2.36	2.04	0.32
7	Guinée Equ.	1.35	1.05	0.30
8	Centrafique	1.35	1.07	0.28
9	Gabon	0.52	0.28	0.24
10	Namibie	2.28	2.04	0.24
11	Lesotho	2.17	1.94	0.23
12	Somalie	2.35	2.16	0.19
13	Ethiopie	2.13	1.95	0.18
14	Djibouti	2.34	2.17	0.17
15	Togo	2.42	2.25	0.17
16	Tchad	2.11	1.94	0.17
17	Angola	1.96	1.81	0.15
18	Botswana	2.96	2.82	0.14
19	Guinée-Bissau	2.29	2.16	0.13
20	Mali	1.85	1.75	0.10
21	Nigeria	2.12	2.03	0.09
22	Ghana	2.53	2.45	0.08
23	Sénégal	1.98	1.93	0.05
24	Swaziland	2.59	2.55	0.04
25	Zimbabwe	3.13	3.09	0.04
26	Cameroun	1.89	1.85	0.04
27	Mozambique	1.68	1.65	0.03
28	Liberia	2.10	2.07	0.03
29	Zambie	2.53	2.50	0.03
30	Comores	2.29	2.28	0.01
31	Congo RDC	2.18	2.18	0.00
32	Ouganda	2.68	2.68	0.00
33	Congo	2.09	2.10	-0.01
34	Tanzanie	2.65	2.66	-0.01
35	Côte d'Ivoire	2.40	2.46	-0.06
36	Malawi	2.07	2.14	-0.07
37	Burundi	2.25	2.33	-0.08
38	Gambie	1.33	1.50	-0.17
39	Burkina Faso	1.64	1.82	-0.18
40	Cap Vert	3.02	3.23	-0.21
41	Niger	2.35	2.72	-0.37

TABLE A.6 – *(1) Taux de croissance intrinsèque de la population stable équivalente aux taux de 1950-1955, (2) taux de croissance estimé par les Nations Unies en 1950-1955, (3) différence entre ces deux séries (en %)*

A.2.4 Notes sur les fonctions splines

Les fonctions splines sont des fonctions polynômiales par morceaux. Elles peuvent être utilisées pour interpoler ou lisser des données. Différentes options sont possibles, en fonction des spécificités des données.

De façon générale, il faut un intervalle $[a, b]$ comptant n points clés, appelés "noeuds", agencés en ordre tels que $a = x_0 < x_1 ... < x_{n+1} = b$. Lorsqu'il s'agit d'une *spline cubique* - ce qui est le cas le plus commun - deux points de jonction définissent un sous-intervalle sur lequel vient s'ajuster un polynôme de type $y = a + bx + cx^2 + dx^3$. Pour tout x tel que $x_0 \leq x_1 \leq x_{n+1}$, la fonction s'établit comme $S(x) = p_i(x)$ où $p_i(x)$ est un polynôme en x. Un certain nombre de contraintes sont imposées à la fonction pour garantir une certaine régularité ; elle doit être *continue* aux points de jonction (les noeuds), elle est *dérivable*, et sa *première et deuxième dérivée* doivent être également continues au niveau des noeuds. Cela va permettre à la pente de la courbe qui précède un noeud de correspondre à la pente de la courbe qui succède à ce noeud, ce qui procure un résultat lissé et continu (Siegel et al. 2004).

Il existe de nombreux types de fonctions splines et nous y recourons souvent pour cette thèse. Les *splines naturelles*, très utilisées, sont par exemple caractérisées par le fait que la seconde dérivée est nulle aux extrémités de l'intervalle d'interpolation, ce qui procure une extrapolation linéaire hors de cet intervalle. A côté des splines d'interpolation, les *splines de lissage* sont également très utilisées. Elles minimisent, pour toute double dérivée S'' en $[a, b]$, la somme suivante : $\sum w_i[y_i - S(x_i)]^2 + \lambda \int_a^b (S''(x))^2 dx$. Il s'agit en fait de minimiser la somme des carrés des résidus, mais en ajoutant un terme $\int_a^b (S''(x))^2 dx$, qui est appelé "pénalité" (*roughness penalty*) (Green et Silverman 1994). Ce terme est multiplié par un paramètre de lissage λ. Comme son nom l'indique, ce dernier définit le degré de lissage de la fonction spline. Plus il est élevé, plus le lissage est important et l'ajustement brut. Plus il est faible, plus le lissage est faible et l'ajustement fin. Le premier terme détermine donc la proximité aux données et le second la régularité de la fonction. Entre ces deux exigences, un équilibre est à chercher. Généralement, on utilise par défaut le critère de "validation croisée" (*cross-validation*) (Faraway 2006).

A.2.5 Distribution des décès infantiles utilisée dans les microsimulations

Comme SOCSIM utilise des taux de mortalité (et pas des quotients), quelques problèmes peuvent se poser pour la mortalité infantile, du fait que les décès ne sont pas uniformément distribués sur la première année mais très concentrés sur le premier mois. Nous répartissons donc les décès infantiles à partir d'une distribution des décès par mois observée dans les enquêtes EDS.

Cette distribution est présentée dans le graphe A.10 sous forme d'un boxplot, en combinant 93 enquêtes EDS menées en Afrique subsaharienne entre 1986 et 2009 [13]. Le boxplot A.11 présente la même information, sous forme du ratio entre le nombre de décès observés par mois sur le nombre de décès observés *dans le premier mois*. Les valeurs médianes sont retenues dans les microsimulations pour distribuer les décès au cours de la première année.

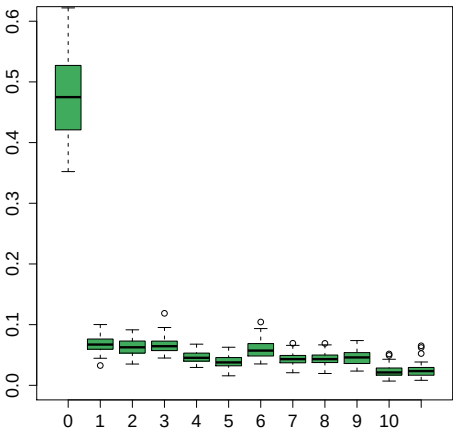

FIGURE A.10 – *Distribution des décès infantiles entre les 12 premiers mois observée dans les enquêtes EDS africaines (n = 93)*

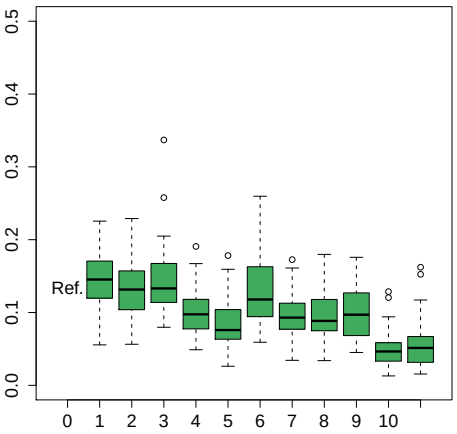

FIGURE A.11 – *Rapports entre le nombre décès observés par mois et le nombre de décès observés dans le 1^{er} mois (n = 93)*

13. Précisons que nous utilisons ici les dates de naissances et de décès déjà imputées.

A.2.6 Notes sur le calcul des taux de fécondité par rang à partir des rapports de taux

Il est clair que chaque taux "tous rangs" $f(x)$ peut s'exprimer comme une moyenne de taux par rang $f_i(x)$, moyenne pondérée par la répartition des femmes par parité atteinte, notée $w_i(x)$ [14] :

$$f(x) = \sum_{i=0}^{n} w_i(x) \times f_i(x) \tag{A.1}$$

Notons $r_1(x)$ le rapport entre la fécondité de rang 1 et la fécondité de rang 2, puis $r_3(x)$ le rapport entre la fécondité de rang 3 et la fécondité de rang 2, etc. La fécondité de rang 2 sert de référence ($r_2(x) = 1$). Ces rapports de taux sont fixes dans le temps. A partir de là, il est possible de calculer la valeur de $f_2(x)$ qui, combinée à la structure par âge des femmes par parité et à ces rapports $r_i(x)$, permet de retomber sur le taux de fécondité tous rangs :

$$\begin{aligned} f(x) &= w_0(x)f_1(x) + w_1(x)f_2(x) + w_2(x)f_3(x) + \dots \\ f(x) &= f_2(x) \times [w_0(x)r_1(x) + w_1(x)r_2(x) + w_2(x)r_3(x) + \dots] \\ f_2(x) &= \frac{f(x)}{[w_0(x)r_1(x) + w_1(x) + w_2(x)r_3(x) + \dots]} \end{aligned} \tag{A.2}$$

Cette valeur de $f_2(x)$ est ensuite combinée aux rapports de taux $r_i(x)$ pour générer les différents taux par rang.

Dans la pratique, il faut donc estimer les différentes valeurs de $w_i(x)$ pour pouvoir calculer les taux par rang. Au cours du premier segment des simulations (1850-1900), la fécondité est constante quelque soit la parité atteinte. Ceci va permettre de distribuer les femmes par parité. Pour tous les segments suivants (c'est-à-dire tous les cinq ans jusqu'à 1950 puis tous les ans ensuite), il s'agit d'estimer, à partir des populations simulées, la valeur de $w_i(x)$ pour la mi-période du segment suivant.

Cette valeur dépend bien sûr des taux introduits précédemment. Quand la fécondité ne change pas, la répartition des femmes par parité (pour chaque âge) peut simplement être observée au terme de chaque année, et les taux recalculés en conséquence. C'est le cas jusqu'en 1950. Après cela, comme la fécondité va varier dans le temps, la valeur $w_i(x,t)$ observée au temps t n'est plus une bonne approximation de la valeur qu'il faut introduire, c'est-à-dire de la valeur à la mi-période de l'année suivante, $w_i(x, t + 0.5)$. Par exemple, la structure par parité pour mi-1975 ne peut plus être directement déduite de la structure observée à la fin de l'année 1974.

Pour résoudre ce problème, au terme de chaque segment, nous estimons $w_i(x, t + 1)$ de la façon suivante, après avoir fixé des valeurs tempo-

14. Avec $\sum_{i=0}^{n} w_i(x) = 1$

raires de $f_i(x,t)$ [15] (Jasilioniene et al. 2009) :

$$\begin{aligned} w_0(x+1,t+1) &= w_0(x,t) - f_1(x,t) \\ w_i(x+1,t+1) &= w_i(x,t) - f_i(x,t) + f_{i+1}(x,t) \; si \; i = 1,2,3,... \\ w_{8+}(x+1,t+1) &= w_{8+}(x,t) + f_8(x,t) \end{aligned} \tag{A.3}$$

En démarrant à 14 ans ($w_0(14) = 1$) puis en procédant âge par âge, les différentes valeurs de $w_i(x,t+1)$ peuvent être calculés de cette façon. A partir de là, les valeurs de $w_i(x,t+0.5)$ peuvent être estimées (en prenant la moyenne de $w_i(x,t)$ et $w_i(x,t+1)$). Les différents taux $f_i(x,t)$ sont ensuite réajustés de façon à ce que la somme des $w_i(x) \times f_i(x)$ corresponde aux taux tous rangs.

15. En approximant $w_i(x,t+0.5)$ par $w_i(x,t)$.

A.2.7 Notes sur l'intégration de la nuptialité dans les micro-simulations africaines

Cette section présente les choix méthodologiques retenus pour intégrer la nuptialité dans les micro-simulations. Nous prévoyons des taux de nuptialité notamment parce que SOCSIM n'enregistre l'identifiant des pères que lorsque les parents sont mariés. A ce stade, seuls les premiers mariages sont considérés, faute de données ou de modèle disponible pour le divorce et le remariage.

Utilisation du modèle de Coale et McNeil (1972)

Il n'existe pas de base de données procurant des schémas de primo-nuptialité par âge au niveau national, couvrant l'ensemble des pays sur la période 1950-2010. Les différentes publications des Nations Unies fournissent uniquement des taux bruts de nuptialité, ainsi que des proportions de femmes et d'hommes déjà mariés par âge, ou bien directement des valeurs de l'âge moyen au premier mariage (UN 1988; 1990; 2000; 2004; 2009b) [16].

L'objectif consiste ici à calculer des taux de primo-nuptialité pour les différents segments des simulations, et pour chaque pays. Nous recourons au modèle de Coale et McNeil (1972), après avoir estimé les différents paramètres du modèle sur base de données empiriques. Il s'agit des valeurs de l'âge moyen au premier mariage, de l'écart-type de cet âge au mariage, et des fréquences du célibat définitif (encadré 13).

Encadré 13 Schéma de primo-nuptialité de Coale et McNeil (1972)

Coale (1971) a comparé les **évolutions par âge de la proportion de femmes mariées** pour de nombreuses cohortes. Il observe que ces proportions cumulées sont en fait assez proches les unes des autres lorsqu'elles subissent certaines transformations : il suffit de diviser les proportions de femmes mariées par âge par la proportion de femmes mariées à 50 ans puis de standardiser ces distributions de sorte qu'elles aient les mêmes moyennes et les mêmes écarts-types. Dans ce cas, les proportions cumulées par âge ainsi que les distributions de fréquences des premiers mariages sont quasiment identiques, quelles que soient les cohortes dont elles sont issues. Un **standard** est alors construit à partir des premiers mariages suédois observés en 1865-1869. Coale et McNeil (1972) cherchent ensuite à exprimer ce standard suédois sous forme mathématique. Ils constatent que la densité des premiers mariages peut s'exprimer sous la forme d'une double exponentielle :

$$G_s(x) = 0.19465 \times exp\left[-0.174(x - 6.06) - exp\left\{-0.288(x - 6.06)\right\}\right] \quad \text{(A.4)}$$

Tout comme le système des logits de Brass (*cf.* p.112), Coale (1971) suggère un modèle relationnel, qui relie la proportion de femmes ayant déjà eu un premier

16. On trouve des estimations récentes dans Hertrich (2007) et Tabutin et Schoumaker (2004), mais sous la forme d'âges *médians* au premier mariage, ce qui ne permet pas de reconstruire facilement des schémas par âge.

mariage $F(x)$, à ce standard (G_s) au moyen de quelques paramètres :

$$F(x) = C \times G_s \times (\frac{x - a_0}{\kappa}) \tag{A.5}$$

Le paramètre a_0 correspond à l'**âge auquel démarrent les premiers mariages** (c'est-à-dire l'âge auquel à peu près 1% des premiers mariages ont eu lieu). C correspond au complément de la **proportion de célibataires définitifs**, tandis que κ traduit la **rapidité des premiers mariages** après l'âge a_0 [17]. En connaissant les paramètres a_0 et κ, il est donc possible de reconstruire d'autres schémas de primo-nuptialité, en utilisant l'équation suivante :

$$g(x) = \frac{0.19465}{\kappa} \times exp\left[\frac{-0.174}{\kappa}(x - a_0 - 6.06\kappa) - exp\left\{\frac{-0.288}{\kappa}(x - a_0 - 6.06\kappa)\right\}\right] \tag{A.6}$$

La moyenne de la fonction de densité, qui est aussi l'âge moyen au premier mariage, vaut $S = a_0 + 11.36\kappa$. Sa variance vaut $Var(x) = 43.34\kappa^2$. Cela signifie qu'il suffit de définir un âge moyen au premier mariage et une valeur de a_0 pour obtenir un schéma de premier mariage par âge (graphe A.12). En effet, on peut établir l'équation suivante : $(S - a_0)/11.36 = \kappa$.

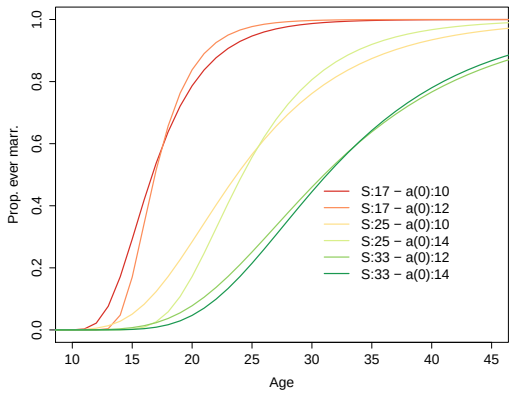

FIGURE A.12 – *Schémas de primo-nuptialité construits à partir de différentes valeurs de S et a_0, d'après le modèle de Coale et McNeil (1972)*

Afin de pouvoir recourir au modèle de Coale et McNeil (1972), nous avons rassemblé un grand nombre d'estimations, en combinant des distributions de célibataires par âge, ou - quand elles ne sont pas disponibles - des valeurs de l'âge moyen au premier mariage ainsi que des proportions de célibataires à 50 ans. Ces estimations proviennent de diverses sources de données, pour les 41 pays retenus dans les simulations (tableau A.7) :

1. les enquêtes EDS fournissent 89 distributions de célibataires par âge pour les femmes et 57 pour les hommes, pour les 41 pays qui nous concernent,

17. Il correspond plus précisément à la longueur de l'intervalle durant lequel la plupart des premiers mariages sont conclus. Il exprime cette longueur par rapport au standard dans lequel la moitié de la population qui va se marier est déjà mariée dix ans après l'âge a_0.

2. les enquêtes MICS (dont les fichiers individuels sont accessibles) fournissent 30 distributions par année d'âge pour les femmes [18],
3. une centaine d'estimations ont également été collectées dans diverses publications ; les travaux de Brass et al. (1968) et Van de Walle (1996), diverses publications des Nations Unies (UN 1988; 1990; 2000; 2004; 2009b), ainsi que deux suppléments historiques au *Demographic Yearbook* (UN 1979; 1997). Il s'agit pour l'essentiel de données de recensements ou d'autres programmes de collecte (WFS, etc.).

Les différentes valeurs collectées couvrent la période 1944-2008, mais la majorité des estimations sont postérieures à 1990. En moyenne, on dispose de 6 distributions par âge et par pays pour les femmes et de 4 pour les hommes sur l'ensemble de la période. Au total, nous avons rassemblé plus de 250 estimations différentes de l'âge moyen au premier mariage pour les femmes, et 181 pour les hommes.

Nous ne discuterons pas ici de la fiabilité de ces données collectées rétrospectivement sur le statut matrimonial, une question abordée notamment par Van de Walle (1996) et Hertrich (2007) [19].

	Femmes			Hommes		
Type	Dist	$prop_{50}$	SMAFM	Dist	$prop_{50}$	SMAFM
EDS	89	89	89	57	57	57
MICS	30	30	30	0	0	0
Autres	109	125	139	101	117	124
<1975	33	40	45	30	37	41
1975-90	50	59	65	38	47	48
1990-2008	145	145	148	90	90	92
Moyenne par pays	6	6	6	4	4	4
Total	228	244	258	158	174	181

TABLE A.7 – *Nombre de distributions de célibataires par âge (Dist), nombre de valeurs de l'âge moyen au premier mariage (SMAFM) et nombre de proportions de célibataires à 50 ans ($prop_{50}$), par sexe et selon les sources de données, pour les 41 pays retenus (de 1944 à 2010).*

Les cartes A.13 présentent **les valeurs de l'âge moyen au premier mariage**, pour les femmes puis les hommes, et ce, pour trois périodes différentes ; avant 1975, de 1975 à 1990 puis après 1990. Elles combinent les valeurs que nous avons estimées par la méthode d'Hajnal (1953) lorsque

18. L'enquête MICS Côte d'Ivoire 2006 n'est pas retenue en raison d'un manque de clarté sur le statut matrimonial.

19. Rappelons simplement que Van de Walle (1996) montre combien la définition du mariage et la place réservée aux unions consensuelles varient sensiblement d'une enquête EDS à l'autre, avec notamment de fortes disparités entre l'Afrique francophone et l'Afrique anglophone. Les définitions varient également selon le type d'opération de collecte, par exemple entre les enquêtes EDS et les recensements. Au niveau méthodologique, rappelons également que le calcul de l'âge au mariage par la méthode d'Hajnal (1953) suppose (1) l'indépendance entre nuptialité et mortalité ainsi qu'entre nuptialité et migration, (2) une constance des probabilités par année d'âge lorsque les données sont présentées par groupe d'âge, et (3) une constance des taux de primo-nuptialité sur les 35 dernières années. Dans les faits, les valeurs obtenues sont souvent issues d'un mélange entre les taux qui prévalaient dans le passé proche et ceux du passé lointain (Preston et al. 2001).

nous disposons des proportions de célibataires par âge [20] et les valeurs publiées lorsque nous n'en disposons pas. Chez les femmes, on observe un recul de l'âge au premier mariage quasi généralisé, qui survient d'abord en Afrique australe puis s'étend progressivement vers l'Afrique de l'Est et le Golfe de Guinée. Les changements sont par contre timides dans la région sahélienne ainsi qu'au Malawi et au Mozambique, où l'âge au mariage reste précoce. Au total, 7 pays présentent toujours des âges moyens au premier mariage inférieurs à 19 ans après 1990 [21]. Au final, on observe un découpage régional très net sur la carte en haut à droite du graphe A.13 (c'est-à-dire chez les femmes après 1990).

Les évolutions sont plus contrastées pour les hommes. Les niveaux observés avant 1975 sont plus hétérogènes que chez les femmes, ce qui rend moins aisée la lecture des tendances. C'est davantage une impression de stabilité qui se dégage sur la période, avec des schémas de primonuptialité très tardifs en Afrique australe, ainsi que dans quelques pays d'Afrique de l'Ouest (Nigeria, Côte d'Ivoire, Sénégal). L'âge au mariage semble par contre légèrement plus précoce en Afrique de l'Est et en Afrique sahélienne. Dans l'ensemble, la quasi-totalité des pays présentent des valeurs supérieures à 23 ans. Le fait que les évolutions que connaît l'âge d'entrée en union soient moins rapides que chez les femmes se traduit par un resserrement des écarts d'âge au mariage. On trouvera dans Hertrich (2007) une présentation similaire à partir des âges médians ainsi qu'une discussion plus complète de ces tendances [22].

20. En retenant 45 comme âge limite supérieur du premier mariage, puisque les données MICS et EDS se limitent aux femmes de 45-49 ans.

21. Centrafrique, Guinée, Malawi, Mali, Mozambique, Niger, Tchad.

22. Lesthaeghe et al. (1989) présentent également des cartes au niveau régional pour la période 1960-1980.

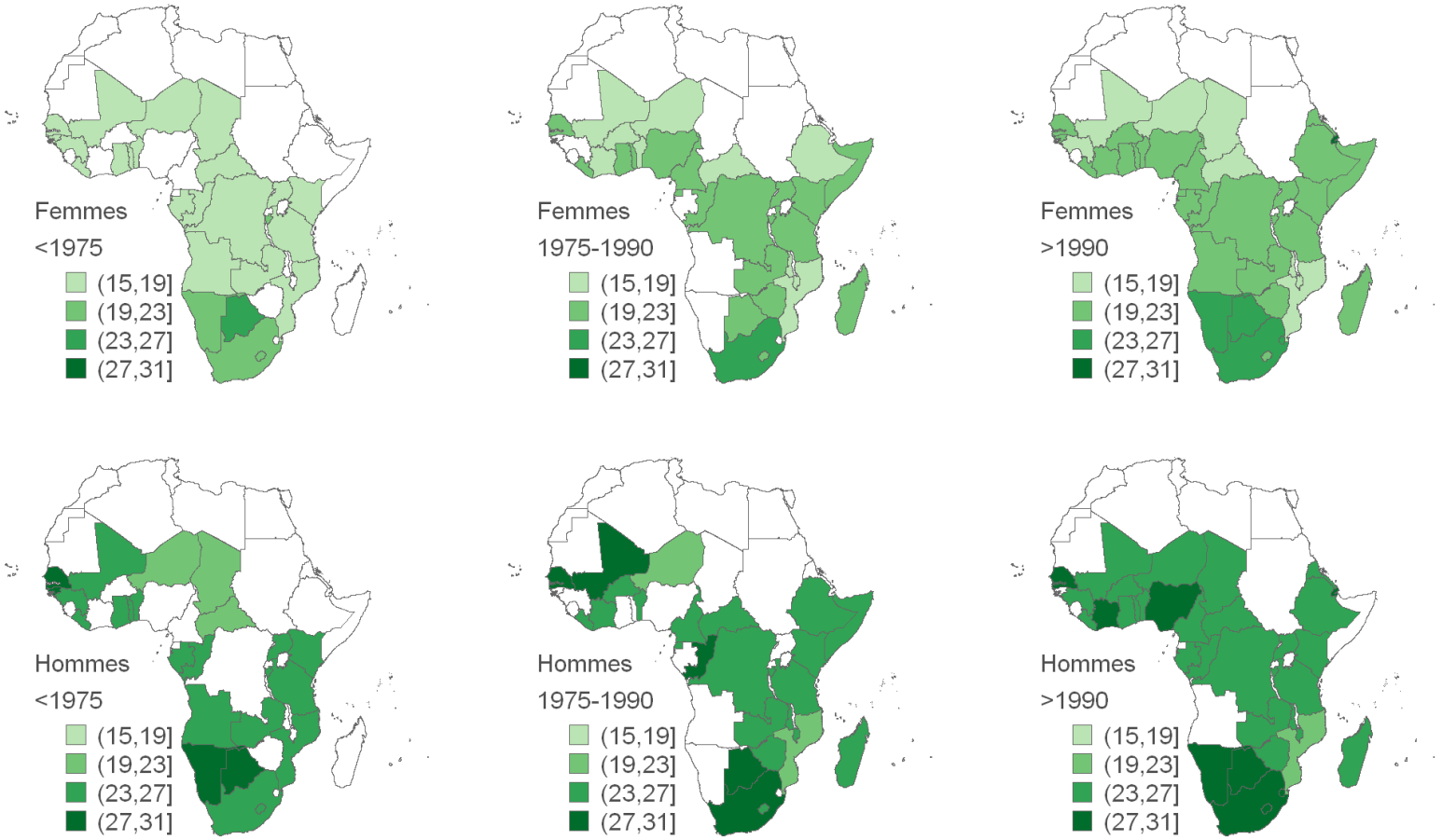

FIGURE A.13 – *Valeurs de l'âge moyen au premier mariage (féminin et masculin) issu de différentes sources, avant 1975, entre 1975 et 1990 et depuis 1990*

En plus de ces valeurs de l'âge moyen, il nous faut, pour reconstruire des schémas de primo-nuptialité, des informations sur la **variation des âges au mariage**. En fait, il est plus aisé de re-paramétriser le modèle de Coale et McNeil (1972), afin d'estimer l'écart-type de l'âge au premier mariage (σ), plutôt que κ ou a_0. Pour ce faire, Rodriguez et Trussell (1980) proposent un nouveau standard, qui a une moyenne nulle et une variance de 1. La fonction de densité de ce nouveau standard vaut :

$$G_0(x) = 1.2813 \times exp\left[-1.145(x + 0.805) - exp\left\{-1.896(x + 0.805)\right\}\right] \tag{A.7}$$

Rodriguez et Trussell (1980) montrent aussi que la fonction de répartition de ce standard peut être exprimée par le biais d'une fonction gamma [23], ce qui simplifie considérablement les calculs. Cette fonction peut en effet être utilisée pour estimer, par maximisation de la vraisemblance, les valeurs de l'écart-type σ ainsi que de la proportion de célibataires définitifs C, en recourant à des procédures d'optimisation numérique [24]. Combinés aux valeurs de l'âge moyen au premier mariage, celles-ci permettent ensuite de reconstituer des schémas par âge [25]. Les proportions observées et ajustées à partir des enquêtes EDS et MICS sont présentées dans le graphe A.14. L'ajustement est bon à la fois pour les pays où le mariage est *précoce* et *concentré aux premiers âges* (par exemple le Burundi en 1987 ou le Tchad en 1997), et pour les pays où il est plus *tardif* et *davantage réparti entre les âges* (p.ex. la Namibie en 1992 ou l'Afrique du Sud en 1998).

Le graphe A.15 procure un aperçu de la *variation* des âges au premier mariage, sous la forme du coefficient de variation σ/S. On observe globalement une hausse de la dispersion des âges au premier mariage chez les femmes, qui est moins nette chez les hommes. Chez les femmes, l'Afrique australe, où l'entrée en union paraît sensiblement plus tardive, sort à nouveau du lot et se caractérise également par une plus grande dispersion des âges au mariage. On retrouve chez les hommes des valeurs plus élevées en Afrique australe (surtout en Namibie et au Botswana) ainsi que sur une région qui part du Sénégal et rejoint le Congo RDC via le Golfe de Guinée.

23. La fonction s'exprime comme ceci :

$$G_0(x) = 1 - I\left[exp(-1.896(x + 0.805)); -0.396\right] \tag{A.8}$$

Avec un paramètre a/λ valant 0.604.

24. Voir Rodriguez et Trussell (1980) pour l'expression du log de vraisemblance. La méthode procure également des valeurs de S, très proches de celles calculées par la méthode d'Hajnal (1953).

25. Nous utilisons la procédure `optim` de R pour procéder à l'optimisation numérique. Cela suppose que soient introduites des valeurs initiales ; la valeur de l'âge moyen au premier mariage calculé par la méthode d'Hajnal (1953) est utilisée comme valeur de départ pour S, une valeur de 5 est fixée pour σ et une valeur de 0.9 est fixée pour C, la proportion de femmes qui vont se marier à terme. Ces valeurs de départ n'ont pas d'impact sur les estimations obtenues.

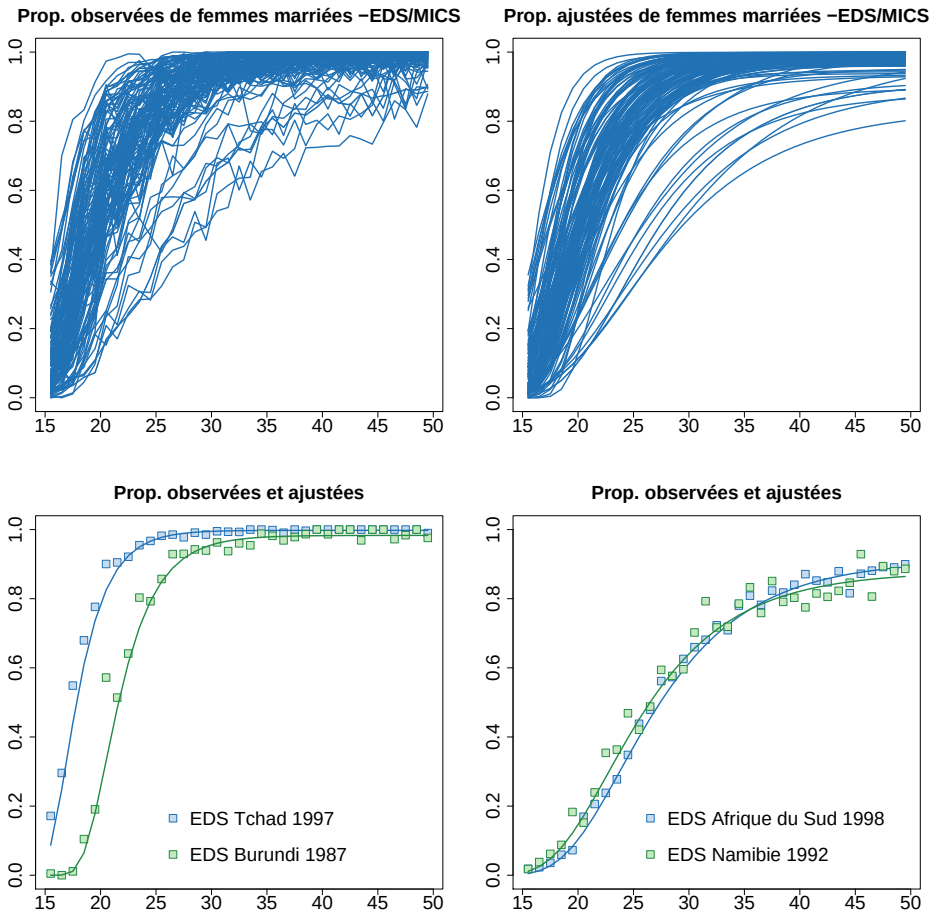

FIGURE A.14 – *Proportions observés et ajustées de femmes mariées, enquêtes EDS et MICS (126 enquêtes pour 42 pays.). Exemples du Tchad (1997), Burundi (1987), Afrique du Sud (1998) et Namibie (1992) (Panel du bas)*

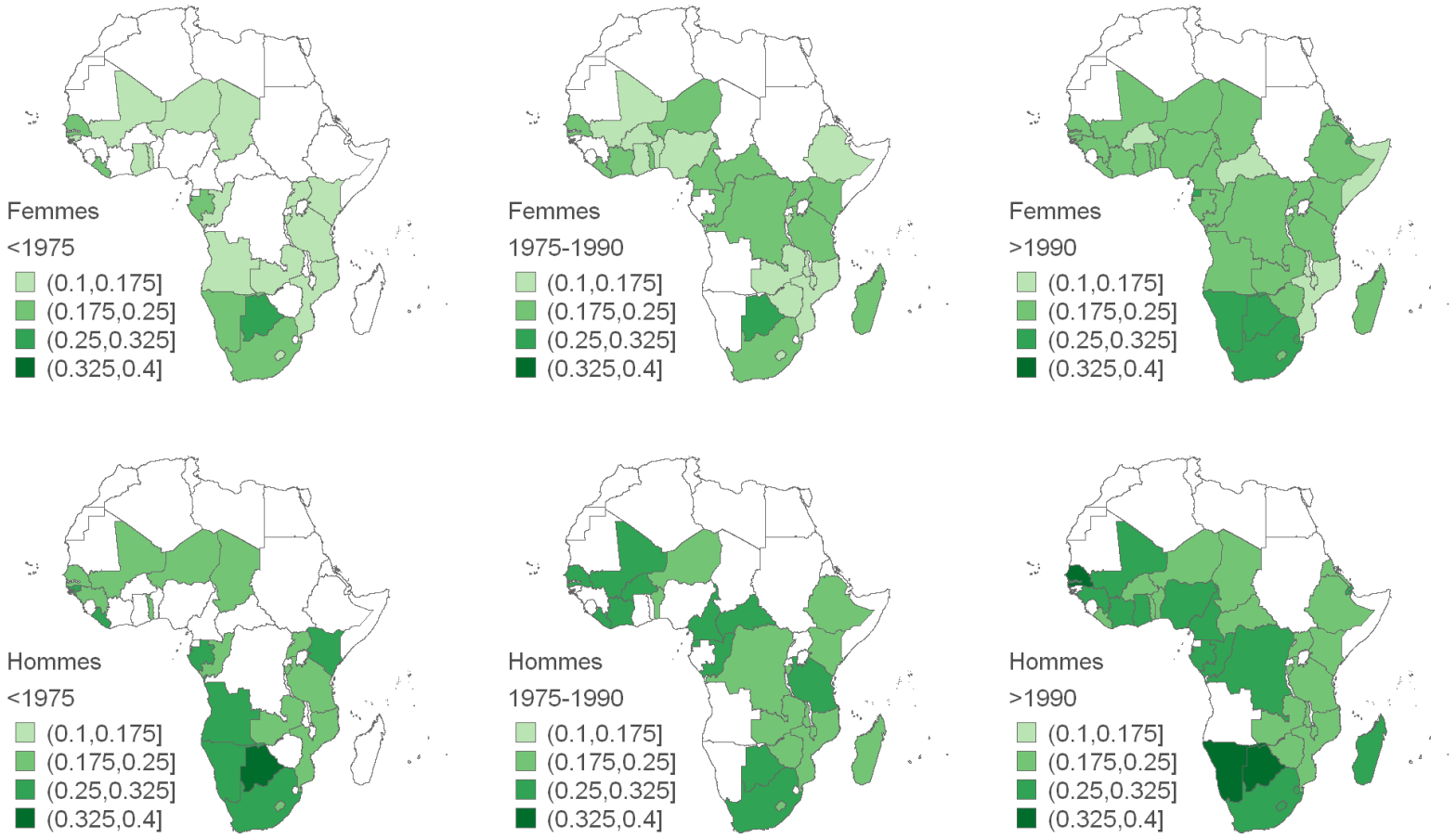

FIGURE A.15 – *Valeurs de σ/SMAFM estimés par la méthode de Rodriguez et Trussell (1980) à partir des proportions de célibataires par âge, avant 1975, entre 1975 et 1990 et depuis 1990*

Nous disposons donc maintenant d'une série d'estimations de S, de σ et de C. Ces estimations sont soit obtenues directement des rapports d'enquêtes et de recensements, soit calculées par la méthode de Rodriguez et Trussell (1980) à partir des distributions de célibataires par âge. Reconstruire des schémas par âge pour les simulations à partir de ces valeurs suppose de disposer d'estimations qui varient *dans le temps* et *par pays* depuis le début des simulations jusqu'à 2010. Ceci nécessite de recourir à une certaine forme d'interpolation. Lisser directement les données nationales pose problème, en raison de la rareté des données pour certains d'entre eux. Nous recourons donc plutôt à un modèle linéaire multi-niveaux, où la courbe de la pente (c-à-d. la tendance) ainsi que la constante (c-à-d. le niveau de départ) sont autorisés à varier d'un pays à l'autre. Cette approche présente l'avantage de pouvoir *"emprunter" de l'information aux différents pays pour ceux où seules quelques estimations sont disponibles*.

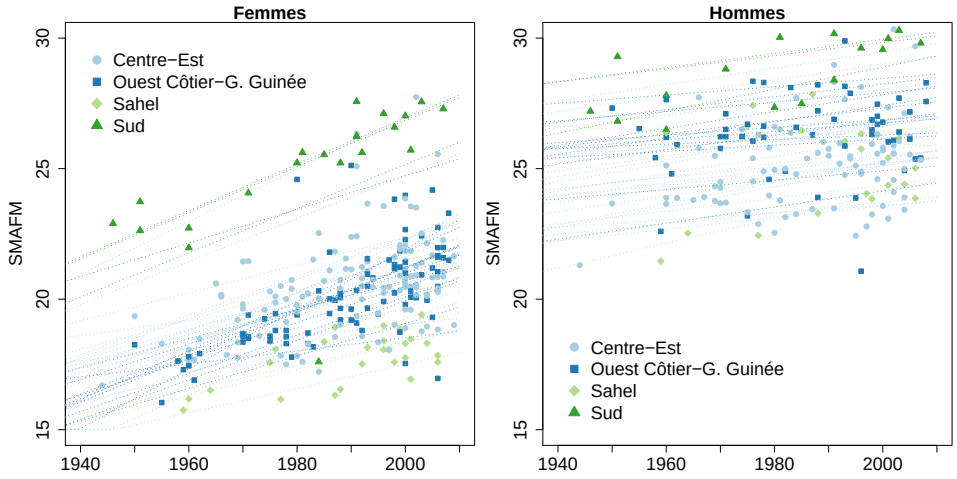

FIGURE A.16 – *Valeurs de SMAFM observées (les points) et estimées (les droites) par le modèle linéaire multi-niveaux (41 pays), entre 1944 et 2010*

Le graphe A.16 présente les valeurs observées de l'âge moyen $SMAFM$ et les résultats de l'ajustement [26]. On y retrouve les grandes tendances dégagées précédemment ; un recul de l'âge d'entrée en union chez les femmes entamé relativement tôt en Afrique australe [27], une évolution plus lente et plus diversifiée dans les pays de l'Ouest Côtier et du Golfe de Guinée [28] ainsi que dans la région Centre-Est [29]. Les rythmes de changements sont par contre moins rapides dans la région sahélienne [30]. Comme on l'a indiqué précédemment, les tendances sont moins évidentes chez les hommes, avec une stabilité dans la région Centre-Est ainsi que dans l'Ouest Côtier et le Golfe de Guinée, une hausse plus prononcée en Afrique australe, et enfin, des tendances contrastées, voire parfois diver-

26. On notera que les valeurs ajustées de $SMAFM$ sont limitées à 15 ans au mimimum.
27. Botswana, Namibie, Afrique du Sud.
28. Benin, Cameroun, République Centrafricaine, Côte d'Ivoire, Gambie, Ghana, Guinée, Guinée-Bissau, Liberia , Nigeria , Sénégal, Togo, Cap Vert.
29. Angola, Burundi, Comores, Congo, RDC, Guinée équatoriale ,Érythrée, Éthiopie, Gabon, Kenya, Madagascar, Malawi, Mozambique, Somalie, Ouganda, Tanzanie, Zambie, Zimbabwe, Swaziland, Lesotho, Djibouti.
30. Burkina Faso, Mali, Niger, Tchad.

gentes, dans les pays du Sahel [31].

Le résultat de l'ajustement du même modèle multi-niveau linéaire sont présentés pour la *proportion de mariés à 50 ans* dans le graphe A.17. Tandis que cette proportion décline légèrement chez les femmes dans la zone Centre-Est, elle baisse fortement au Botswana, en Afrique du Sud, en Namibie ainsi qu'au Cap Vert. Elle semble par contre connaître une hausse chez les hommes.

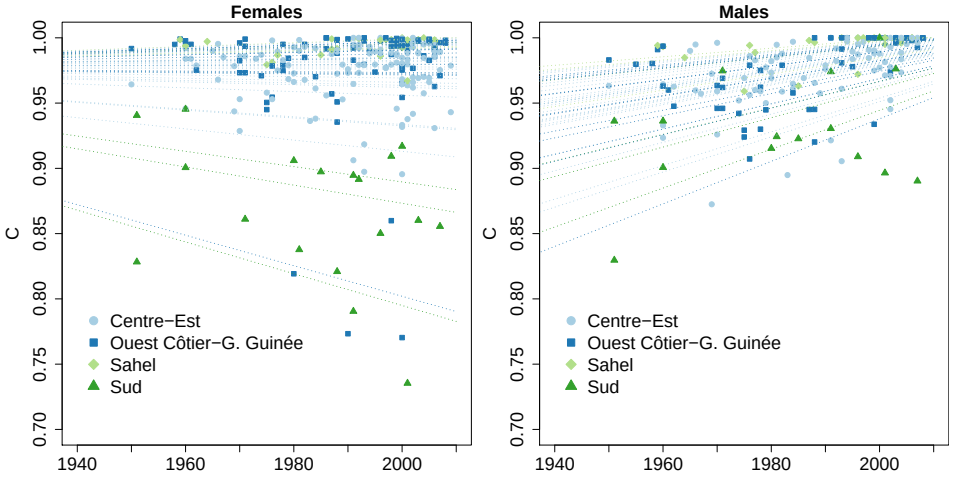

FIGURE A.17 – *Valeurs de C (% de mariés à 50 ans) observées et estimées par le modèle linéaire multi-niveaux*

A partir de ces estimations interpolées de $SMAFM$, σ et C, nous reconstruisons des schémas de primo-nuptialité pour chaque pays et chaque segment, et nous construisons les tables de "survie" correspondantes à partir de la courbe de survivants au premier mariage [32].

A ce stade, l'intégration de la nuptialité dans les microsimulations reste assez sommaire. En raison du manque de biographies maritales complètes, il est difficile de disposer d'estimations sur la rapidité des divorces et les taux de re-mariage, à l'exception de quelques pays ou de quelques zones géographiques restreintes (Reniers 2003). *Nous supposons dans les simulations que le mariage se maintient jusqu'au décès de l'un des deux partenaires, et que les taux de re-mariage sont nuls. Nous évitons également les nombreux problèmes associés à la modélisation de la polygamie*, ce qui constitue une simplification qui devrait être corrigée par la suite. Il s'agit là d'une piste de développements futurs pour un modèle de micro-simulations plus abouti.

Qu'obtient-on au niveau des populations simulées ? Le graphe A.18 présente les proportions de célibataires par âge *estimés à partir du modèle multi-niveaux* en 2009 pour la Namibie et le Lesotho. Ces proportions sont

31. L'entrée en union semble ainsi se faire de plus en plus précocement au Mali et de plus en plus tardivement au Niger, par exemple.

32. Nous utilisons donc des données de cohortes pour construire des schémas de primo-nuptialité "du moment". Nous faisons également l'hypothèse que les risques de décès ne varient pas avec la situation matrimoniale. Nous posons en outre ${}_na_x = n/2$, ce qui signifie que les mariages sont répartis uniformément entre les âges exacts.

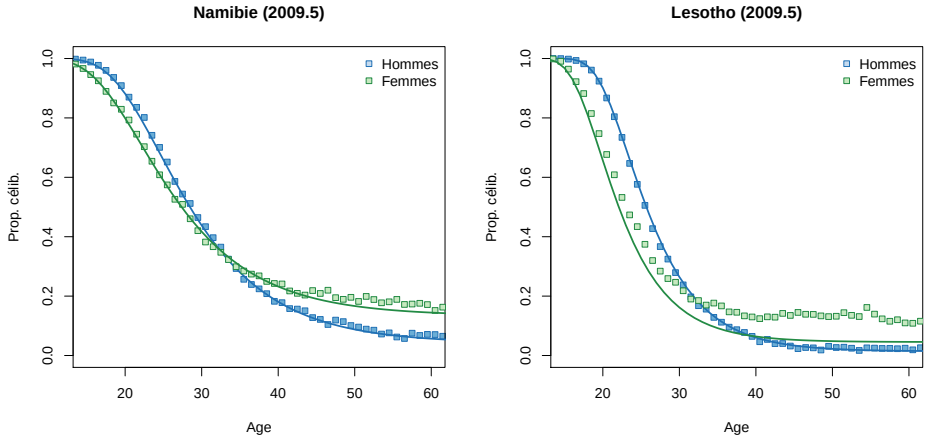

FIGURE A.18 – *Proportions de célibataires (par âge) observées dans les populations simulées du Lesotho et de la Namibie (Année 2009)*

représentées par les courbes. Elles sont comparées aux proportions *observées à partir des populations issues des simulations* pour ces deux pays (représentées par les carrés). Ces proportions observées dans les simulations devraient idéalement cadrer avec celles issues du modèle multi-niveaux, puisque c'est à partir de ce modèle que sont conçus les taux de primo-nuptialités. Or qu'observe-t-on ? En Namibie, les taux de nuptialité introduits permettent bien de retomber sur les proportions de célibataires estimées à partir des enquêtes et des recensements. Par contre, au Lesotho, les proportions de *femmes célibataires* sont trop élevées dans les simulations.

Comment expliquer ces écarts ? Rappelons que lorsque les taux de nuptialité indiquent qu'un individu doit se marier à tel âge pour telle année, SOCSIM ne programme *pas directement* de mariage pour cet individu. Il lance plutôt une recherche de partenaires qui pourraient lui correspondre. Si personne n'est disponible pour épouser cet individu, aucun mariage n'aura lieu. Au terme des simulations, les taux de nuptialité que nous pouvons recalculer à partir des populations finales ne correspondront donc pas toujours à ceux qui ont été introduits avant de lancer ces simulations. *Tout dépend de la vitesse avec laquelle les individus en recherche de partenaires parviennent à trouver l'"âme sœur"*. C'est pourquoi les proportions de célibataires observées sur le graphe A.18 ne cadrent pas avec ce que nous attendons sur base des taux introduits.

Revenons un instant sur le cas du Lesotho. Les femmes s'y marient beaucoup plus tôt que les hommes, sur globalement toute la période des simulations. Les hommes arrivent donc sur un marché matrimonial dans lequel la *demande de partenaires masculins* est élevée. En effet, de nombreux mariages ont été programmés pour des femmes plus jeunes qui n'ont jusque là pas trouvé de partenaires. Dès qu'un homme arrive sur le marché matrimonial (parce que les taux introduits en ont décidé ainsi à son âge), il trouve immédiatement l'un des partenaires féminins qui l'attendaient. Dans cette configuration, les hommes sont directement mariés lorsqu'une recherche de partenaire est entamée par SOCSIM, tandis que les femmes doivent patienter plus longtemps. C'est ce qui explique que les simulations ne parviennent pas à reproduire les proportions de

célibataires par âge estimées à partir des enquêtes et des recensements.

Ceci signifie également qu'un nombre trop important de naissances surviennent hors mariage. Or seuls les enfants nés au sein d'un mariage ont un père identifié dans les simulations.

Nous n'avons pas trouvé de solution satisfaisante pour régler ce problème. La meilleure option serait sans doute d'intégrer dans les simulations des probabilités d'entrer en *union polygame*, ou bien d'intégrer des *taux de divorce et de remariages*. Ceci permettrait d'augmenter la disponibilité des hommes, et les femmes trouveraient plus facilement des partenaires. Mais le calcul des taux de nuptialité à introduire dans les simulations en est rendu plus complexe.

Nous en sommes donc restés à la simulation des premiers mariages. Comme une proportion élevée d'enfants naissent hors mariage, nous leur allouons des pères au terme des simulations. Les pères potentiels sont choisis en fonction des écarts d'âge avec la mère[33]. Nous retenons une distribution des écarts d'âge entre partenaires sexuels, qui approxime une distribution log-logistique (graphe A.19). Cette dernière est commune à tous les groupes d'âge des femmes, et issue de l'ajustement des données d'une enquête menée dans le Zimbabwe rural entre 1998 et 2000 (Hallett et al. 2007).

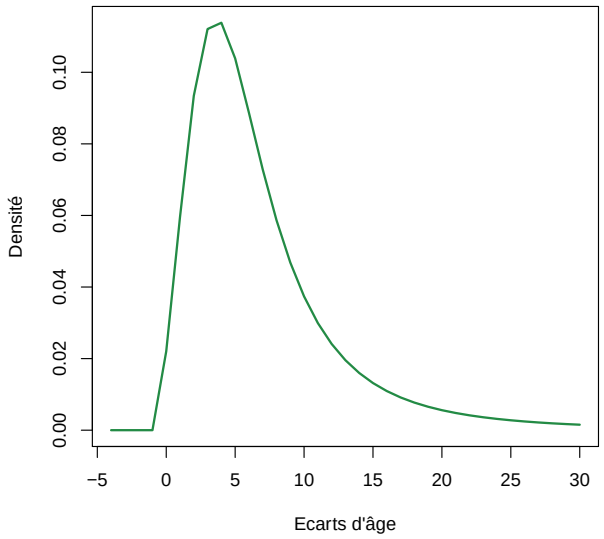

FIGURE A.19 – *Distribution des écarts d'âge entre partenaires sexuels retenus pour allouer un père aux enfants nés hors mariage - Sc : approximé depuis Hallett et al. (2007)*

En définitive, cette section relative à la nuptialité montre qu'il est possible de combiner de nombreuses sources de données pour reconstituer les tendances de la primo-nuptialité. Le recours à un modèle multi-niveaux permet de dégager des estimations qui varient dans le temps pour tous les pays considérés. Les premiers mariages peuvent donc être simulés. Cela étant, comme le mariage est un événement qui concerne deux individus,

33. Seuls ceux qui étaient encore en vie au moins à 9 mois avant la naissance sont considérés comme des pères potentiels.

il peut y avoir un décalage dans les simulations entre les taux introduits pour chaque sexe et les taux recalculés. C'est ce que nous avons observé ici. Les écarts d'âge importants entre les sexes observés en Afrique en matière d'âge au mariage rendent la simulation des divorces et de la polygamie nécessaires, mais ceci sortait du cadre de cette recherche.

A.2.8 Notes sur les distributions "phase-type"

Toute distribution de probabilité prenant des valeurs positives peut être approximée par une distribution appelée "phase-type" (Aalen 1995). Il s'agit d'une distribution qui résulte de la combinaison de différentes transitions entre étapes, ou phases. La probabilité d'opérer chaque transition dépend uniquement de l'état courant du processus (c'est-à-dire de l'étape atteinte, et pas du temps passé dans cette étape ou dans les étapes antérieures). Cela signifie qu'au sein de chaque étape (ou sous-groupe), la probabilité d'opérer une transition est constante et exponentielle *en fonction du temps*.

Ces distributions "phase-type", qui résultent donc de la combinaison d'exponentielles, présentent une grande flexibilité, car les étapes peuvent être combinées de nombreuses façons. Le plus simple consiste à faire passer les individus de façon séquentielle d'une étape à l'autre en vertu d'une probabilité identique dans toutes les étapes, en prévoyant que chaque étape n'est visitée qu'une fois. La dernière étape de ce processus est absorbante ; elle correspond dans notre cas au décès dû au sida. La distribution des durées de survie correspond alors à la *durée écoulée dans la séquence des étapes avant d'atteindre cette étape absorbante*.

Des modèles plus complexes peuvent également être utilisés, notamment quand :

- les probabilités de transition d'une phase à l'autre sont spécifiques à chaque phase,
- plusieurs phases donnent accès directement à l'état absorbant,
- des effets de feedback vers les étapes antérieures sont introduits.

On peut en trouver des exemples dans la littérature consacrée au VIH-sida. Par exemple, Longini et al. (1989) analysent la progression du VIH avec des taux de transition variables, ce qui permet, à partir d'une cohorte de séropositifs américains, d'estimer la durée de survie moyenne passé dans quatre étapes séquentielles :

- infectés mais séronégatifs (soit en moyenne 2.2 mois),
- séropositifs mais asymptotiques (52.6 mois),
- présentant des symptômes pré-sida (62.9 mois),
- ayant atteint le sida clinique (23.6 mois).

Pour d'autres exemples de distributions "phase-type" adaptées à l'analyse de survie, on se réfèrera à Aalen (1995).

Quel est l'intérêt de ces distributions "phase-type" pour notre modélisation de la survie des personnes infectées dans SOCSIM ? Simplement, nous pouvons décomposer la survie en plusieurs étapes, en prévoyant que la durée passée dans la dernière étape renvoie à la survie au sein

du stade du sida clinique. Comme nous modélisons la survie au sein du stade du sida clinique à partir d'une distribution exponentielle, nous pouvons ensuite fixer le paramètre de la dernière exponentielle et chercher à estimer les différents paramètres des exponentielles qui précèdent. En d'autres termes, nous supposons que :

- la survie des personnes infectées est un processus par étapes,
- le passage d'une étape à l'autre se fait à un taux constant (mais variable d'une étape à l'autre),
- le taux de la dernière transition est connu.

Nous cherchons ensuite les taux de transitions entre les différentes étapes qui, combinés entre eux, ajustent au mieux les courbes de survie de l'infection au décès.

Nous utilisons cette méthode notamment pour approcher au plus près des distributions de Weibull détaillées dans le tableau 2.1 (repris en page 90). Pratiquement, nous avons approximés numériquement les paramètres des exponentielles. Les taux de transition sont calculés de façon à minimiser la somme des carrés des distances entre la fonction de répartition de la distribution de Weibull et la fonction de répartition de la distribution "phase-type" qui résulte de la combinaison des différentes exponentielles [34]. Les taux de transition sont variables d'une étape à l'autre.

Nous avons retenu 4 étapes, car c'est l'option qui procure le meilleur ajustement. Chaque étape n'est visitée qu'une fois et il n'y a pas d'allers-retours entre les étapes [35].

Le graphe 2.18 (repris en page 92) et la comparaison des médianes dans le tableau 2.1 (p.90) permettent de juger de la qualité de l'ajustement.

34. Nous utilisons la procédure `optim` de R.

35. Il serait possible d'obtenir un meilleur ajustement en prévoyant des allers-retours, ou en autorisant les personnes infectées à sauter une étape, mais ceci complique un peu l'estimation des paramètres.

A.3 Annexes au chapitre 3

A.3.1 Calcul des probabilités ${}_np_{25}$ à partir de la survie des mères : 6 recensements est-africains

Note : Le recensement rwandais de 2002 n'a interrogé que les moins de 25 sur la survie des mères.

	Année	Age	${}_5S_{n-5}$	M	${}_np_{25}$	${}_{45}q_{15}$
10	1991	5-9	0.967	28.9	0.961	0.255
15	1991	10-14	0.947	28.9	0.946	0.230
20	1991	15-19	0.915	28.9	0.919	0.236
25	1991	20-24	0.880	28.9	0.892	0.227
30	1991	25-29	0.828	28.9	0.850	0.227
35	1991	30-34	0.769	28.9	0.805	0.214
40	1991	35-39	0.675	28.9	0.727	0.215
45	1991	40-44	0.547	28.9	0.606	0.221
50	1991	45-49	0.398	28.9	0.447	0.225
10	2002	5-9	0.935	28.5	0.920	0.432
15	2002	10-14	0.865	28.5	0.854	0.484
20	2002	15-19	0.823	28.5	0.821	0.443
25	2002	20-24	0.786	28.5	0.793	0.394

Table A.8 – *Application de la méthode des orphelins aux recensements rwandais de 1991 et 2002 (mortalité féminine)*

	Année	Age	${}_5S_{n-5}$	M	${}_nP_{25}$	${}_{45}q_{15}$
10	1989	5-9	0.986	26.4	0.982	0.134
15	1989	10-14	0.976	26.4	0.973	0.126
20	1989	15-19	0.957	26.4	0.955	0.141
25	1989	20-24	0.929	26.4	0.928	0.157
30	1989	25-29	0.890	26.4	0.893	0.167
35	1989	30-34	0.836	26.4	0.843	0.173
40	1989	35-39	0.765	26.4	0.779	0.170
45	1989	40-44	0.649	26.4	0.662	0.181
50	1989	45-49	0.550	26.4	0.556	0.156
10	1999	5-9	0.974	25.0	0.965	0.236
15	1999	10-14	0.955	25.0	0.946	0.229
20	1999	15-19	0.935	25.0	0.926	0.218
25	1999	20-24	0.911	25.0	0.901	0.209
30	1999	25-29	0.883	25.0	0.872	0.196
35	1999	30-34	0.835	25.0	0.823	0.194
40	1999	35-39	0.780	25.0	0.769	0.179
45	1999	40-44	0.685	25.0	0.669	0.177
50	1999	45-49	0.581	25.0	0.552	0.158

TABLE A.9 – *Application de la méthode des orphelins aux recensements kenyans de 1989 et 1999 (mortalité féminine)*

	Année	Age	${}_5S_{n-5}$	M	${}_nP_{25}$	${}_{45}q_{15}$
10	1988	5-9	0.980	26.6	0.975	0.178
15	1988	10-14	0.965	26.6	0.961	0.174
20	1988	15-19	0.939	26.6	0.936	0.192
25	1988	20-24	0.902	26.6	0.901	0.210
30	1988	25-29	0.855	26.6	0.858	0.217
35	1988	30-34	0.780	26.6	0.788	0.234
40	1988	35-39	0.694	26.6	0.705	0.234
45	1988	40-44	0.563	26.6	0.569	0.250
50	1988	45-49	0.469	26.6	0.465	0.212
10	2002	5-9	0.974	26.6	0.967	0.221
15	2002	10-14	0.948	26.6	0.942	0.242
20	2002	15-19	0.914	26.6	0.909	0.259
25	2002	20-24	0.877	26.6	0.875	0.257
30	2002	25-29	0.842	26.6	0.844	0.236
35	2002	30-34	0.795	26.6	0.802	0.217
40	2002	35-39	0.736	26.6	0.750	0.195
45	2002	40-44	0.647	26.6	0.664	0.180
50	2002	45-49	0.549	26.6	0.558	0.155

TABLE A.10 – *Application de la méthode des orphelins aux recensements tanzaniens de 1988 et 2002 (mortalité féminine)*

A.3.2 Coefficients utilisés pour les méthodes indirectes

Facteurs de pondération de Brass & Hill (1973) - mortalité féminine

TABLE A.11 – *Facteurs de pondération $W(n)$ pour estimer la mortalité féminine*

n	22	23	24	25	26	27	28	29	30
10	0,420	0,470	0,517	0,557	0,596	0,634	0,674	0,717	0,758
15	0,418	0,489	0,556	0,618	0,678	0,738	0,800	0,863	0,924
20	0,404	0,500	0,590	0,673	0,756	0,838	0,921	1,004	1,085
25	0,366	0,485	0,598	0,704	0,809	0,913	1,016	1,118	1,218
30	0,303	0,445	0,580	0,708	0,834	0,957	1,080	1,203	1,323
35	0,241	0,401	0,554	0,701	0,844	0,986	1,128	1,270	1,412
40	0,125	0,299	0,467	0,630	0,791	0,950	1,111	1,274	1,442
45	0,007	0,186	0,361	0,535	0,708	0,884	1,063	1,250	1,447
50	-0,190	-0,017	0,158	0,334	0,514	0,699	0,890	1,095	1,318
55	-0,368	-0,220	-0,059	0,101	0,270	0,456	0,645	0,856	1,083
60	-0,466	-0,352	-0,217	-0,084	0,053	0,220	0,378	0,579	0,800

Coefficients de régression du Manuel X (1983)

n	β_0	β_1	β_2
20	-0.1798	0.00476	1.0505
25	-0.2267	0.00737	1.0291
30	-0.3108	0.01072	1.0287
35	-0.4259	0.01473	1.0473
40	-0.5566	0.01903	1.0818
45	-0.6676	0.02256	1.1228
50	-0.6981	0.02344	1.1454

TABLE A.12 – *Coefficients utilisés pour convertir les proportions de mères survivantes en probabilités de survie - UN (1983b)*

Facteurs de pondération de Brass & Hill (1973) - mortalité masculine

Ces facteurs sont destinés à estimer ${}_{n+2.5}p_{32.5}$ et ${}_{n+2.5}p_{37.5}$.

Si l'âge paternel à la naissance des enfants qui est calculé est inférieur à 36 ans, 32.5 est l'âge retenu comme âge de départ des probabilités de survie qui seront estimées. Par contre, si cet âge paternel à la naissance est supérieur à 36, alors 37.5 ans est retenu comme âge de départ.

Par ailleurs, les calculs se font pour $n + 2.5$ puisque la durée d'exposition masculine doit être rallongée (en raison du délai entre la conception et la naissance). Il faudrait théoriquement décaler les classes d'âge des enquêtés pour pouvoir travailler avec n+0.75. Mais pour des raisons pratiques, Brass et Hill (1973) prévoient des facteurs de pondération pour n+2.5.

TABLE A.13 – *Facteurs de pondération $W(n)$ pour estimer la mortalité masculine*

[32.5]					M				
N	28	29	30	31	32	33	34	35	36
10	0,192	0,258	0,322	0,388	0,455	0,521	0,587	0,650	0,714
15	0,151	0,243	0,336	0,429	0,522	0,613	0,702	0,790	0,877
20	0,043	0,166	0,287	0,406	0,523	0,638	0,750	0,861	0,969
25	-0,093	0,051	0,194	0,335	0,474	0,611	0,744	0,877	1,007
30	-0,327	-0,161	0,001	0,162	0,319	0,475	0,627	0,779	0,931
35	-0,640	-0,408	-0,211	-0,047	0,109	0,269	0,438	0,610	0,782
40	-0,856	-0,714	-0,554	-0,379	-0,203	-0,034	0,133	0,303	0,480
45	-1,120	-0,963	-0,806	-0,651	-0,495	-0,340	-0,183	-0,024	0,141
50	-1,162	-1,030	-0,903	-0,776	-0,651	-0,524	-0,396	-0,264	-0,128
55	-1,040	-0,943	-0,850	-0,758	-0,667	-0,576	-0,486	-0,397	-0,304

[32.5]					M				
N	36	37	38	39	40	41	42	43	44
10	0,384	0,460	0,537	0,613	0,687	0,384	0,384	0,384	0,384
15	0,378	0,484	0,588	0,690	0,790	0,378	0,378	0,378	0,378
20	0,324	0,455	0,582	0,708	0,833	0,324	0,324	0,324	0,324
25	0,164	0,315	0,465	0,613	0,759	0,164	0,164	0,164	0,164
30	-0,043	0,122	0,286	0,450	0,614	-0,043	-0,043	-0,043	-0,043
35	-0,359	-0,183	-0,015	0,152	0,321	-0,359	-0,359	-0,359	-0,359
40	-0,624	-0,473	-0,316	-0,157	0,003	-0,624	-0,624	-0,624	-0,624
45	-0,757	-0,631	-0,503	-0,372	-0,237	-0,757	-0,757	-0,757	-0,757
50	-0,742	-0,650	-0,559	-0,471	-0,742	-0,742	-0,742	-0,742	-0,742
55	-0,559	-0,541	-0,485	-0,425	-0,559	-0,559	-0,559	-0,559	-0,559

Coefficients de régression de Timaeus (1991b)

n	β_0	β_1	β_2	β_3	R^2	CV
25	-0.0554	0.00757	0.0239	0.8080	0.984	0.0063
30	-0.7539	0.01558	0.6452	0.6498	0.990	0.0114
35	-1.0809	0.02273	0.9289	0.4807	0.991	0.0187
40	-1.1726	0.02647	0.9381	0.4372	0.979	0.0448

TABLE A.14 – *Coefficients utilisés pour convertir les proportions de non-orphelins de père (parmi les non-orphelins de père à 20 ans) en probabilités de survie masculines - Timaeus (1991b)*

Coefficients utilisés pour dater les estimations

M + n	Z	M + n	Z	M + n	Z	M + n	Z	M + n	Z
26	0.090	36	0.092	46	0.149	56	0.274	66	0.452
27	0.090	37	0.093	47	0.160	57	0.289	67	0.473
28	0.090	38	0.095	48	0.171	58	0.305	68	0.495
29	0.090	39	0.099	49	0.182	59	0.321	69	0.518
30	0.090	40	0.104	50	0.193	60	0.338	70	0.542
31	0.090	41	0.109	51	0.205	61	0.356	71	0.568
32	0.090	42	0.115	52	0.218	62	0.374	72	0.595
33	0.090	43	0.122	53	0.231	63	0.392	73	0.622
34	0.090	44	0.130	54	0.245	64	0.411	74	0.650
35	0.091	45	0.139	55	0.259	65	0.431	75	0.678

TABLE A.15 – *Coefficients utilisés pour dater les probabilités de survie obtenues à partir de la survie des parents - Brass et Bamgboye (1981)*

	β_0	β_1	β_2	β_3	β_4	β_5	R^2	CV
5	-339.7	13.53	-0.25	2.27	355.27	-13.99	0.986	0.0117
10	-167	6.84	0.11	2.69	181.69	-7.2	0.987	0.0078
15	-105.4	4.49	0.29	3.07	121.19	-4.81	0.985	0.0077
20	-72.2	3.28	0.51	3.42	90.95	-3.65	0.981	0.0078
25	-46.9	2.33	0.75	3.61	68.99	-2.76	0.976	0.0074
30	-18.3	1.26	0.52	2.43	44.33	-1.77	0.976	0.0044
35	14.3	0.11	0.66	1.81	3.93	-0.28	0.834	0.0061
40	52	-1.22	0.77	0.54	-60.05	1.98	0.860	0.0185
45	94.9	-2.71	0.25	-1.66	-167.51	5.69	0.852	0.0440

TABLE A.16 – *Coefficients utilisés pour dater les probabilités de survie féminines obtenues à partir de la survie des mères (Coefficients calculés sur base du Standard général des Nations Unies) - Palloni et al. (1984)*

A.3.3 Modèle polynomial de Brass

Le modèle polynomial de fécondité développé par Brass est un modèle très simple qui s'exprime sous la forme suivante :

$$f(a) = c(a - s)(s + w - a)^2$$

Le paramètre c renvoie au niveau de fécondité. Ses hausses se répercutent de façon multiplicative sur les taux à chaque âge et il est directement lié à la valeur de l'ISF. Le paramètre s correspond à l'âge au début de la période reproductive. Enfin, le paramètre w renvoie à la durée de la période de reproduction (fixée à 33). Un changement de s se traduit par une translation de la structure par âge de la fécondité, sans la modifier. Il y a donc une relation linéaire entre s et l'âge moyen à la maternité (Retherford 1979) : $M = s + 13.2$. Pour les hommes, la fonction est légèrement modifiée pour tenir compte de la plus grande répartition de la fécondité aux âges élevés (w est fixé à 60) :

$$f(a) = c(a - s)(s + w - a)^3$$

Les deux schémas, masculins et féminins, sont présentés dans le graphe A.20.

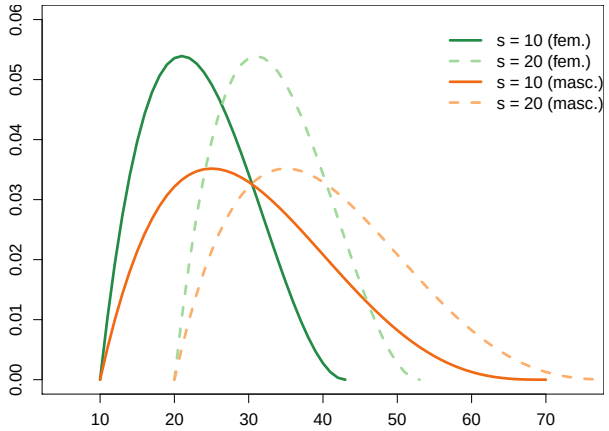

FIGURE A.20 – *Modèle polynomial de Brass pour la fécondité avec deux valeurs de s = 10 et 20*

A.3.4 Notes sur l'estimation indirecte de l'âge moyen à la naissance des enfants des pères

Nous avons calculé, à partir des enquêtes EDS, les valeurs de M_{σ} en comparant les deux options indirectes détaillées en page 131. Ces calculs révèlent des écarts importants. Tout d'abord, l'ajustement de l'âge médian des hommes mariés pour tenir compte de la polygamie (en pondérant simplement le nombre d'hommes par âge par leur nombre moyen de femmes) a pour effet de relever l'âge moyen des pères de 1 an environ (Figure A.21.A). Cette correction est surtout importante en Afrique de l'Ouest (Benin, Burkina Faso, Guinée). Par ailleurs, on note des différences conséquentes entre les estimations basées sur les *âges médians* et celles basées sur les *âges moyens au premier mariage* (Figure A.21.B). Sur l'ensemble des 66 enquêtes considérées ici, les deux estimations sont séparées de 3 ans en moyenne. Dans certaines enquêtes (Benin 1996, Sénégal 1993-97, Burkina Faso 1993), elles sont même séparées de plus de 9 ans. Quand un calcul direct n'est pas possible, le choix de la méthode d'estimation indirecte n'est donc pas neutre, il est susceptible de se répercuter sur les valeurs des probabilités de survies calculées.

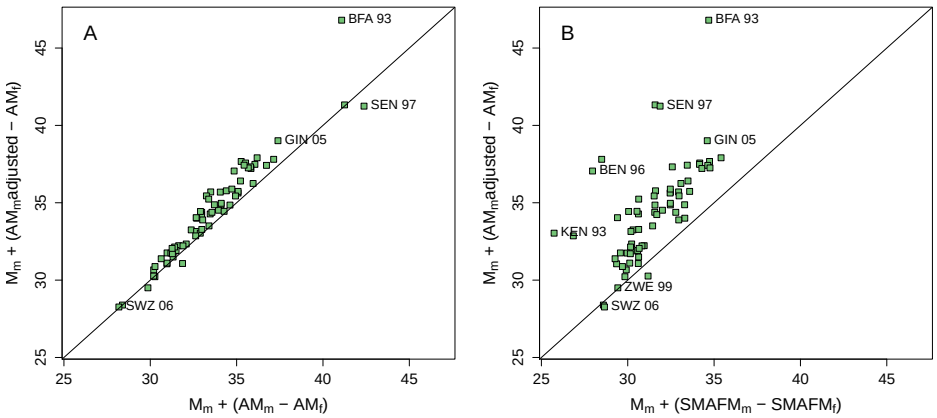

FIGURE A.21 – *Comparaisons entre (en ordonnées) l'âge moyen des pères calculé à partir des âges médians des mariés et ajusté pour la polygamie avec (en abcisses) l'âge moyen des pères calculé sans ajustement pour la polygamie (A) et l'âge moyen des pères calculé à partir des âges moyens au premier mariage (B) (66 enquêtes EDS)*

Notons également que la seconde option, basée sur les âges médians, est problématique pour les EDS, car dans la plupart des enquêtes, le statut matrimonial n'est collecté qu'à partir des questionnaires individuels. Les âges médians sont donc calculés *sur une plage d'âge réduite* ; 15-49 chez les femmes et généralement 15-54 ou 15-59 chez les hommes. Ils sont de ce fait sous-estimés. Dans une quinzaine d'enquêtes, une question sur le statut matrimonial a également été posée *au niveau du ménage* (Benin 2006, Congo RDC 2007, etc.). Ces enquêtes permettent d'estimer que l'âge médian des hommes mariés, tel qu'il est calculé à partir des fichiers individuels, est sous-estimé de 4 ans en moyenne (soit 10% environ), par rapport à celui obtenu à partir du questionnaire ménage. Ceci est présenté dans la

figure A.22, pour les hommes (à gauche) et les femmes (à droite) [36]. Le calcul de l'âge médian des femmes mariées à partir du questionnaire individuel sous-estime cet âge de 2.7 ans en moyenne (soit également à peu près 10%). Ces sous-estimations n'ont toutefois qu'un impact limité sur les probabilités de survie estimées, puisque c'est la *différence entre ces âges médians* qui intervient dans les calculs, et celle-ci n'est que légèrement surestimée (d'un peu plus d'1 an en moyenne sur les 16 enquêtes considérées ici) [37].

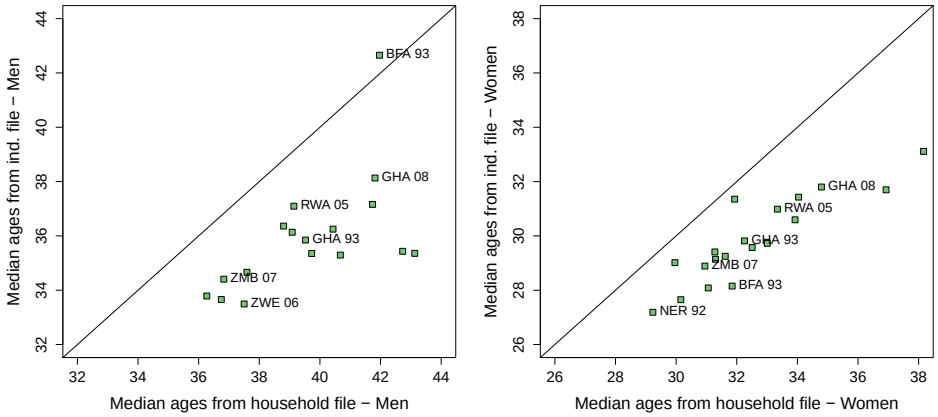

FIGURE A.22 – *Comparaison des âges médians des femmes et des hommes mariés calculés à partir des fichiers ménages (en abcsisses) et des fichiers individuels (en ordonnées) - 16 enquêtes EDS*

Enfin, en plus du calcul direct et de ces deux approches indirectes, signalons que les données d'enquêtes EDS permettent d'estimer $M_{♂}$ en adoptant un quatrième angle d'approche. En effet, à partir de 1995, plus de 50 enquêtes ont intégré une question sur l'âge des partenaires des femmes actuellement *en union*. En supposant que les naissances qui sont survenues dans l'année sont issues de ces unions actuelles, elles peuvent être tabulées par âge des partenaires. On obtient alors des valeurs de $M_{♂}$ estimées plus "directement". Ces estimations sont présentées en ordonnées dans les graphes A.23.A et A.23.B, et elles sont croisées avec les deux autres types d'estimation. Les valeurs obtenues sont à nouveau nettement supérieures à celles déduites des âges au premier mariage, de 4 ans en moyenne sur les 59 enquêtes considérés. Par contre, on note une meilleure concordance avec les valeurs issues des âges médians, sauf dans certains cas comme en Namibie et au Swaziland. Il faut garder en tête que ces estimations sont également biaisées vers le haut, car elles sont basées uniquement sur les naissances survenues en union, dont la proportion est plus importante chez les femmes plus âgées (et donc chez les pères plus âgés également).

En définitive, il nous semble préférable d'exploiter les données de façon directe, quand les informations nécessaires sont disponibles. Il serait souhaitable que les questions relatives à la date de la dernière naissance

36. La seule exception est celle de l'enquête burkinabé de 1993, où le questionnaire individuel a été administré aux hommes de plus de 18 ans. Cette enquête n'est pas reprise dans le calcul des moyennes présentées ici

37. La plus forte sous-estimation de cet écart est observée à partir de l'enquête menée au Liberia en 2007, elle se chiffre à 2.7 ans.

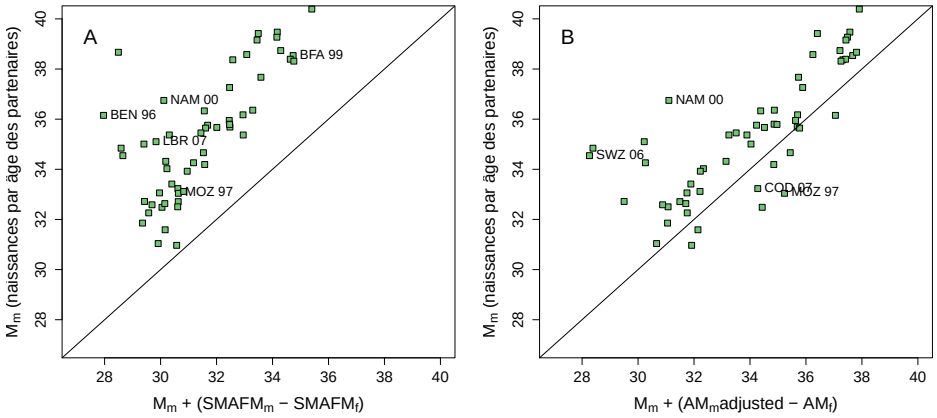

FIGURE A.23 – *Comparaisons entre (en ordonnées) l'âge moyen des pères calculé en tabulant les naissances par l'âge des partenaires avec (en abcisses) l'âge moyen des pères calculé à partir des âges médians des mariés (ajustés) (A) ou à partir des âges moyens au premier mariage (B) (59 enquêtes EDS)*

soient intégrées plus systématiquement dans les enquêtes EDS menées auprès des hommes.

A.4 Annexes au chapitre 4

A.4.1 Formules de Timaeus et Nunn (1997)

Nous détaillons ici les formules élaborées par Timaeus et Nunn (1997) pour calculer le facteur d'ajustement permettant de réduire la proportion de non-orphelins pour tenir compte des biais associés au VIH-sida.

1. Notons $N(a)$ le nombre de femmes qui ont donné naissance il y a a années, et qui ont survécu. Si $n(x,a)$ renvoie au nombre de femmes d'âge x il y a a années (qu'elles aient survécu ou non), $N(a)$ s'obtient simplement en faisant intervenir une probabilité de survie et des taux de fécondité et en intégrant sur tous les âges de reproduction (cfr. p.103) :

$$N(a) = \int_{\alpha}^{\beta} n(x,a)\ {}_{a}p_x\ m(x)dx \tag{A.9}$$

2. Parmi ces femmes, nous pouvons distinguer trois catégories :
   (a) celles qui étaient séropositives quand elles ont donné naissance : $N^*(a)$
   (b) celles qui n'étaient *pas* séropositives au moment de cette naissance et qui n'ont *pas* été infectées par la suite : $N^-(a)$
   (c) celles qui n'étaient *pas* séropositives au moment de la naissance mais le sont devenues par la suite : $N^i(a)$.

   Ces trois nombres, $N^*(a)$, $N^-(a)$ et $N^i(a)$, peuvent être exprimés en faisant intervenir des probabilités de décès et d'infection au VIH.

   (a) Le nombre de femmes qui étaient *séropositives* quand elles ont donné naissance est noté $N^*(a)$. Ces femmes font face à des probabilités de décès plus élevées : ${}_{a}p_x^+$. Elles connaissent également une réduction de leur fécondité. Notons $F(a,x)$ le ratio, par âge, de la fécondité des séropositives sur les séronégatives. On peut alors exprimer $N^*(a)$ comme ceci :

$$N^*(a) = \int_{\alpha}^{\beta} n^+(x,a)\ {}_{a}p_x^+\ F(a,x)m(x)dx \tag{A.10}$$

   Dans cette équation, $n^+(x,a)$ correspond au nombre de femmes séropositives d'âge x observé il y a a années.

   Si ces femmes *n'avaient pas* connu une fécondité différente de celle des séronégatives, leur nombre aurait pu s'exprimer en éliminant simplement le facteur $F(a,x)$:

$$N^+(a) = \int_{\alpha}^{\beta} n^+(x,a)\ {}_{a}p_x^+\ m(x)dx \tag{A.11}$$

   (b) Le nombre de femmes qui n'étaient pas *séropositives* au moment de la naissance et qui n'ont *pas été infectées* par la suite est noté $N^-(a)$. La probabilité d'être restée séronégative, noté ${}_{a}v_x$, peut s'exprimer à partir du taux instantané d'infection ($\mu^i(z)$) :

$${}_{a}v_x = exp\left[-\int_{x}^{x+a} \mu^i(z)dz\right] \tag{A.12}$$

A partir de là, si $n^-(x,a)$ correspond au nombre de femmes séronégatives d'âge x observé il y a a années, on peut écrire :

$$N^-(a) = \int_\alpha^\beta n^-(x,a)\ {}_ap_x^-\ {}_av_x\ m(x)dx \tag{A.13}$$

(c) Le nombre de femmes qui n'étaient *pas séropositives* mais qui *le sont devenues* depuis cette naissance est noté $N^i(a)$. Il s'exprime en faisant intervenir le nombre $N^-(a)$, et le taux instantané d'infection : [38]

$$N^i(a) = \int_\alpha^\beta \int_x^{x+a} N^-(z)\ \mu^i(z)\ {}_{x+a-z}p_z^+\ m(x)dz\ dx \tag{A.14}$$

3. Notons h l'intensité de la transmission verticale. Supposons que tous les enfants infectés verticalement décèdent avant cinq ans (ce qui aura pour effet de sur-estimer le biais de sélection). Supposons également que les orphelins ne font pas face à une surmortalité spécifique.
4. Supposons que $F(a,x)$ prend une valeur unique de 0.8, quelque soit l'âge des femmes.
5. Tous les éléments sont réunis pour exprimer la proportion de mères survivantes qui est observée *en présence à la fois d'une moindre fécondité des séropositives et d'une transmission verticale*. Le numérateur correspond au total entre (1) le nombre de mères séropositives à la naissance de leur enfant (multipliée par $1-h$ en raison de la transmission verticale), et (2) le nombre de mères non séropositives à la naissance. Le dénominateur correspond à la somme entre, d'une part, (1) le nombre de femmes séropositives qui ont donné naissance à l'âge x (ce qui correspond à $n^+(x,a)\ m(x)dx$) - que l'on multiplie par $1-h$ pour tenir compte de la transmission verticale et par F pour tenir compte de leur moindre fécondité - et d'autre part, (2) le nombre de femmes qui étaient séronégatives au moment de la naissance. On obtient donc [39] :

$$S^*(a) = \frac{(1-h)N^*(a) + N^-(a) + N^i(a)}{\int_\alpha^\beta \left[(1-h)\times F\times n^+(a,x) + n^-(a,x)\right]\ m(x)dx} \tag{A.15}$$

6. Pour calculer l'ampleur des biais introduits, il nous faut estimer la même proportion, mais *en l'absence de ces deux mécanismes de sélection*. Pour rappel, le nombre de femmes survivantes qui étaient séropositives quand elles ont donné naissance dans le cas où la séropositivité n'a pas d'impact sur la fécondité est noté $N^+(a)$. A partir de là, la proportion qui serait observée en l'absence des deux mécanismes de sélection s'établit comme :

$$S(a) = \frac{N^+(a) + N^-(a) + N^i(a)}{\int_\alpha^\beta \left[n^+(a,x) + n^-(a,x)\right]\ m(x)dx} \tag{A.16}$$

38. On notera que dans cette équation, les probabilités de décès des séropositives varient uniquement par âge (et pas selon la durée écoulée depuis l'infection).

39. Pour des valeurs de a supérieures à 5.

7. C'est là qu'intervient l'ajustement proposé par Timaeus et Nunn (1997). Ils montrent que la prévalence du VIH parmi les mères s'exprime comme ceci :

$$P = \frac{\int_{\alpha}^{\beta} n^{+}(a,x) \times F\ mx\ dx}{\int_{\alpha}^{\beta} n^{+}(a,x) \times F\ m(x)\ dx + \int_{\alpha}^{\beta} n^{-}(a,x)\ m(x)\ dx} \tag{A.17}$$

8. Ils supposent ensuite que *toutes les femmes qui étaient séropositives quand elles ont donné naissance décèdent dans les cinq ans*. Dès lors, on peut supprimer $N^{+}(a)$ du numérateur de $S(a)$ et $(1-h)N^{*}(a)$ du numérateur de $S^{*}(a)$. Comme les deux numérateurs sont de ce fait identiques, rapporter $S(a)$ sur $S^{*}(a)$ procure l'équation suivante :

$$\frac{S(a)}{S^{*}(a)} = \frac{\int_{\alpha}^{\beta}(1-h) \times F \times n^{+}(a,x)\ m(x)\ dx + \int_{\alpha}^{\beta} n^{-}(a,x)\ m(x)dx}{\int_{\alpha}^{\beta}\ [n^{+}(a,x) + n^{-}(a,x)]\ m(x)\ dx} \tag{A.18}$$

   Ceci procure une expression du biais introduit.

9. Il est possible de simplifier cette équation en divisant ses numérateur et dénominateur par le dénominateur de l'équation A.17. Ceci permet de calculer le facteur d'ajustement :

$$\frac{S(a)}{S^{*}(a)} = \frac{(1-h) \times P + (1-P)}{\frac{P}{F} + (1-P)} \tag{A.19}$$

10. En supposant une probabilité de transmission de 25% et une réduction de la fécondité de 20% chez les séropositives, ceci se simplifie en :

$${}_5S_n^{adj} = \frac{1 - 0.25\ P}{1 + 0.25\ P} \times_5 S_n^{*} \tag{A.20}$$

A.5 Annexes au chapitre 5

A.5.1 Démonstration de Trussell et Rodriguez (1990)

La somme $\sum_{x=0}^{n-1} x(n-x)f(x)$, qui est au numérateur de l'équation 5.5 (p. 235), est simplifiée en partant d'une situation hypothétique où les fratries sans survivants pourraient être observées :

- Factorisé, ce numérateur peut s'exprimer comme étant la somme partielle des $n-1$ premiers termes de $xnf(x) - x^2f(x)$.

- Si cette somme partielle était prise sur les n premiers termes (c'est-à-dire si on observait les fratries sans survivants et que la somme était faite jusqu'à n plutôt que jusqu'à $n-1$), ce numérateur vaudrait $n\mu - x^2f(x)$. Comme $\sigma^2 = x^2f(x) - \mu^2$, on obtient :

$$n\mu - (\sigma^2 + \mu^2)$$

- Pour le n^e terme de cette somme partielle - qui est omis ici puisque les fratries sans survivants ne sont pas observées - le premier terme de cette différence ($xnf(x)$) vaudrait n^2p^n. Le second terme de cette différence ($x^2f(x)$) vaudrait également n^2p^n. Donc nous pouvons réécrire le numérateur comme étant la somme partielle pour les $n-1$ premiers termes de $(n\mu - n^2p^n) - (\sigma^2 + \mu^2 - n^2p^n)$, ce qui permet d'annuler les deux termes n^2p^n :

$$n\mu - \sigma^2 - \mu^2$$

- Puisque $\mu = np$ et que $\sigma^2 = npq$, lorsque x suit une loi binomiale, le numérateur peut alors être réécrit comme :

$$n^2p - npq - (np)^2, \text{ soit } npq(n-1)$$

On peut procéder de la même façon du côté du dénominateur :
- la somme partielle des $n-1$ premiers termes de $n-1(n-x)f(x)$ vaut :

$$(n^2-n)f(x) - nxf(x) + xf(x)$$

- pour le n^e terme non observé, les différents termes de cette somme vaudraient respectivement $p^n(n^2-n)$, $-n^2p^n$, et np^n.

- Dès lors, lorsque cette somme partielle est prise sur les $n-1$ premiers termes, il est possible de soustraire ces éléments et de réécrire le dénominateur comme :

$$n^2 - n - p^n(n^2-n) - n\mu + n^2p^n + \mu - np^n$$

- Le dénominateur peut donc être réécrit comme :

$$n^2 - n - n\mu + \mu = n^2(1-p) - n(1-q) = nq(n-1)$$

Dès lors, on peut effectivement exprimer la proportion de survivants déclarés comme :

$$PD = \frac{npq(n-1)}{nq(n-1)} = p$$

A.5.2 Méthode des sœurs développée par Graham et al. (1989)

Comme toutes les méthodes indirectes présentées ici, le principe de la méthode des sœurs consiste à convertir une proportion en probabilité de décès. En l'occurrence, il s'agit de convertir *la proportion de sœurs qui, ayant atteint l'âge adulte, sont décédées de cause maternelle*. Cette proportion, notée Π_a, s'obtient à partir des quelques questions qui concernent donc uniquement les plus de 15 ans. Ces questions sont détaillées dans le chapitre 1 (p. 32). La proportion se calcule par âge des enquêtées (a), et elle est convertie en une probabilité de décès pour cause maternelle, elle aussi indexée par âge ($q(a)$). Lorsque le nombre de décès maternels reste limité, on préférera estimer $q(w)$, qui correspond au risque de décéder de cause maternelle *durant toute la période de reproduction*. De façon générale, les quotients $q(a)$ obtenus indirectement peuvent s'avérer très erratiques, et il est préférable de s'en tenir à une valeur calculée sur tous les âges ou sur les 30-49 ans au minimum (Garenne et Friedberg 1997).

Pour faire la conversion entre proportions et probabilités, il faut tenir compte de deux éléments : (1) de la distribution des différences d'âge entre une enquêtée et ses différentes sœurs, et (2) de la distribution par âge des risques de décès maternels.

1. La **distribution des différences entre l'âge de l'enquêtée et l'âge de ses sœurs** : Graham et al. (1989) supposent que dans l'ensemble, cette distribution est *symétrique* et de *moyenne nulle* (pour peu que la mère de l'enquêtée ait atteint la fin de sa période reproductive). Cela n'est qu'une approximation, pas tout à fait exacte car cette distribution est également influencée par la croissance de la population.

   Ce que relèvent par contre ces auteurs, c'est que dans le cas de la mortalité maternelle, cette distribution des différences d'âge est tronquée : d'un côté, les sœurs *plus âgées* sont toutes comptées, parce qu'elles auront toutes au moins 15 ans, alors que de l'autre, les sœurs *plus jeunes* ne sont pas toutes prises en compte, puisque les questions sur les décès maternels ne concernent que celles qui ont atteint 15 ans. Il faut donc procéder à un ajustement, puisque les sœurs qui ne sont pas encore entrées en période de reproduction ne sont pas comptabilisées dans le calcul. Sans cette correction, la proportion Π_a risque d'être trop élevée, car son dénominateur est trop faible par rapport au numérateur.

   Cette correction doit surtout se faire sur les premiers groupes d'âge. En effet, par rapport aux enquêtées plus âgées, les enquêtées plus jeunes ont moins de sœurs déjà entrées en période de reproduction (et donc soumises au risque de décès maternel) et plus de sœurs qui vont bientôt entrer en période d'exposition. Au-delà d'un certain âge, les enquêtées ont des sœurs qui sont pour la plupart déjà soumises au risque de décès maternel. Pratiquement, on rehausse

donc le nombre de sœurs déclarées par les enquêtées de *moins de 25 ans*. Cela est fait simplement en multipliant le nombre d'enquêtées de 15-19 et 20-24 ans par le nombre moyen de sœurs des enquêtées de plus de 25 ans. Si N_{15-19} correspond au nombre d'enquêtées de 15-19 ans et S_{15-19} au nombre de leurs sœurs, le nombre "ajusté" de sœurs de 15-19 ans se calcule comme :

$$S^{adj}_{15-19} = N_{15-19} \times \frac{S_{25-49}}{N_{25-49}} \tag{A.21}$$

2. **La distribution par âge des risques de décès maternels** : la relation entre la proportion de sœurs décédées de cause maternelle et la probabilité de décéder de cause maternelle dépend bien entendu aussi de l'âge des enquêtés. Les enquêtées plus âgées vont en effet fournir des informations sur des femmes qui, en moyenne, ont été exposées aux risques de mortalité sur l'ensemble de leur vie reproductive. Par contre, les enquêtées plus jeunes procurent des informations sur des femmes qui, en moyenne, ont été moins longtemps exposées. Des facteurs de correction (notés βn) ont donc été calculés par Graham et al. (1989), pour tenir compte de la durée variable de l'exposition au risque selon l'âge. Ils sont repris dans le tableau A.17.

n	β_n	n	β_n
12.5	0.048	42.5	0.802
15	0.073	45	0.856
17.5	0.107	47.5	0.900
20	0.151	50	0.934
22.5	0.206	52.5	0.958
25	0.270	55	0.975
27.5	0.343	57.5	0.986
30	0.421	60	0.992
32.5	0.503	62.5	0.996
35	0.585	65	0.998
37.5	0.664	67.5	0.999
40	0.737	70	1.000

TABLE A.17 – *Coefficients de conversion permettant de calculer les unités d'exposition au risque de décéder de cause maternelle. - Graham et al. (1989)*

La valeur de 32.5 s'élève par exemple à 0.503. Cela peut s'interpréter comme si 50.3% des sœurs des répondantes de 30 à 34 ans avaient connu une exposition complète. Appliqués au nombre de sœurs soumises au risque pour chaque groupe d'âge, les facteurs d'ajustement procurent donc des *unités d'exposition au risque de décès maternel*. Logiquement, ces facteurs augmentent avec l'âge et atteignent des niveaux proches de l'unité lorsque la quasi-totalité des sœurs des répondantes ont connu une exposition complète au risque de décès maternel.

Pour obtenir les quotients de mortalité maternelle, il suffit alors de rapporter les décès maternels aux unités d'exposition au risque de décès. Pour obtenir $q(w)$, le risque de décès maternel sur l'ensemble de la période de reproduction, on rapportera la somme des décès survenus entre

15 et 50 ans à la somme des unités d'observations entre ces âges. Une série de coefficients est également disponible pour dater les estimations.

Le rapport de mortalité maternelle (RMM) constitue un autre indice de mortalité maternelle couramment utilisé. Il se calcule comme ceci :

$$RMM = 1 - (1 - q(w))^{1/ISF} \tag{A.22}$$

Multiplié par 100000, il donne le taux de mortalité maternelle, qui correspond au nombre de décès maternels pour 100000 naissances vivantes.

A.5.3 Coefficients utilisés pour dater les estimations indirectes basées sur la survie des frères et sœurs

Age	n	γ_0	γ_1
20-24	25	3.23	1.12
25-29	30	5.46	1.95
30-34	35	7.52	2.78
35-39	40	9.38	3.62
40-44	45	11.00	4.45
45-49	50	12.32	5.28

TABLE A.18 – *Coefficients utilisés par Timaeus et al. (2001) pour dater les estimations indirectes basées sur la survie des frères et sœurs*

A.6 Annexes au chapitre 6

A.6.1 Données manquantes sur la survie des parents

	Enquête	Année	Mères	Pères	Deux parents	Au moins un
1	Ghana	2008	0.2	0.2	0.1	0.3
2	Zambie	2007	0.1	0.3	0.0	0.4
3	Nigeria	2008	0.3	0.3	0.2	0.4
4	Ethiopie	2005	0.2	0.3	0.1	0.4
5	Burkina Faso	1993	0.4	0.4	0.4	0.5
6	Côte d'Ivoire	2005	0.1	0.5	0.1	0.5
7	Tchad	2004	0.4	0.5	0.4	0.5
8	Guinée	2005	0.4	0.5	0.3	0.6
9	Tanzanie	2007	0.1	0.5	0.1	0.6
10	Niger	1992	0.4	0.5	0.4	0.6
11	Bénin	2001	0.3	0.6	0.3	0.6
12	Congo RDC	2007	0.1	0.6	0.1	0.6
13	Ethiopie	2000	0.2	0.6	0.2	0.6
14	Mali	2006	0.5	0.5	0.3	0.6
15	Mali	1996	0.5	0.6	0.5	0.6
16	Malawi	2004	0.1	0.6	0.1	0.7
17	Niger	2006	0.4	0.5	0.2	0.7
18	Togo	1998	0.5	0.6	0.5	0.7
19	Congo	2005	0.4	0.6	0.3	0.7
20	Ouganda	2006	0.2	0.6	0.1	0.7
21	Zambie	1992	0.5	0.7	0.5	0.7
22	Tanzanie	2004	0.1	0.7	0.1	0.8
23	Mozambique	2003	0.3	0.7	0.2	0.8
24	Malawi	2000	0.3	0.8	0.3	0.8
25	Burkina Faso	2003	0.6	0.6	0.4	0.8
26	Ghana	2003	0.4	0.8	0.3	0.8
27	Zambie	2002	0.4	0.7	0.3	0.9
28	Malawi	1992	0.5	0.8	0.5	0.9
29	Ghana	1998	0.4	0.9	0.4	0.9
30	Centrafique	1995	0.8	0.9	0.8	0.9
31	Bénin	1996	0.8	0.9	0.7	1.0
32	Tchad	1997	0.9	1.0	0.9	1.0
33	Ouganda	2001	0.4	0.9	0.3	1.0
34	Côte d'Ivoire	1994	0.8	1.0	0.8	1.0
35	Bénin	2006	0.6	0.7	0.3	1.0
36	Rwanda	1992	0.3	1.1	0.2	1.1
37	Mali	2001	0.9	1.1	0.9	1.1
38	Ghana	1993	0.8	1.1	0.8	1.1
39	Niger	1998	1.1	1.1	1.1	1.2
40	Zambie	1996	0.8	1.1	0.8	1.2
41	Nigeria	2003	1.0	0.7	0.5	1.2
42	Tanzanie	1999	0.5	1.2	0.4	1.3
43	Comores	1996	1.0	1.3	1.0	1.3
44	Madagascar	2009	0.3	1.3	0.3	1.3
45	Cameroun	2004	0.4	1.2	0.2	1.4
46	Liberia	2007	0.5	1.2	0.4	1.4
47	Rwanda	2005	0.3	1.3	0.2	1.5
48	Ouganda	1995	1.0	1.5	0.9	1.6
49	Tanzanie	1996	1.0	1.6	0.9	1.6
50	Cameroun	1991	0.4	1.7	0.4	1.7
51	Madagascar	2004	0.6	1.6	0.4	1.8
52	Namibie	2007	0.6	1.7	0.5	1.8
53	Swaziland	2006	0.4	1.6	0.1	1.9
54	Sénégal	2005	1.3	1.8	1.2	1.9
55	Sierra Leone	2008	1.3	1.8	1.2	1.9

56	Sénégal	1993	1.7	1.9	1.6	1.9
57	Zimbabwe	2006	0.3	1.8	0.1	1.9
58	Zimbabwe	1994	0.8	2.0	0.7	2.0
59	Mozambique	1997	1.6	2.1	1.5	2.2
60	Guinée	1999	2.0	2.2	2.0	2.2
61	Cameroun	1998	0.5	2.3	0.5	2.3
62	Madagascar	1997	0.7	2.4	0.7	2.4
63	Kenya	2003	0.5	2.6	0.4	2.7
64	Namibie	1992	2.1	2.8	2.0	2.9
65	Kenya	1998	0.9	3.0	0.9	3.0
66	Tanzanie	1992	1.8	3.0	1.7	3.0
67	Kenya	1993	1.1	3.0	1.0	3.1
68	Namibie	2000	1.5	3.2	1.5	3.3
69	Rwanda	2000	0.9	3.2	0.7	3.4
70	Gabon	2000	2.6	3.2	2.4	3.5
71	Zimbabwe	1999	1.0	3.5	0.9	3.7
72	Lesotho	2004	0.5	3.7	0.3	3.8
73	Madagascar	1992	1.0	3.8	0.9	3.8
74	Afrique du Sud	1998	2.3	4.6	2.1	4.8
75	Nigeria	1999	4.6	6.0	4.1	6.5

TABLE A.19 – *Proportion (en %) d'enfants de moins de 15 ans pour lesquels l'information est manquante sur la survie des mères, des pères, des deux parents ou sur la survie de l'un d'eux, 75 enquêtes EDS*

A.6.2 Notes sur le traitement des réponses manquantes relatives à la survie des parents

Rappelons que pour traiter des réponses manquantes sur la survie des parents, Bicego et al. (2003) reclassent une certaine proportions d'enfants comme orphelins ou non-orphelins, à partir de seuils pré-établis (section 6.1.2, p. 286). Ces seuils dépendent de la survie de l'*autre* parent, mais ils sont fixes pour toutes les enquêtes.

Ceci paraît arbitraire, car la proportion d'orphelins parmi ceux qui ignorent si l'un de leurs parents survit va logiquement varier avec la mortalité. La proportion de doubles orphelins parmi les orphelins de mères ou de pères va également varier selon les enquêtes. Ceci apparaît dans le graphe A.24, construit à partir de 75 EDS africaines. A gauche, la proportion d'orphelins de père est croisée avec la proportion de doubles orphelins parmi les orphelins de mère [40]. On constate - logiquement - une relation positive entre les deux séries de valeurs, ce qui signifie que les seuils devraient varier avec le niveau global de la mortalité [41].

Une approche alternative consisterait donc à se baser plutôt sur les proportions *observées* à partir des réponses non-manquantes. Par exemple, les enfants orphelins de *mère* dont les données sont manquantes sur la survie du *père* pourraient être reclassés comme *doubles orphelins* en fonction de la proportion observée de doubles orphelins (du même âge) parmi ceux dont la mère est décédée [42].

40. A droite, la proportion d'orphelins de mère est croisée avec la proportion de doubles orphelins parmi les orphelins de père.

41. On note également que les proportions de doubles orphelins parmi ceux qui sont orphelins de père sont inférieures à 50%, tout comme la plupart des proportions de doubles orphelins parmi ceux qui sont orphelins de mère.

42. Si la mère de ces enfants est survivante, la proportion à retenir serait la proportion d'orphelins de pères (du même âge) parmi ceux dont la mère est survivante.

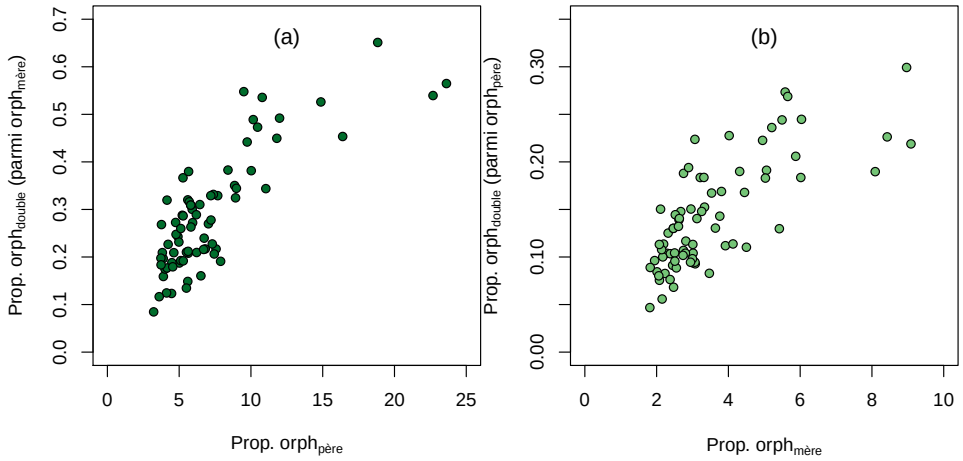

FIGURE A.24 –
(a) Proportion de doubles orphelins parmi les enfants qui sont orphelins de mère (selon la proportion d'orphelins de père) - Enquêtes EDS (n= 75)
(b) Proportion de doubles orphelins parmi les enfants qui sont orphelins de père (selon la proportion d'orphelins de mère) - Enquêtes EDS (n= 75)

Mais ceci a-t-il un réel intérêt ? N'est-il pas préférable d'ignorer simplement les réponses manquantes ? Pour mesurer l'effet de différents types d'imputation, nous avons calculé les proportions d'orphelins à partir de quatre options différentes. Ces proportions sont ensuite comparées à celles calculées *en ignorant les réponses non manquantes*. La figure A.25 présente les résultats de ces comparaisons, toujours à partir des 75 enquêtes EDS.

1. Le graphe en haut à gauche (A.25.1) présente par exemple les *rapports* entre, d'une part, les proportions calculées en supposant que toutes les réponses manquantes renvoient à des orphelins et, d'autre part, les proportions calculées uniquement sur les réponses non manquantes. Logiquement, ce premier calcul maximise les proportions d'orphelins. Celles qui concernent les 0-4 ans sont les plus rehaussées ; elles connaissent un accroissement moyen de plus de 40% dans le cas des orphelins de mère et de père, et les proportions de double orphelins à ces âges sont quasiment multipliées par trois. *Ce premier type d'imputation ne devrait donc pas être utilisé*, car il a un impact très important sur les estimations, et il est peu probable que toutes les réponses manquantes renvoient à des orphelins.
2. Le second graphe, en haut à droite (A.25.2), présente la situation inverse : tous les enfants pour lesquels la survie des parents est inconnue sont reclassés comme non-orphelins. C'est par exemple le choix de Rutstein (2008). Les proportions sont logiquement plus basses que celles obtenues sur les seules réponses non manquantes, mais les deux séries de proportions sont nettement plus proches. La médiane des rapports pour les 5-9 ans se chiffre à 0.98 pour les orphelins de mère, et 0.99 pour les orphelins de père et les doubles orphelins.
3. Le troisième graphe présente le résultat de la correction suggérée par Bicego et al. (2003), à partir des seuils qu'ils ont pré-établis. Celle-ci affecte surtout les proportions de doubles orphelins et les plus jeunes enfants. Entre 0 et 4 ans, les proportions d'orphelins de mère connaissent une hausse de 1.5% en moyenne, 6% chez les orphelins

de père et 12% chez les doubles orphelins. A nouveau, ceci nous paraît trop important ; les seuils retenus pour reclasser les réponses manquantes comme orphelins s'avèrent en fait nettement plus élevés que lorsqu'ils sont calculés à partir des données [43].

4. Enfin, le dernier graphe présente le résultat du recodage des données manquantes à partir des proportions calculées directement dans les enquêtes, en tenant compte de la survie du parent pour lequel nous disposons d'informations. Comme nous l'avons précisé p.286, cette correction revient à supposer que la méconnaissance de la survie de son père ou de sa mère n'est pas associée à une plus grande probabilité d'être orphelin de père ou de mère. C'est en somme la même hypothèse que lorsque les données manquantes sont simplement ignorées. Par contre, cette approche tient compte de l'information fournie sur la survie de l'autre parent. Nous observons sur le graphe que *les proportions ne sont modifiées que de façon marginale pour les orphelins de mères ou de pères*, par rapport à la situation où on ignore tout simplement les données manquantes [44].

En définitive, les trois premières options nous semblent "forcer" les données. La quatrième option, qui consiste à reclasser les réponses manquantes en tenant compte de la survie de l'autre parent nous paraît plus appropriée, mais elle ne change quasiment rien par rapport à la solution qui consiste à ne retenir que les réponses non manquantes. Pour l'ensemble des proportions d'orphelins présentées dans ce document, *nous avons donc simplement ignoré les réponses non-manquantes.*

43. Ceci peut s'observer à partir du graphe A.24. A titre indicatif, la médiane des proportions de doubles orphelins (de 0-14 ans) parmi les orphelins de mère se chiffre à 0.29 sur l'ensemble des enquêtes EDS considérées ici. 29% des enfants orphelins de mère - pour lesquels l'information est manquante sur la survie du père - pourraient donc être considérés comme doubles orphelins. Par comparaison, la correction de Bicego et al. (2003) reclasse 50% de ces enfants comme doubles orphelins. La différence est encore plus marquée pour les orphelins de pères qui ignorent si leur mère a survécu : une correction basée sur les proportions observées de doubles orphelins parmi les orphelins de pères reclasserait en moyenne 14% de ces enfants comme doubles orphelins, contre 50% si l'on suit Bicego et al. (2003).

44. Les proportions de doubles orphelins sont légèrement rehaussées, au maximum de 2.5% en moyenne chez les doubles orphelins de 0-4 ans.

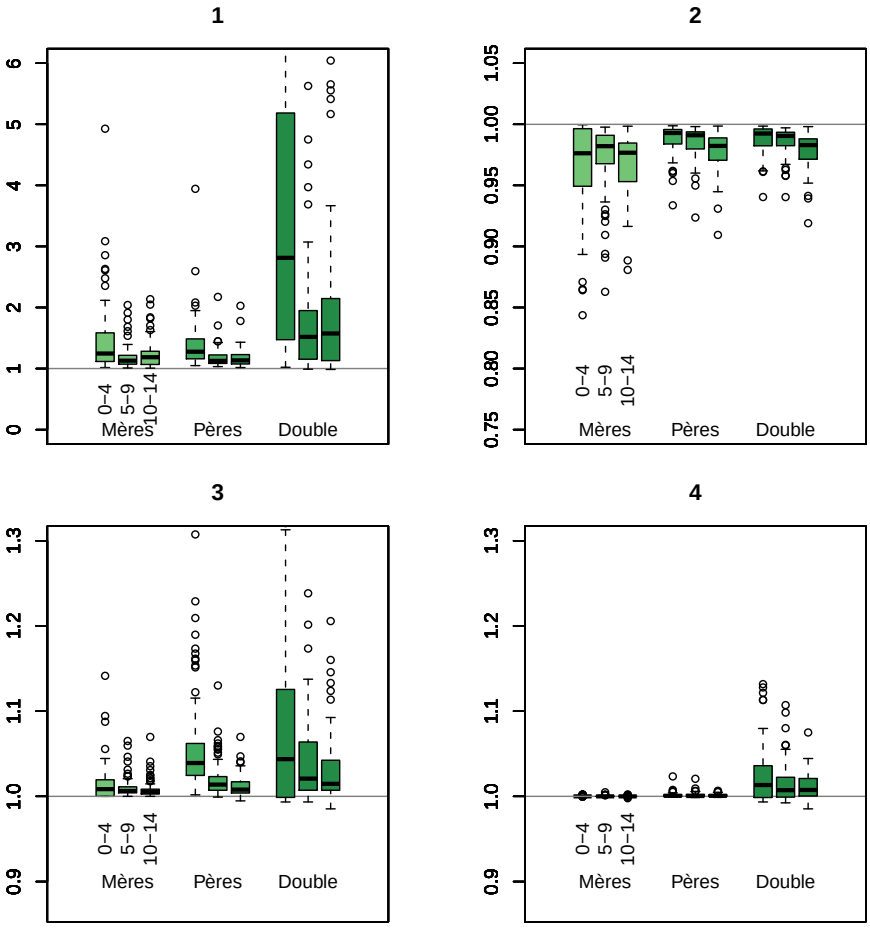

FIGURE A.25 –
*(1) Rapports entre les proportions d'orphelins calculées en retenant **toutes les réponses manquantes comme orphelins** et les proportions d'orphelins calculées sur les réponses non-manquantes*
*(2) Rapports entre les proportions d'orphelins calculées en retenant **toutes les réponses manquantes comme non-orphelins** et les proportions d'orphelins calculées sur les réponses non-manquantes*
*(3) Rapports entre les proportions d'orphelins calculées **à partir de l'imputation de Bicego et al. (2003)** et les proportions d'orphelins calculées sur les réponses non-manquantes*
*(4) Rapports entre les proportions d'orphelins calculées **en imputant les réponses manquantes selon la survie de l'autre parent** et les proportions d'orphelins calculées sur les réponses non-manquantes*
Nos calculs, à partir des enquêtes EDS (n = 75)

A.6.3 Données manquantes sur les frères et sœurs

Données manquantes sur le sexe des frères et sœurs

	Total	Décédés	Survivants	NSP	Pays	Année	Type
1	<0.1	0.1	<0.1	58.9	Cameroun	2004	♀
2	<0.1	<0.1	<0.1	29.0	Cameroun	2004	♂
3	<0.1	<0.1	<0.1	<0.1	Centrafique	1995	♀
4	<0.1	<0.1	<0.1	<0.1	Tchad	2004	♀
5	<0.1	<0.1	<0.1	<0.1	Côte d'Ivoire	1994	♀
6	<0.1	<0.1	<0.1	<0.1	Madagascar	1992	♀
7	<0.1	0.1	<0.1	<0.1	Mali	1996	♀
8	<0.1	<0.1	<0.1	<0.1	Mozambique	1997	♀
9	0.1	<0.1	<0.1	97.4	Congo	2005	♀
10	0.1	<0.1	<0.1	92.0	Congo	2005	♂
11	0.1	0.1	<0.1	48.0	Congo RDC	2007	♀
12	0.1	<0.1	<0.1	76.7	Guinée	2005	♀
13	0.1	0.3	<0.1	15.2	Lesotho	2004	♀
14	0.1	<0.1	<0.1	42.5	Madagascar	2009	♀
15	0.1	0.1	<0.1	58.4	Malawi	2004	♀
16	0.1	0.3	<0.1	26.5	Malawi	2000	♀
17	0.1	<0.1	<0.1	64.2	Niger	2006	♀
18	0.1	0.3	<0.1	52.6	Sénégal	1993	♀
19	0.1	0.1	<0.1	66.0	Tanzanie	2004	♀
20	0.1	<0.1	<0.1	52.7	Zambie	2007	♂
21	0.1	0.1	<0.1	80.9	Zambie	1996	♀
22	0.1	0.1	<0.1	84.6	Zambie	2002	♀
23	0.1	0.5	<0.1	36.6	Zimbabwe	1994	♀
24	0.2	0.1	0.1	64.2	Guinée	1999	♀
25	0.2	0.1	<0.1	15.9	Rwanda	2005	♀
26	0.2	0.4	0.1	37.1	Tanzanie	1996	♀
27	0.2	0.1	<0.1	64.8	Zambie	2007	♀
28	0.2	0.2	0.1	56.1	Zimbabwe	2006	♀
29	0.2	0.4	<0.1	41.3	Zimbabwe	1999	♀
30	0.3	0.7	<0.1	59.0	Niger	1992	♀
31	0.3	0.5	<0.1	74.4	Sénégal	2005	♀
32	0.3	0.3	<0.1	66.3	Swaziland	2006	♀
33	0.3	0.7	0.1	60.5	Tanzanie	1996	♂
34	0.3	0.2	0.1	58.9	Zimbabwe	2006	♂
35	0.4	0.4	0.2	85.8	Bénin	2006	♀
36	0.4	0.8	0.1	82.7	Tchad	1997	♀
37	0.4	0.3	<0.1	48.2	Ethiopie	2000	♀
38	0.4	1.0	0.1	83.3	Kenya	2003	♀
39	0.4	0.5	<0.1	80.4	Kenya	2009	♀
40	0.4	0.5	0.2	87.1	Madagascar	2004	♀
41	0.4	0.4	<0.1	65.8	Mozambique	2003	♀
42	0.4	0.4	0.2	74.7	Namibie	2000	♀
43	0.4	0.7	0.1	78.5	Togo	1998	♀
44	0.4	0.8	0.1	82.5	Zimbabwe	1994	♂
45	0.5	0.1	0.1	69.9	Ethiopie	2005	♀
46	0.5	1.2	0.1	93.1	Kenya	1998	♀
47	0.5	0.3	0.1	89.2	Mali	2006	♀
48	0.5	0.2	<0.1	34.2	Ouganda	2006	♀
49	0.6	2.0	0.4	<0.1	Gabon	2000	♀
50	0.6	1.6	<0.1	78.7	Madagascar	1997	♀
51	0.6	0.6	<0.1	84.8	Nigeria	1999	♀
52	0.6	0.6	0.2	61.9	Nigeria	2008	♂
53	0.6	0.7	0.2	63.3	Nigeria	2008	♀
54	0.6	1.1	0.1	70.9	Ouganda	1995	♀
55	0.7	1.9	0.1	90.8	Burkina Faso	2003	♀

56	0.7	0.4	0.2	80.7	Mali	2001	♀
57	0.8	0.8	0.3	75.6	Malawi	1992	♂
58	0.8	1.2	0.2	57.1	Namibie	1992	♀
59	0.8	0.7	<0.1	80.8	Nigeria	1999	♂
60	0.8	0.4	0.1	26.5	Rwanda	2000	♀
61	0.9	0.9	0.3	75.0	Malawi	1992	♀
62	1.1	2.2	0.3	83.4	Bénin	1996	♀
63	1.2	0.9	0.1	77.8	Ouganda	2001	♀
64	1.2	0.5	0.1	62.3	Ouganda	1995	♂
65	1.8	4.3	0.1	77.0	Burkina Faso	1999	♀
66	1.9	3.1	0.7	94.7	Cameroun	1998	♀
67	1.9	2.8	0.4	83.2	Afrique du Sud	1998	♀
68	9.0	0.1	0.1	96.1	Sierra Leone	2008	♀

TABLE A.20 – *Proportion d'observations dont le sexe est inconnu - par statut de survie (en %) - enquêtes EDS*

Données manquantes sur la survie des frères et sœurs

	Total	Frères	Soeurs	NSP	Pays	Année	Type
1	<0.1	<0.1	<0.1	65.5	Cameroun	2004	♀
2	<0.1	<0.1	<0.1	10<0.1	Cameroun	2004	♂
3	<0.1	<0.1	<0.1	49.8	Tchad	2004	♀
4	<0.1	<0.1	<0.1	88.2	Côte d'Ivoire	1994	♀
5	<0.1	<0.1	<0.1	<0.1	Gabon	2000	♀
6	<0.1	<0.1	<0.1	5.8	Malawi	2000	♀
7	<0.1	<0.1	<0.1	<0.1	Mali	1996	♀
8	<0.1	<0.1	<0.1	92.0	Mozambique	1997	♀
9	0.1	<0.1	0.1	44.1	Centrafique	1995	♀
10	0.1	<0.1	<0.1	10<0.1	Congo	2005	♂
11	0.1	<0.1	<0.1	94.5	Congo	2005	♀
12	0.1	0.1	0.1	20.8	Lesotho	2004	♀
13	0.1	0.1	<0.1	62.2	Madagascar	2009	♀
14	0.1	0.1	0.1	76.3	Madagascar	1992	♀
15	0.1	<0.1	<0.1	40.7	Malawi	2004	♀
16	0.1	0.1	<0.1	24.4	Niger	1992	♀
17	0.1	0.1	<0.1	77.4	Niger	2006	♀
18	0.1	0.1	<0.1	43.5	Sénégal	1993	♀
19	0.1	<0.1	<0.1	62.7	Tanzanie	2004	♀
20	0.1	0.1	<0.1	2<0.1	Tanzanie	1996	♀
21	0.1	<0.1	<0.1	76.9	Zambie	2002	♀
22	0.1	<0.1	<0.1	43.2	Zambie	1996	♀
23	0.1	0.1	<0.1	22.2	Zimbabwe	1994	♀
24	0.1	0.1	<0.1	38.3	Zimbabwe	2006	♀
25	0.2	<0.1	<0.1	44.0	Bénin	2006	♀
26	0.2	<0.1	0.1	42.5	Tchad	1997	♀
27	0.2	0.1	0.1	85.2	Congo RDC	2007	♀
28	0.2	<0.1	<0.1	96.3	Guinée	2005	♀
29	0.2	0.1	0.1	64.4	Guinée	1999	♀
30	0.2	<0.1	<0.1	48.8	Kenya	2003	♀
31	0.2	<0.1	<0.1	42.9	Madagascar	2004	♀
32	0.2	0.1	<0.1	55.7	Sénégal	2005	♀
33	0.2	<0.1	<0.1	32.3	Togo	1998	♀
34	0.2	<0.1	0.1	30.5	Tanzanie	1996	♂
35	0.2	0.1	0.1	79.1	Zambie	2007	♂
36	0.2	0.1	0.1	74.2	Zambie	2007	♀
37	0.2	0.1	<0.1	38.3	Zimbabwe	1994	♂
38	0.3	<0.1	0.1	24.2	Bénin	1996	♀
39	0.3	<0.1	<0.1	38.6	Burkina Faso	2003	♀
40	0.3	<0.1	<0.1	57.5	Kenya	1998	♀

41	0.3	0.1	0.1	44.2	Namibie	2000	♀
42	0.3	0.2	0.2	60.7	Zimbabwe	1999	♀
43	0.4	<0.1	0.1	82.1	Kenya	2009	♀
44	0.4	0.1	<0.1	78.4	Mali	2006	♀
45	0.4	0.2	0.1	69.0	Mozambique	2003	♀
46	0.4	0.1	0.1	75.2	Swaziland	2006	♀
47	0.4	0.2	0.1	48.6	Ouganda	1995	♀
48	0.4	0.2	0.1	73.7	Zimbabwe	2006	♂
49	0.5	0.2	0.1	76.0	Ethiopie	2005	♀
50	0.5	0.1	0.1	59.1	Madagascar	1997	♀
51	0.5	0.2	0.1	44.8	Malawi	1992	♂
52	0.5	0.1	0.1	59.7	Mali	2001	♀
53	0.6	0.5	0.1	75.3	Ethiopie	2000	♀
54	0.6	0.2	0.1	48.8	Malawi	1992	♀
55	0.6	0.2	0.2	60.9	Nigeria	2008	♀
56	0.6	0.1	0.1	84.1	Nigeria	1999	♀
57	0.6	0.2	0.2	59.0	Nigeria	2008	♂
58	0.8	0.1	<0.1	37.7	Cameroun	1998	♀
59	0.8	0.1	0.2	81.4	Nigeria	1999	♂
60	0.9	0.2	0.2	39.0	Burkina Faso	1999	♀
61	0.9	0.5	0.3	62.9	Namibie	1992	♀
62	1.1	0.3	0.2	72.6	Ouganda	2001	♀
63	1.2	1.4	0.6	86.4	Rwanda	2005	♀
64	1.2	0.8	0.8	84.3	Ouganda	2006	♀
65	1.4	0.3	0.2	62.4	Afrique du Sud	1998	♀
66	1.6	0.8	0.4	83.1	Ouganda	1995	♂
67	2.1	2.1	1.0	71.7	Rwanda	2000	♀
68	9.3	0.4	0.4	99.0	Sierra Leone	2008	♀

TABLE A.21 – *Proportion d'observations dont le statut de survie est inconnu - par sexe (en %) - enquêtes EDS*

Données manquantes sur l'âge à l'enquête des frères et sœurs survivants

	Total	Frères	Soeurs	NSP	Pays	Année	Type
1	<0.1	<0.1	<0.1	10<0.1	Centrafique	1995	♀
2	<0.1	<0.1	<0.1	37.0	Ethiopie	2000	♀
3	<0.1	<0.1	<0.1	<0.1	Guinée	1999	♀
4	<0.1	<0.1	0.1	<0.1	Malawi	2000	♀
5	0.1	0.1	0.1	<0.1	Tchad	1997	♀
6	0.1	0.1	0.1	43.9	Côte d'Ivoire	1994	♀
7	0.1	0.1	0.1	<0.1	Malawi	1992	♀
8	0.1	<0.1	0.1	<0.1	Malawi	1992	♂
9	0.1	0.1	0.1	1.5	Togo	1998	♀
10	0.1	0.1	<0.1	<0.1	Tanzanie	2004	♀
11	0.1	0.1	0.1	<0.1	Zambie	2002	♀
12	0.1	0.1	0.1	<0.1	Zimbabwe	1994	♀
13	0.2	0.2	0.2	<0.1	Bénin	1996	♀
14	0.2	0.2	0.2	<0.1	Kenya	2009	♀
15	0.2	0.3	0.2	2.0	Malawi	2004	♀
16	0.2	0.2	0.1		Mali	1996	♀
17	0.2	0.3	0.2	<0.1	Ouganda	2006	♀
18	0.2	0.2	0.1	<0.1	Zimbabwe	1994	♂
19	0.3	0.3	0.4	0.5	Bénin	2006	♀
20	0.3	0.4	0.3		Congo	2005	♂
21	0.3	0.3	0.4	<0.1	Mozambique	2003	♀
22	0.3	0.2	0.4	<0.1	Rwanda	2000	♀
23	0.3	0.3	0.3	<0.1	Sénégal	1993	♀
24	0.4	0.5	0.4	16.9	Burkina Faso	2003	♀
25	0.4	0.4	0.4	0.7	Cameroun	1998	♀

26	0.4	0.4	0.4	<0.1	Ethiopie	2005	♀
27	0.4	0.4	0.4	<0.1	Niger	2006	♀
28	0.5	0.5	0.5	<0.1	Congo RDC	2007	♀
29	0.5	0.5	0.5		Guinée	2005	♀
30	0.5	0.5	0.5	<0.1	Madagascar	2009	♀
31	0.5	0.5	0.4	16.1	Swaziland	2006	♀
32	0.5	0.5	0.4	<0.1	Tanzanie	1996	♀
33	0.5	0.4	0.5	<0.1	Zambie	1996	♀
34	0.5	0.5	0.5	<0.1	Zambie	2007	♂
35	0.6	0.6	0.6		Cameroun	2004	♀
36	0.6	0.6	0.6		Congo	2005	♀
37	0.6	0.6	0.7	<0.1	Sénégal	2005	♀
38	0.6	0.7	0.6	<0.1	Tanzanie	1996	♂
39	0.6	0.7	0.6	<0.1	Zambie	2007	♀
40	0.6	0.6	0.6	<0.1	Zimbabwe	2006	♀
41	0.7	0.6	0.7	12.2	Tchad	2004	♀
42	0.7	0.7	0.6	<0.1	Lesotho	2004	♀
43	0.8	0.7	0.9		Cameroun	2004	♂
44	0.8	0.7	0.7	32.7	Ouganda	2001	♀
45	0.8	0.7	0.8	15.4	Zimbabwe	2006	♂
46	0.9	0.9	0.8	7.9	Mali	2006	♀
47	0.9	1.0	0.8	<0.1	Rwanda	2005	♀
48	0.9	1.0	0.8	<0.1	Zimbabwe	1999	♀
49	1.0	1.0	0.9	22.2	Kenya	1998	♀
50	1.0	1.0	1.0	4.3	Nigeria	2008	♂
51	1.1	1.1	1.1	1.7	Nigeria	2008	♀
52	1.1	1.0	1.1	2.6	Ouganda	1995	♀
53	1.4	1.3	1.5	<0.1	Ouganda	1995	♂
54	1.5	1.4	1.5	25.9	Burkina Faso	1999	♀
55	1.5	1.6	1.4	54.9	Kenya	2003	♀
56	1.7	1.6	1.3	76.5	Gabon	2000	♀
57	2.2	2.3	2.1	4.7	Madagascar	2004	♀
58	2.3	2.2	2.5	<0.1	Nigeria	1999	♂
59	2.7	2.7	2.6	<0.1	Nigeria	1999	♀
60	2.9	3.1	2.7	1.2	Mali	2001	♀
61	3.1	3.0	3.3	<0.1	Niger	1992	♀
62	3.9	3.9	3.9	8.8	Sierra Leone	2008	♀
63	4.0	3.9	4.1	6.0	Namibie	1992	♀
64	4.7	4.8	4.7	2.5	Afrique du Sud	1998	♀
65	6.6	6.3	6.8	23.6	Madagascar	1997	♀
66	6.6	6.5	6.6	34.8	Namibie	2000	♀
67	8.0	8.0	7.9	94.4	Mozambique	1997	♀
68	9.6	9.5	9.7	66.8	Madagascar	1992	♀

TABLE A.22 – *Proportion de frères et soeurs survivants dont l'âge à l'enquête est inconnu - par sexe (en %) - enquêtes EDS*

Données manquantes sur les âges au décès et/ou les années écoulées depuis le décès

Enquête		Tot.	Frères	Soeurs	NSP
Malawi 1992 ♂	Age au décès	0.10	0.10	0.10	0.00
	Années depuis le décès	0.00	0.00	0.00	0.00
	Age + Années	0.10	0.20	0.00	0.00
	Total	0.20	0.30	0.10	0.00
Centrafique 1995 ♀	Age au décès	0.30	0.40	0.20	12.50
	Années depuis le décès	0.10	0.10	0.20	57.60
	Age + Années	0.00	0.00	0.00	100.00
	Total	0.50	0.50	0.40	35.10

Ethiopie 2000 ♀	Age au décès	0.10	0.10	0.10	0.00
	Années depuis le décès	0.00	0.00	0.00	0.00
	Age + Années	0.50	0.30	0.20	93.10
	Total	0.70	0.40	0.40	93.10
Zambie 2002 ♀	Age au décès	0.70	0.70	0.60	0.00
	Années depuis le décès	0.00	0.00	0.10	0.00
	Age + Années	0.20	0.20	0.10	23.10
	Total	0.90	0.90	0.80	23.10
Malawi 1992 ♀	Age au décès	0.30	0.40	0.30	0.00
	Années depuis le décès	0.40	0.30	0.40	1.80
	Age + Années	0.30	0.20	0.10	16.50
	Total	1.00	0.80	0.80	18.40
Zimbabwe 1994 ♀	Age au décès	0.60	0.70	0.40	0.00
	Années depuis le décès	0.30	0.40	0.20	0.00
	Age + Années	0.10	0.10	0.10	2.40
	Total	1.00	1.20	0.80	2.40
Tanzanie 2004 ♀	Age au décès	0.30	0.30	0.30	0.00
	Années depuis le décès	0.20	0.20	0.20	0.00
	Age + Années	0.80	0.90	0.60	24.10
	Total	1.30	1.40	1.10	24.10
Zimbabwe 1994 ♂	Age au décès	0.50	0.70	0.40	0.00
	Années depuis le décès	0.10	0.00	0.20	0.00
	Age + Années	0.60	0.20	0.30	48.30
	Total	1.30	0.90	0.90	48.30
Tchad 2004 ♀	Age au décès	0.40	0.40	0.40	35.80
	Années depuis le décès	0.80	0.90	0.80	33.20
	Age + Années	0.10	0.10	0.20	95.40
	Total	1.40	1.40	1.40	20.40
Sénégal 1993 ♀	Age au décès	0.30	0.30	0.20	0.00
	Années depuis le décès	0.40	0.40	0.30	0.00
	Age + Années	0.80	1.00	0.50	13.60
	Total	1.40	1.80	0.90	13.60
Niger 2006 ♀	Age au décès	0.50	0.50	0.50	0.00
	Années depuis le décès	0.60	0.60	0.70	6.40
	Age + Années	0.60	0.80	0.30	0.00
	Total	1.70	1.90	1.50	6.40
Malawi 2004 ♀	Age au décès	0.70	0.60	0.80	0.00
	Années depuis le décès	0.40	0.40	0.30	0.00
	Age + Années	0.70	0.90	0.50	0.00
	Total	1.80	2.00	1.70	0.00
Guinée 2005 ♀	Age au décès	0.80	0.80	0.80	0.00
	Années depuis le décès	0.70	0.70	0.60	0.00
	Age + Années	0.50	0.60	0.30	0.00
	Total	1.90	2.00	1.70	0.00
Ouganda 2006 ♀	Age au décès	0.80	1.00	0.60	0.00
	Années depuis le décès	0.30	0.30	0.40	0.00
	Age + Années	0.70	0.70	0.50	50.70
	Total	1.90	2.00	1.40	50.70
Zambie 2007 ♂	Age au décès	0.70	0.60	0.90	0.00
	Années depuis le décès	0.20	0.30	0.10	0.00
	Age + Années	1.00	1.20	0.80	0.00
	Total	1.90	2.10	1.80	0.00
Zambie 2007 ♀	Age au décès	0.50	0.60	0.50	0.00
	Années depuis le décès	0.20	0.20	0.10	0.00
	Age + Années	1.20	1.50	0.90	58.00
	Total	2.00	2.30	1.50	58.00
Ethiopie 2005 ♀	Age au décès	0.80	0.80	0.70	0.00
	Années depuis le décès	0.10	0.10	0.10	11.50
	Age + Années	1.20	1.20	1.10	25.60
	Total	2.10	2.10	1.90	37.10

Rwanda 2000 ♀	Age au décès	0.60	0.60	0.60	2.40
	Années depuis le décès	0.70	0.70	0.60	10.40
	Age + Années	1.10	0.90	0.80	54.50
	Total	2.40	2.20	1.90	67.30
Zimbabwe 2006 ♂	Age au décès	1.40	1.30	1.60	0.00
	Années depuis le décès	0.20	0.20	0.20	0.00
	Age + Années	0.90	0.60	0.90	62.60
	Total	2.50	2.10	2.60	62.60
Zimbabwe 2006 ♀	Age au décès	1.40	1.40	1.30	0.00
	Années depuis le décès	0.50	0.50	0.40	16.80
	Age + Années	0.70	0.70	0.70	0.00
	Total	2.50	2.60	2.40	16.80
Congo 2005 ♀	Age au décès	0.50	0.30	0.70	0.00
	Années depuis le décès	0.60	0.70	0.40	0.00
	Age + Années	1.60	1.50	1.70	0.00
	Total	2.70	2.50	2.80	0.00
Côte d'Ivoire 1994 ♀	Age au décès	0.20	0.30	0.20	21.50
	Années depuis le décès	1.20	1.30	1.10	8.70
	Age + Années	1.30	1.90	0.70	0.00
	Total	2.80	3.50	2.00	0.00
Bénin 2006 ♀	Age au décès	0.90	0.90	0.90	3.30
	Années depuis le décès	1.40	1.30	1.50	0.00
	Age + Années	0.70	0.60	0.50	22.60
	Total	2.90	2.90	2.80	25.90
Madagascar 2009 ♀	Age au décès	1.40	1.40	1.40	0.00
	Années depuis le décès	0.40	0.50	0.30	58.80
	Age + Années	1.20	1.10	1.20	15.40
	Total	3.00	3.10	2.80	74.20
Cameroun 2004 ♀	Age au décès	0.40	0.40	0.40	0.00
	Années depuis le décès	1.50	1.60	1.20	0.00
	Age + Années	1.30	1.30	1.20	100.00
	Total	3.20	3.40	2.80	100.00
Congo 2005 ♂	Age au décès	0.20	0.20	0.20	
	Années depuis le décès	0.70	0.60	0.90	
	Age + Années	2.40	3.10	1.70	
	Total	3.40	3.90	2.80	
Cameroun 2004 ♂	Age au décès	0.50	0.60	0.30	
	Années depuis le décès	1.00	1.20	0.90	
	Age + Années	1.90	1.80	2.00	
	Total	3.40	3.60	3.20	
Kenya 2009 ♀	Age au décès	1.20	1.30	1.10	0.40
	Années depuis le décès	0.50	0.50	0.60	0.00
	Age + Années	2.00	1.40	1.70	81.00
	Total	3.70	3.20	3.40	81.40
Congo RDC 2007 ♀	Age au décès	1.30	1.70	0.90	0.00
	Années depuis le décès	1.30	1.50	1.00	0.00
	Age + Années	1.10	1.20	1.10	0.00
	Total	3.80	4.40	3.10	0.00
Mali 2006 ♀	Age au décès	0.90	1.00	0.90	0.00
	Années depuis le décès	0.90	0.90	1.00	0.00
	Age + Années	2.00	2.30	1.40	16.20
	Total	3.80	4.20	3.30	16.20
Sénégal 2005 ♀	Age au décès	1.20	1.20	1.10	10.60
	Années depuis le décès	1.30	1.10	1.20	14.80
	Age + Années	1.80	1.70	1.60	23.20
	Total	4.20	4.10	3.90	48.60
Swaziland 2006 ♀	Age au décès	2.00	2.30	1.60	14.30
	Années depuis le décès	0.90	1.10	0.70	0.00
	Age + Années	1.40	1.50	1.30	0.00
	Total	4.30	4.90	3.60	14.30

Burkina Faso 2003 ♀	Age au décès	0.20	0.20	0.20	0.30
	Années depuis le décès	0.70	0.80	0.50	2.30
	Age + Années	4.20	2.80	2.60	79.10
	Total	5.10	3.80	3.30	81.60
Rwanda 2005 ♀	Age au décès	0.90	0.90	0.80	0.00
	Années depuis le décès	3.20	3.00	3.40	14.30
	Age + Années	1.10	1.10	1.00	30.60
	Total	5.20	5.10	5.20	44.90
Lesotho 2004 ♀	Age au décès	2.20	2.40	1.90	5.00
	Années depuis le décès	1.10	0.90	1.40	0.00
	Age + Années	2.00	1.70	2.00	39.20
	Total	5.30	5.00	5.30	44.20
Nigeria 2008 ♂	Age au décès	1.60	1.60	1.50	7.40
	Années depuis le décès	1.00	1.00	0.80	11.70
	Age + Années	2.80	3.10	2.00	25.70
	Total	5.30	5.70	4.40	44.90
Nigeria 2008 ♀	Age au décès	2.30	2.50	2.20	1.80
	Années depuis le décès	1.00	1.00	0.80	5.70
	Age + Années	2.10	2.00	1.80	31.70
	Total	5.40	5.50	4.80	39.30
Kenya 1998 ♀	Age au décès	0.70	0.90	0.50	0.00
	Années depuis le décès	2.00	1.70	1.60	32.90
	Age + Années	3.00	2.90	2.00	43.80
	Total	5.70	5.50	4.10	76.70
Madagascar 2004 ♀	Age au décès	3.30	3.00	3.70	3.10
	Années depuis le décès	0.90	1.00	0.50	34.10
	Age + Années	1.60	1.00	2.10	10.20
	Total	5.80	5.00	6.30	47.30
Gabon 2000 ♀	Age au décès	0.60	0.90	0.30	0.00
	Années depuis le décès	3.10	2.20	3.70	10.30
	Age + Années	2.40	2.00	1.70	29.60
	Total	6.10	5.10	5.80	39.90
Kenya 2003 ♀	Age au décès	1.00	1.30	0.60	2.60
	Années depuis le décès	3.20	3.30	2.70	14.90
	Age + Années	2.80	2.70	1.90	54.30
	Total	7.00	7.30	5.20	71.70
Mali 2001 ♀	Age au décès	1.60	1.60	1.70	4.50
	Années depuis le décès	1.70	1.90	1.40	6.60
	Age + Années	3.80	4.00	3.60	2.20
	Total	7.20	7.50	6.70	13.30
Niger 1992 ♀	Age au décès	0.80	0.90	0.70	0.00
	Années depuis le décès	4.70	4.00	5.50	0.00
	Age + Années	5.40	5.80	3.80	93.70
	Total	10.90	10.70	10.10	93.70
Sierra Leone 2008 ♀	Age au décès	5.40	6.20	4.30	10.60
	Années depuis le décès	2.30	2.10	2.60	0.00
	Age + Années	5.80	6.30	5.20	20.70
	Total	13.50	14.60	12.10	31.30
Namibie 2000 ♀	Age au décès	4.60	4.70	4.50	5.50
	Années depuis le décès	5.60	5.70	5.50	0.00
	Age + Années	12.60	13.20	11.50	57.60
	Total	22.80	23.60	21.50	63.00
Namibie 1992 ♀	Age au décès	3.90	4.80	2.90	0.00
	Années depuis le décès	9.10	9.30	8.50	18.30
	Age + Années	11.70	11.80	11.10	30.10
	Total	24.80	25.90	22.50	48.40
Madagascar 1992 ♀	Age au décès	2.40	2.50	2.20	9.70
	Années depuis le décès	22.30	22.20	22.40	1.10
	Age + Années	9.00	9.70	8.10	0.00
	Total	33.60	34.40	32.70	0.00

Burkina Faso 1999 ♀	Age au décès	0.80	0.80	0.80	0.20
	Années depuis le décès	28.20	28.60	29.60	9.50
	Age + Années	11.90	9.10	8.40	80.20
	Total	40.90	38.50	38.80	90.00
Bénin 1996 ♀	Age au décès	0.30	0.10	0.20	4.80
	Années depuis le décès	48.40	48.60	48.30	45.80
	Age + Années	1.70	1.20	1.10	24.40
	Total	50.30	49.90	49.60	75.10
Mali 1996 ♀	Age au décès	0.10	0.10	0.00	0.00
	Années depuis le décès	51.70	51.00	52.40	79.10
	Age + Années	0.20	0.40	0.00	0.00
	Total	52.00	51.50	52.40	79.10
Togo 1998 ♀	Age au décès	0.10	0.10	0.00	0.00
	Années depuis le décès	52.50	53.10	52.00	39.00
	Age + Années	0.40	0.40	0.20	11.10
	Total	53.00	53.60	52.30	50.10
Cameroun 1998 ♀	Age au décès	0.10	0.10	0.10	0.00
	Années depuis le décès	63.90	63.40	64.10	68.80
	Age + Années	1.20	1.00	0.70	12.10
	Total	65.20	64.50	64.90	80.90
Tchad 1997 ♀	Age au décès	0.10	0.00	0.10	0.00
	Années depuis le décès	83.50	83.80	83.80	45.30
	Age + Années	0.90	0.70	0.40	51.30
	Total	84.50	84.50	84.30	96.70
Ouganda 2001 ♀	Age au décès	0.30	0.30	0.30	0.00
	Années depuis le décès	77.50	77.20	78.70	30.90
	Age + Années	7.50	7.90	5.90	63.70
	Total	85.20	85.40	84.90	94.60
Guinée 1999 ♀	Age au décès	0.00	0.00	0.00	0.00
	Années depuis le décès	85.40	85.60	85.30	54.10
	Age + Années	0.00	0.00	0.00	0.00
	Total	85.50	85.70	85.30	54.10
Madagascar 1997 ♀	Age au décès	1.60	1.70	1.60	0.00
	Années depuis le décès	72.10	73.10	73.00	13.80
	Age + Années	12.20	11.20	10.90	82.40
	Total	85.90	86.00	85.50	96.10
Mozambique 1997 ♀	Age au décès	1.90	1.70	2.10	62.30
	Années depuis le décès	62.60	59.90	65.50	18.90
	Age + Années	22.20	23.70	20.40	0.00
	Total	86.60	85.40	88.00	0.00
Mozambique 2003 ♀	Age au décès	0.20	0.30	0.10	0.00
	Années depuis le décès	87.20	87.40	87.20	66.20
	Age + Années	3.70	3.90	3.30	19.00
	Total	91.10	91.50	90.70	85.20
Ouganda 1995 ♂	Age au décès	0.10	0.00	0.10	0.00
	Années depuis le décès	84.70	84.20	85.60	77.80
	Age + Années	6.60	6.60	6.50	22.20
	Total	91.40	90.80	92.10	100.00
Ouganda 1995 ♀	Age au décès	0.10	0.10	0.20	0.00
	Années depuis le décès	87.40	88.50	87.90	21.70
	Age + Années	5.80	5.00	5.20	73.00
	Total	93.40	93.50	93.30	94.70
Afrique du Sud 1998 ♀	Age au décès	1.50	1.60	1.30	0.60
	Années depuis le décès	69.60	70.70	70.70	33.70
	Age + Années	22.30	21.20	20.90	65.10
	Total	93.40	93.50	92.90	99.40
Nigeria 1999 ♀	Age au décès	0.90	0.80	0.90	3.00
	Années depuis le décès	66.70	66.20	68.20	13.40
	Age + Années	26.40	27.40	24.30	83.60
	Total	94.00	94.40	93.30	100.00

Zimbabwe 1999 ♀	Age au décès	0.40	0.40	0.30	0.00
	Années depuis le décès	89.50	88.30	91.20	48.40
	Age + Années	4.50	4.80	3.80	42.00
	Total	94.30	93.60	95.20	90.40
Tanzanie 1996 ♂	Age au décès	0.10	0.20	0.00	0.00
	Années depuis le décès	92.30	92.20	92.40	100.00
	Age + Années	2.20	2.40	2.10	0.00
	Total	94.70	94.70	94.50	100.00
Zambie 1996 ♀	Age au décès	0.00	0.00	0.10	0.00
	Années depuis le décès	91.60	91.00	92.20	87.10
	Age + Années	3.10	3.30	2.90	5.10
	Total	94.70	94.40	95.20	92.10
Malawi 2000 ♀	Age au décès	0.00	0.00	0.00	0.00
	Années depuis le décès	95.50	95.40	95.60	91.30
	Age + Années	0.20	0.20	0.20	0.00
	Total	95.70	95.60	95.80	91.30
Tanzanie 1996 ♀	Age au décès	0.10	0.10	0.20	0.00
	Années depuis le décès	92.60	92.10	93.40	79.10
	Age + Années	3.40	3.80	2.90	15.50
	Total	96.20	96.00	96.50	94.50
Nigeria 1999 ♂	Age au décès	1.00	1.50	0.40	0.00
	Années depuis le décès	79.00	77.20	81.90	48.20
	Age + Années	16.80	18.70	13.70	51.80
	Total	96.80	97.40	96.00	100.00

TABLE A.23 – *Proportion de frères et soeurs décédés dont l'âge au décès et/ou le nombre d'années écoulées depuis le décès est inconnu (en %) - enquêtes EDS*

Quelle est la véritable proportion de réponses manquantes sur la datation des décès des frères et sœurs ?

Dans l'enquête menée au Malawi en 2000, nous estimons que pour la *quasi-totalité* des décédés, le temps écoulé depuis le décès est une information manquante (95.8%). Or dans le rapport de cette enquête, c'est l'inverse ; les auteurs indiquent que cette information est manquante pour moins d'1% des frères et sœurs décédés. Comment cela s'explique-t-il ?

Parcourir visuellement les fichiers d'enquête nous permet de penser que **la proportion de réponses manquantes sur ces variables est bien plus élevée que ce qui est publié dans les rapports d'enquête**. Nous ne pouvons pas affirmer avec certitude avoir trouvé la source de ces divergences [45], mais nous proposons ici une explication plausible : les rapports d'enquête calculeraient en fait les proportions de réponses manquantes en tenant compte uniquement des réponses *codées comme manquantes* (c'est-à-dire codées comme 98). Or il y a un très grand nombre de frères et sœurs décédés qui apparaissent dans les fichiers, mais pour lesquels *l'âge au décès ou la durée écoulée depuis sont manquantes sans être codées comme tel*. Elles ne sont pas codées (" " au lieu de "98").

Prenons encore le cas du Malawi en 2000. D'après nos calculs, sur 22280 frères et sœurs décédés (données pondérées), seuls 5 frères et sœurs ont des données manquantes sur la datation de leurs décès *qui sont codées comme telles* (tableau A.24). Dès lors, pour plus de 99% des décédés, il semble que les données soient complètes, et c'est ce qui est publié dans

45. Parce que nous ne parvenons pas à retomber exactement sur les chiffres publiés dans les rapports, notamment en raison des choix de pondérations à appliquer aux frères et sœurs.

le rapport. Mais si on intégre les données manquantes *qui ne sont pas codées comme telles* , les choses changent radicalement : pour seuls 4% des décédés, l'âge au décès et la datation de ce décès sont enregistrés (tableau A.25). Il nous semble donc que *les rapports d'enquêtes surestiment la complétude des données sur les frères et sœurs décédés*.

		Age au décès	
		Non manquant	Manquant
Années depuis le décès	Non Manquant	22275	0
	Manquant	0	5

TABLE A.24 – *Nombre de décédés pour lesquels des données manquent sur la datation des décès (à partir des observations **codées comme 98**) - enquête EDS Malawi 2000*

		Age au décès	
		Non Manquant	Manquant
Années depuis le décès	Non Manquant	961	5
	Manquant	21267	48

TABLE A.25 – *Nombre de décédés pour lesquels des données manquent sur la datation des décès (à partir des observations **manquantes ou codées comme 98**) - enquête EDS Malawi 2000*

A.7 Annexes au chapitre 7

A.7.1 Bénin

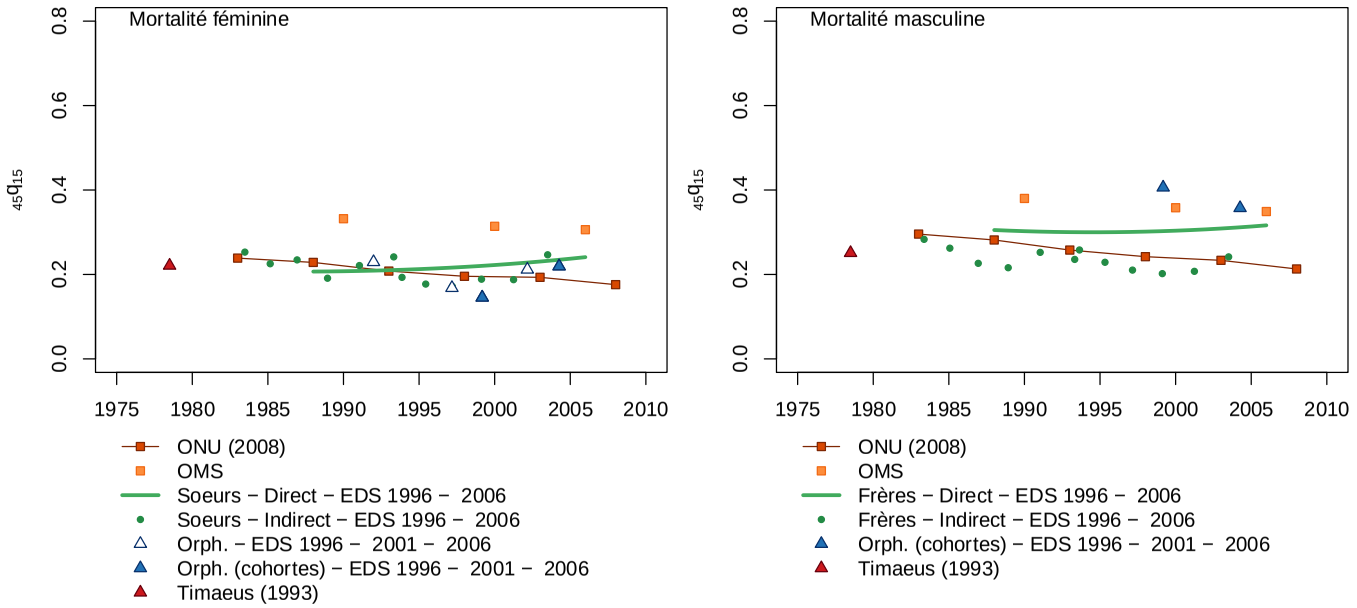

Figure A.26 – *Bénin : Estimations du quotient $_{45}q_{15}$ à partir de la survie des frères, des soeurs et des parents / estimations UN et OMS*

A.7.2 Burkina Faso

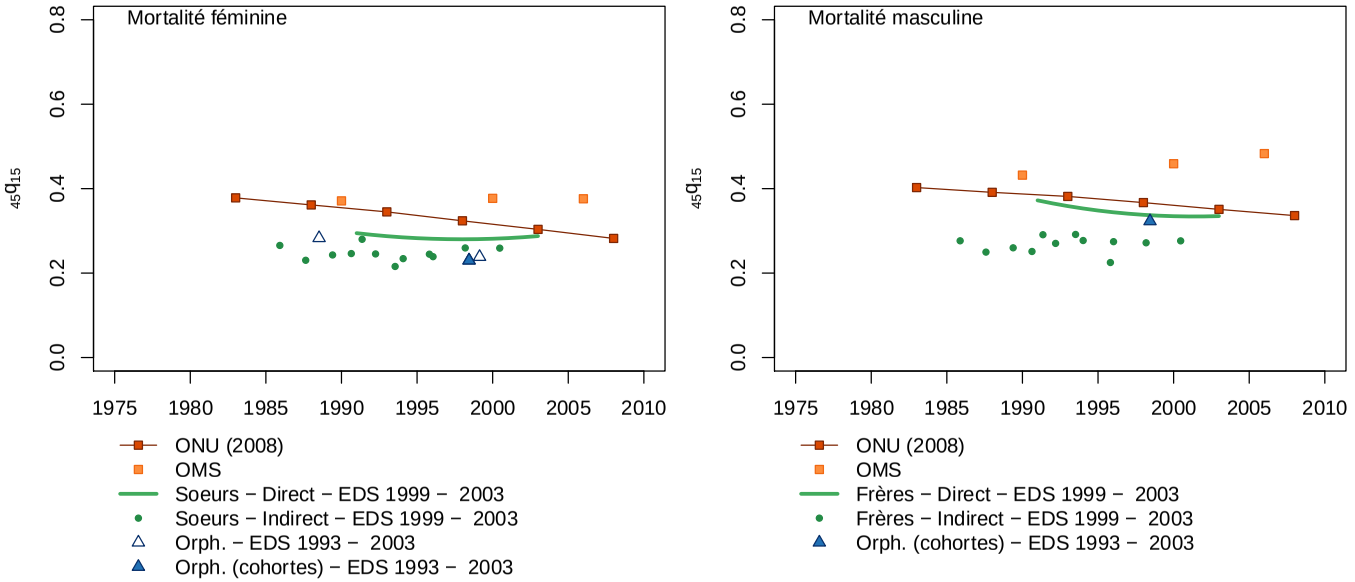

FIGURE A.27 – *Burkina Faso : Estimations du quotient ${}_{45}q_{15}$ à partir de la survie des frères, des soeurs et des parents / estimations UN et OMS*

A.7.3 Centrafique

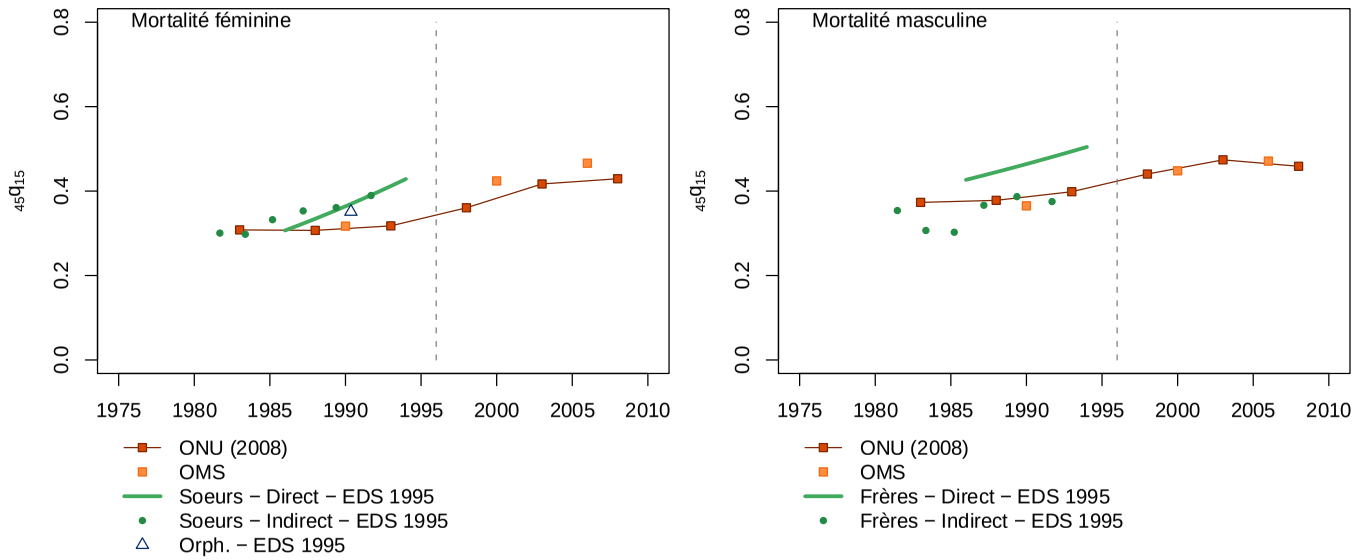

FIGURE A.28 – *Centrafique : Estimations du quotient $_{45}q_{15}$ à partir de la survie des frères, des soeurs et des parents / estimations UN et OMS*

A.7.4 Comores

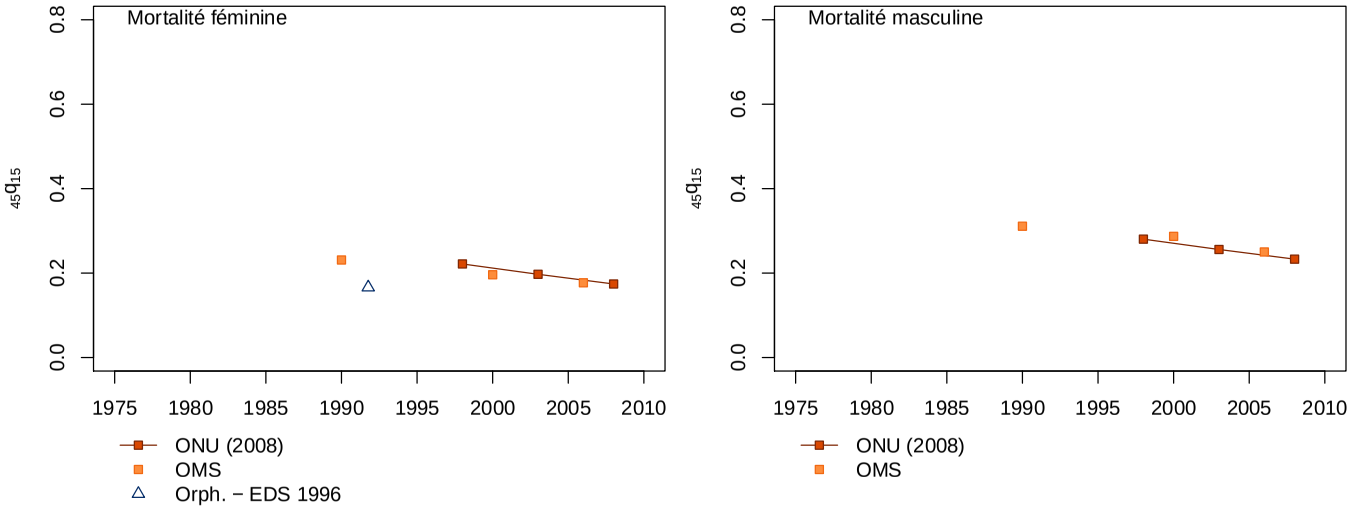

FIGURE A.29 – *Comores : Estimations du quotient $_{45}q_{15}$ à partir de la survie des frères, des soeurs et des parents / estimations UN et OMS*

A.7.5 Congo

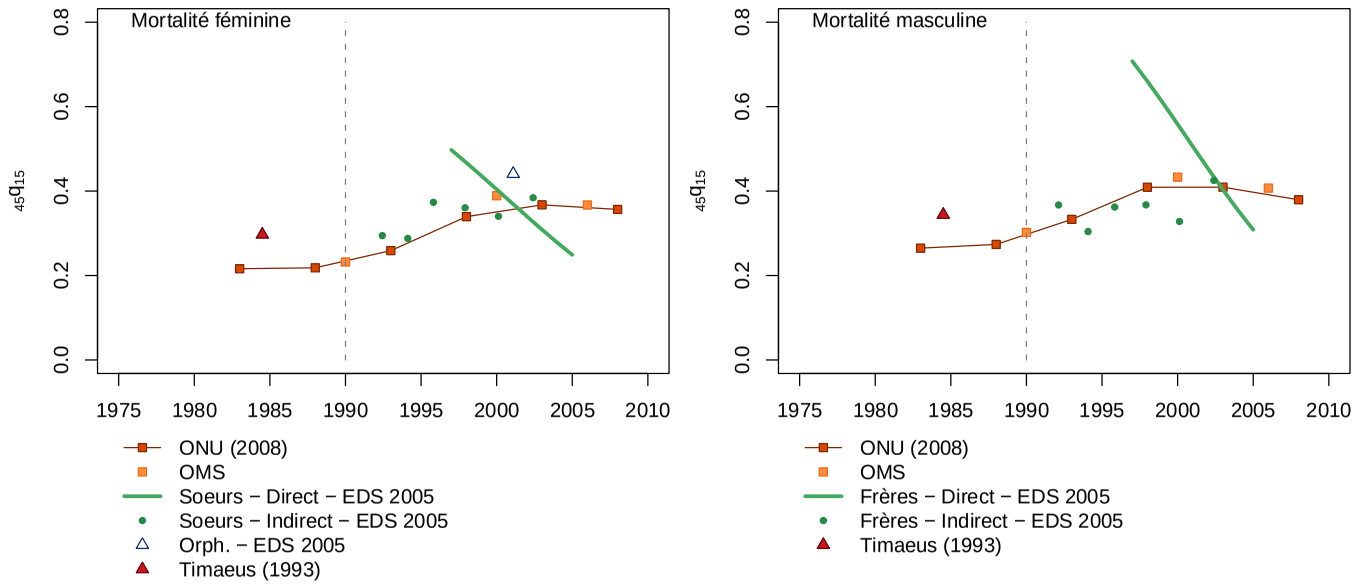

FIGURE A.30 – *Congo : Estimations du quotient $_{45}q_{15}$ à partir de la survie des frères, des soeurs et des parents / estimations UN et OMS*

A.7.6 Congo RDC

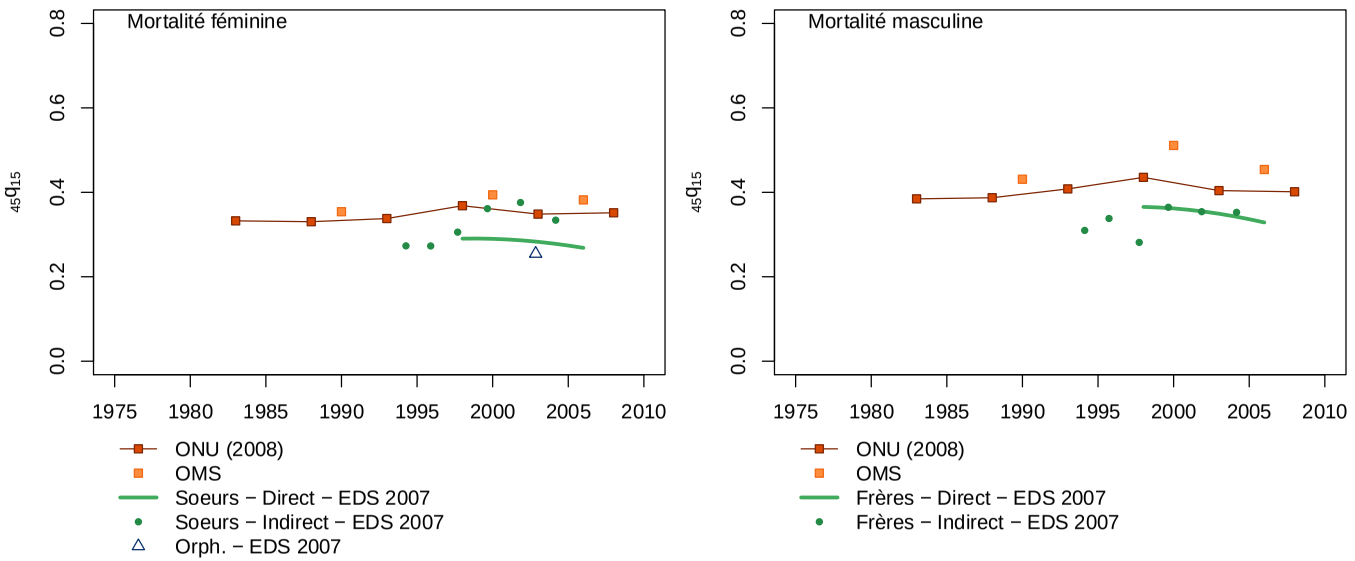

FIGURE A.31 – *Congo RDC : Estimations du quotient $_{45}q_{15}$ à partir de la survie des frères, des soeurs et des parents / estimations UN et OMS*

A.7.7 Côte d'Ivoire

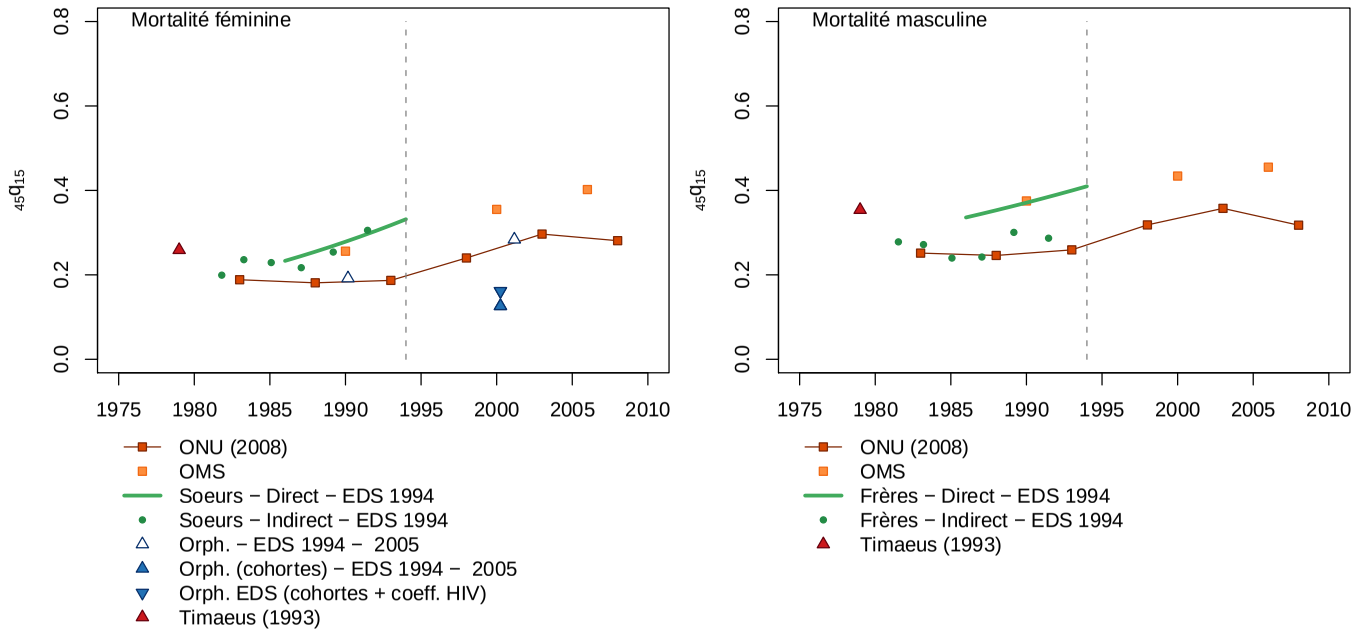

FIGURE A.32 – *Côte d'Ivoire : Estimations du quotient $_{45}q_{15}$ à partir de la survie des frères, des soeurs et des parents / estimations UN et OMS*

A.7.8 Ethiopie

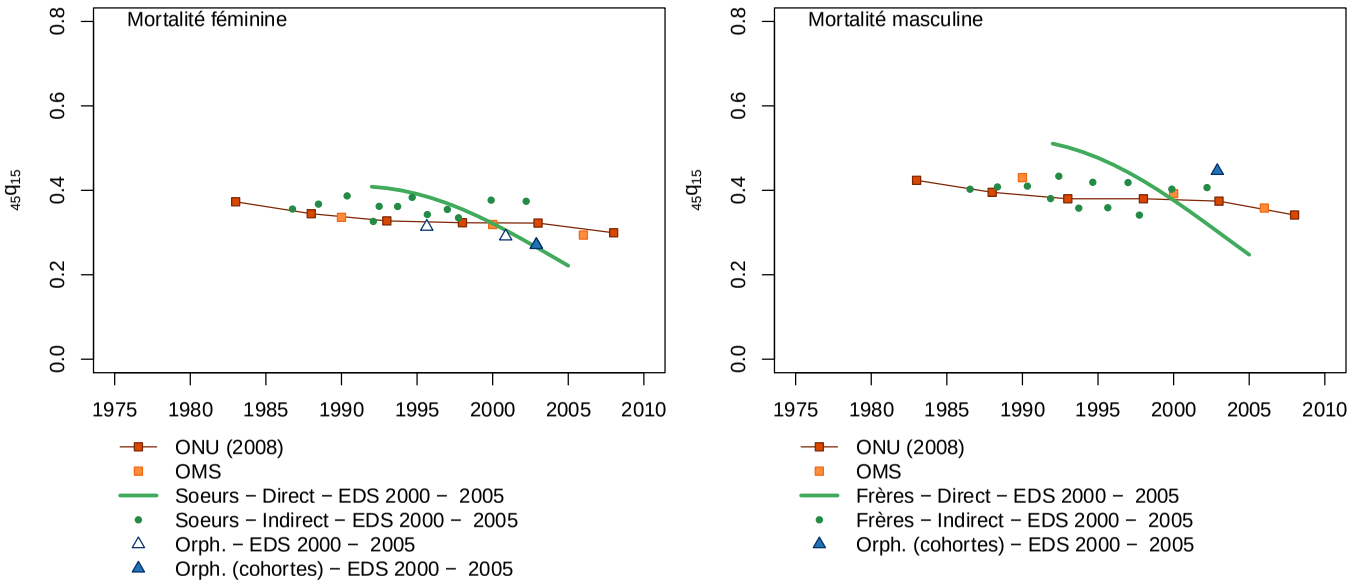

FIGURE A.33 – *Ethiopie : Estimations du quotient $_{45}q_{15}$ à partir de la survie des frères, des soeurs et des parents / estimations UN et OMS*

A.7.9 Gabon

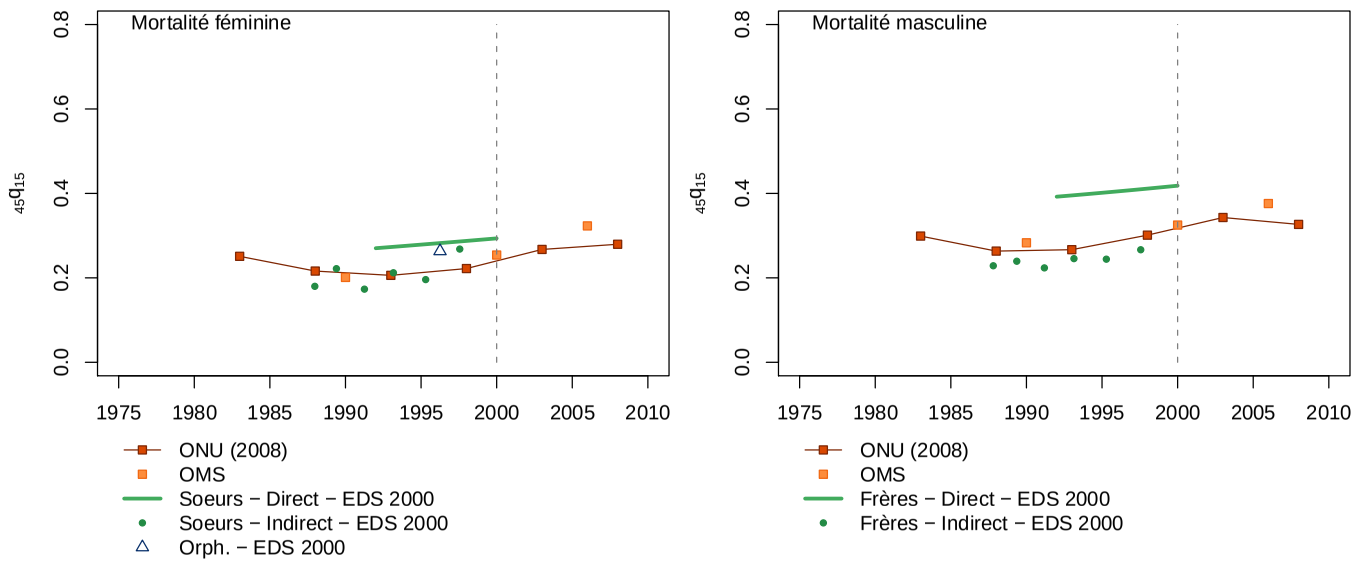

FIGURE A.34 – *Gabon : Estimations du quotient $_{45}q_{15}$ à partir de la survie des frères, des soeurs et des parents / estimations UN et OMS*

A.7.10 Ghana

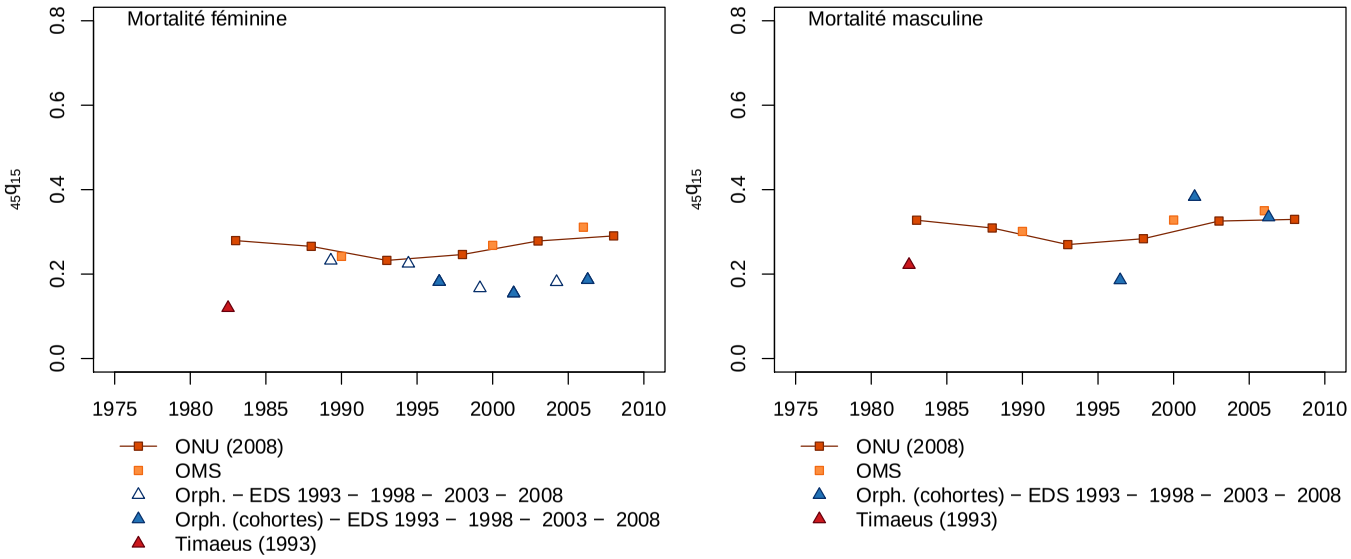

FIGURE A.35 – *Ghana : Estimations du quotient $_{45}q_{15}$ à partir de la survie des frères, des soeurs et des parents / estimations UN et OMS*

A.7.11 Guinée

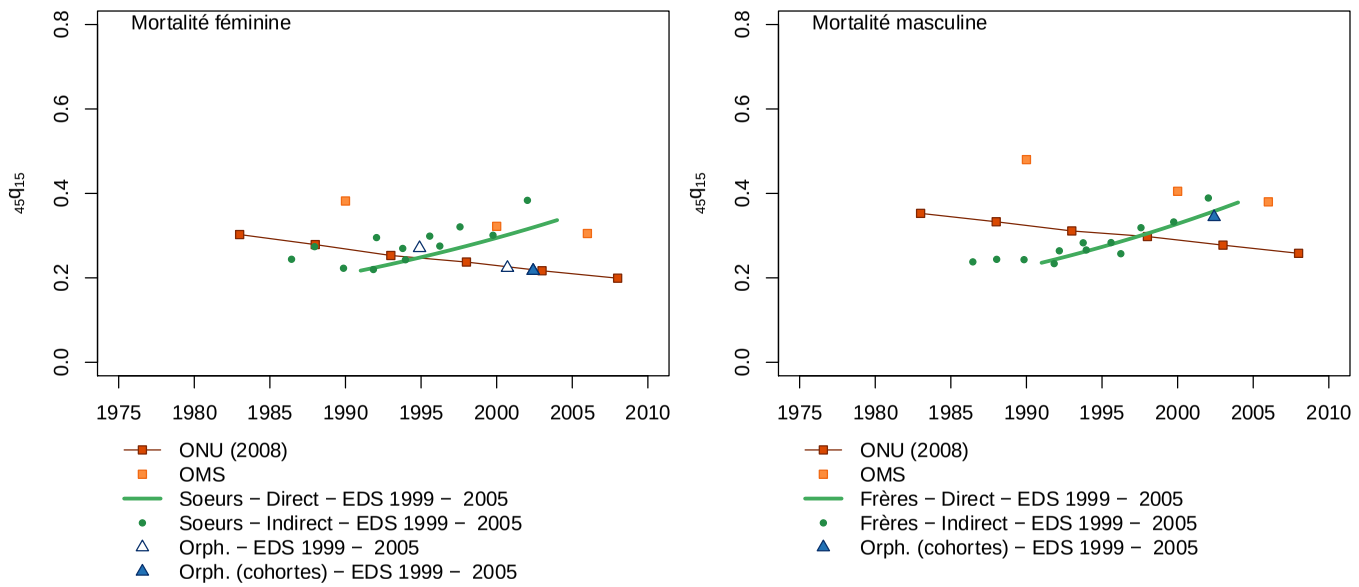

FIGURE A.36 – *Guinée : Estimations du quotient ${}_{45}q_{15}$ à partir de la survie des frères, des soeurs et des parents / estimations UN et OMS*

A.7.12 Kenya

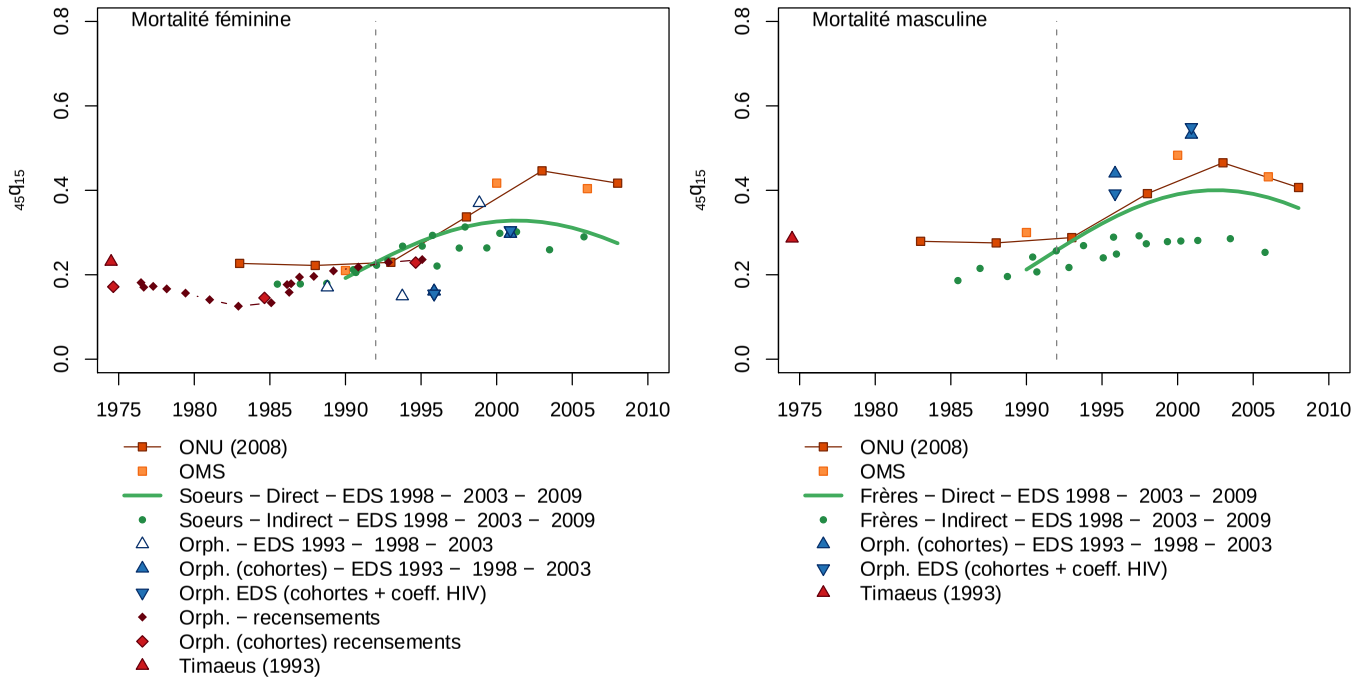

FIGURE A.37 – *Kenya : Estimations du quotient $_{45}q_{15}$ à partir de la survie des frères, des soeurs et des parents / estimations UN et OMS*

A.7.13 Lesotho

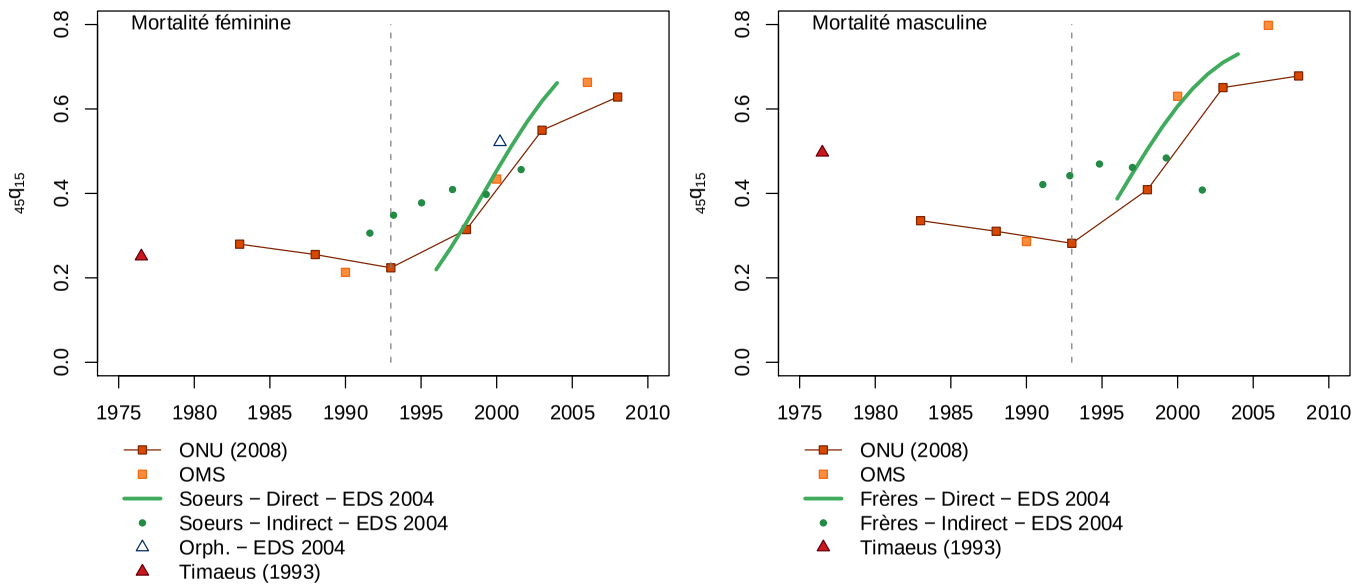

FIGURE A.38 – *Lesotho : Estimations du quotient $_{45}q_{15}$ à partir de la survie des frères, des soeurs et des parents / estimations UN et OMS*

A.7.14 Liberia

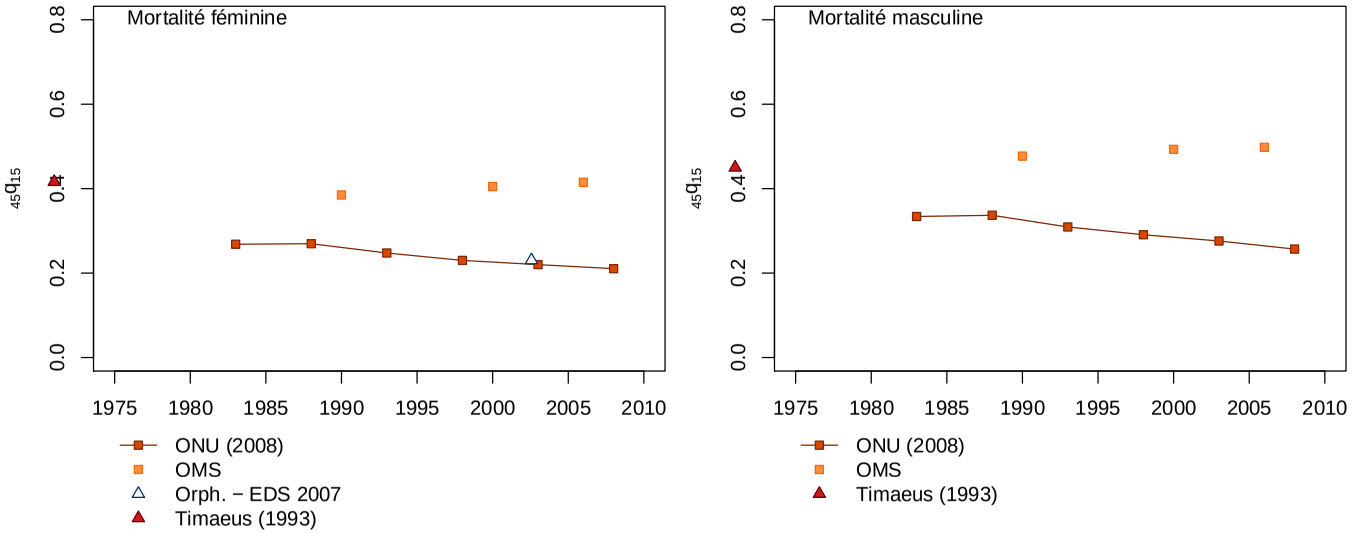

FIGURE A.39 – *Liberia : Estimations du quotient $_{45}q_{15}$ à partir de la survie des frères, des soeurs et des parents / estimations UN et OMS*

A.7.15 Mozambique

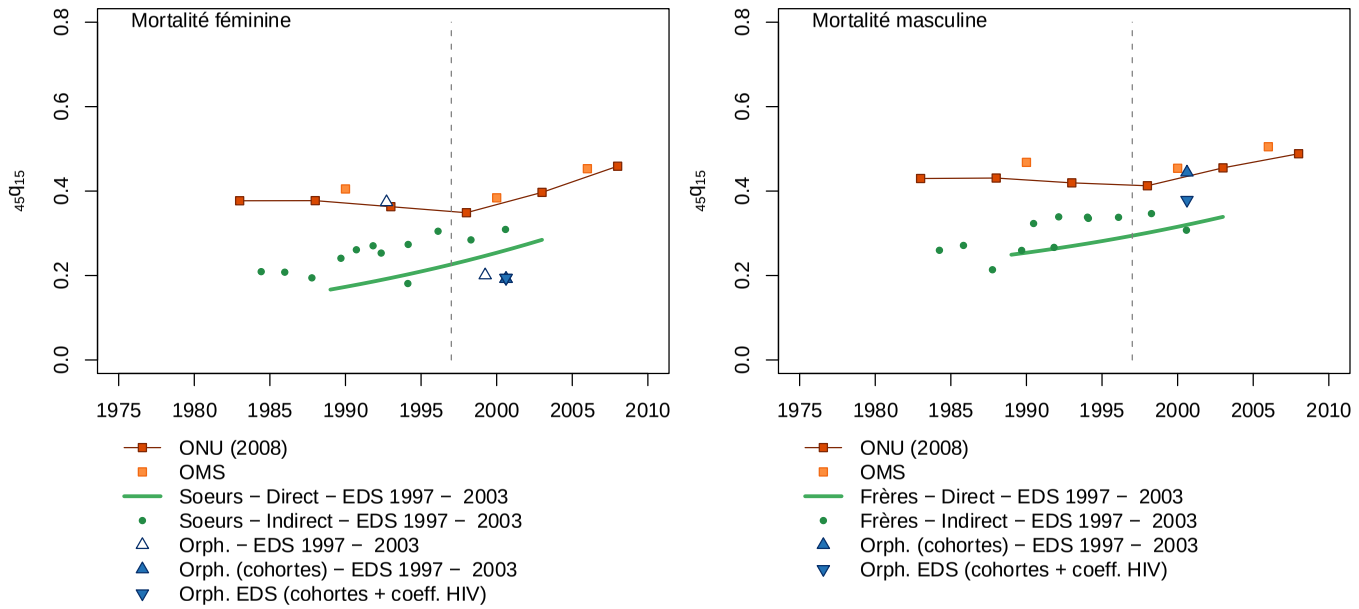

FIGURE A.40 – *Mozambique : Estimations du quotient ${}_{45}q_{15}$ à partir de la survie des frères, des soeurs et des parents / estimations UN et OMS*

A.7.16 Namibie

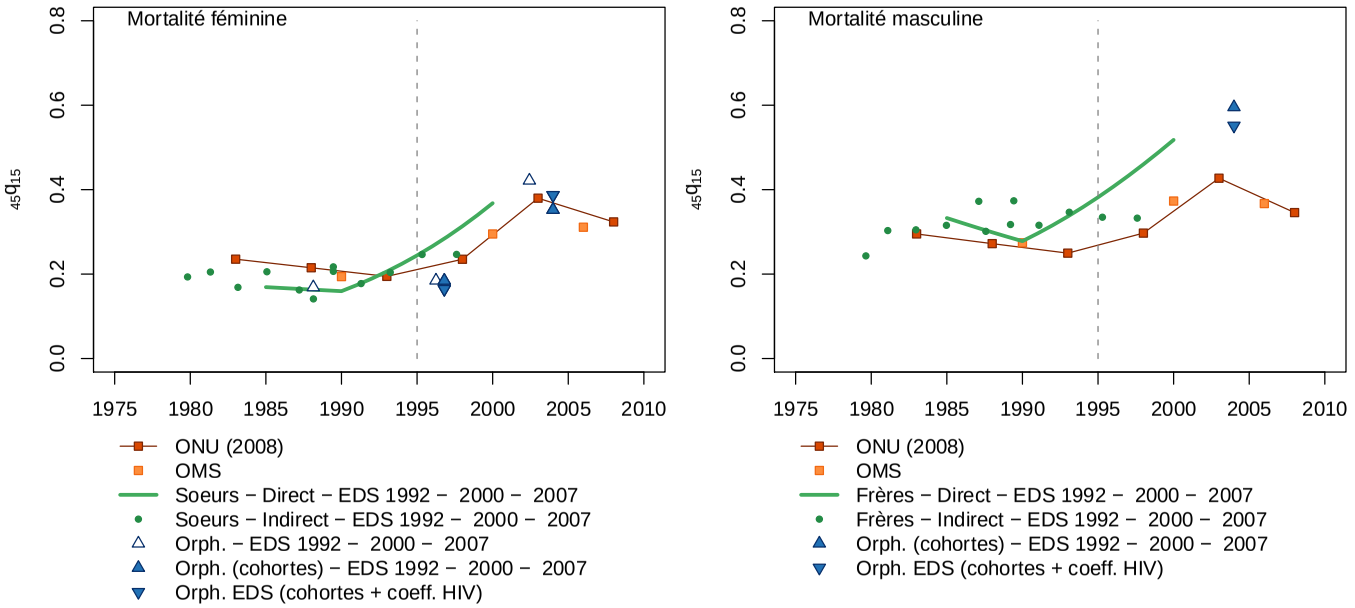

FIGURE A.41 – *Namibie : Estimations du quotient $_{45}q_{15}$ à partir de la survie des frères, des soeurs et des parents / estimations UN et OMS*

A.7.17 Nigeria

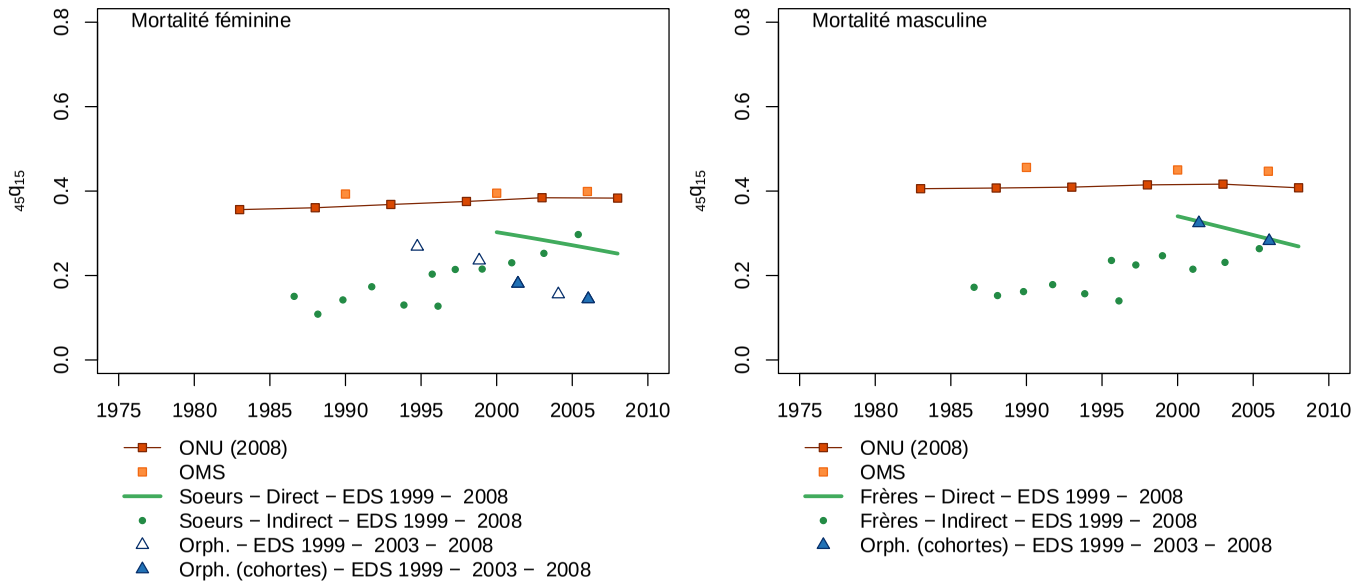

FIGURE A.42 – *Nigeria : Estimations du quotient $_{45}q_{15}$ à partir de la survie des frères, des soeurs et des parents / estimations UN et OMS*

A.7.18 Sierra Leone

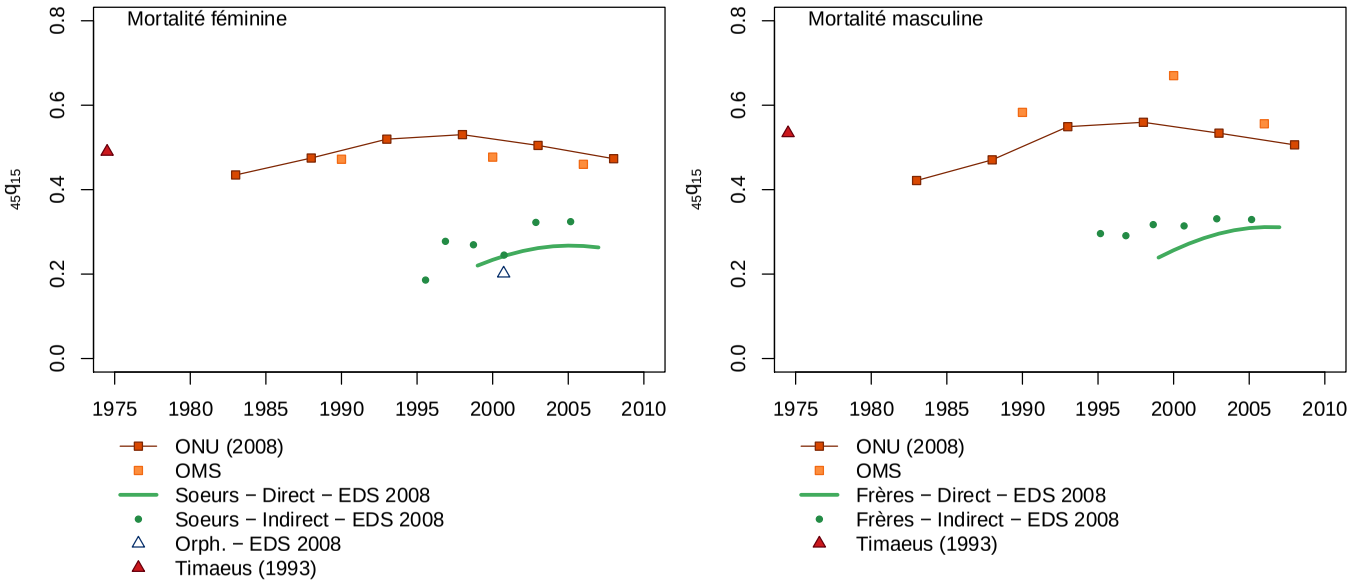

FIGURE A.43 – *Sierra Leone : Estimations du quotient $_{45}q_{15}$ à partir de la survie des frères, des soeurs et des parents / estimations UN et OMS*

A.7.19 Swaziland

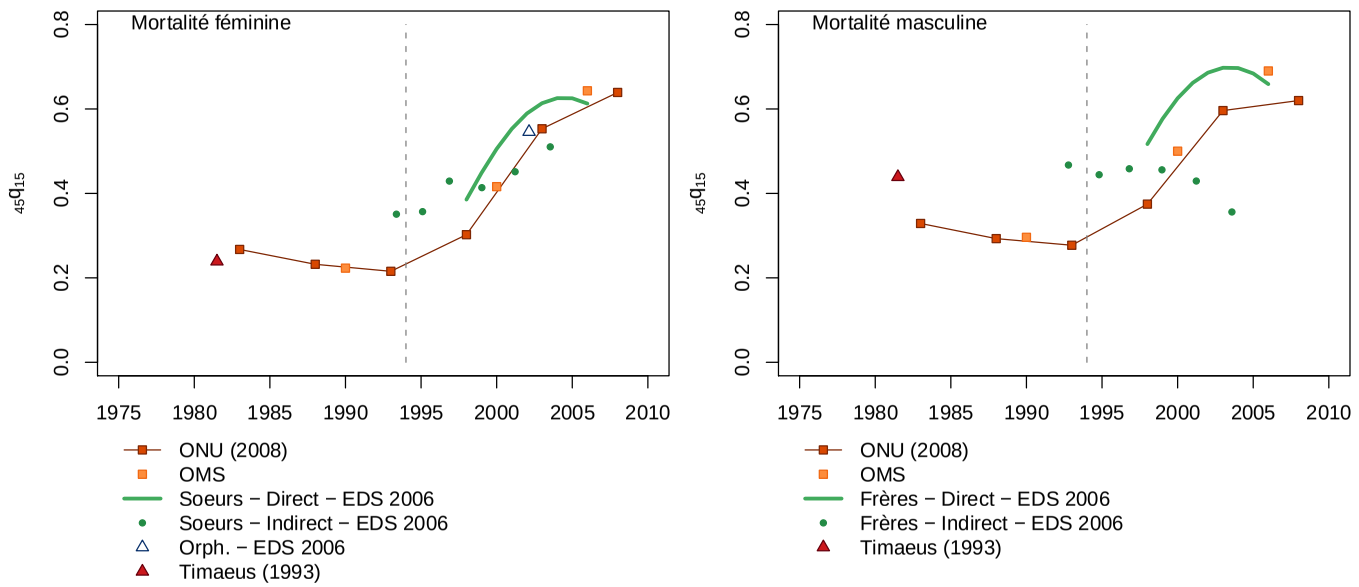

FIGURE A.44 – *Swaziland : Estimations du quotient ${}_{45}q_{15}$ à partir de la survie des frères, des soeurs et des parents / estimations UN et OMS*

A.7.20 Tanzanie

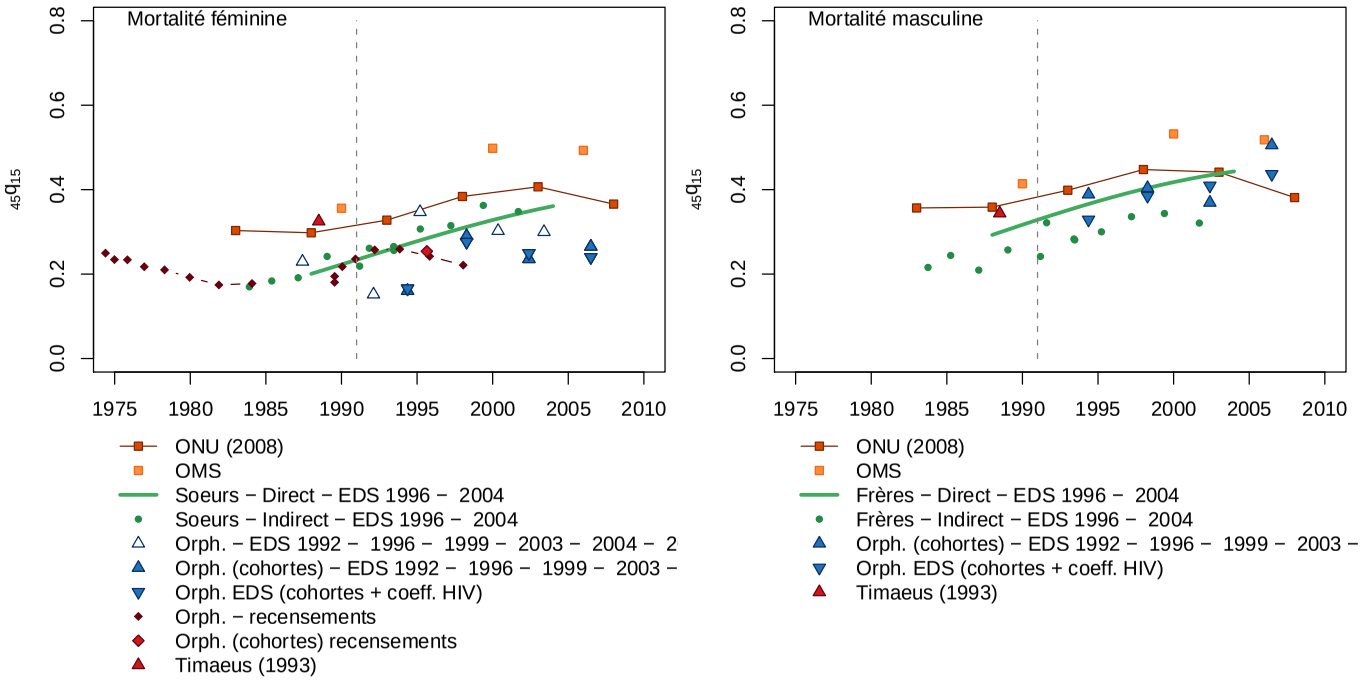

FIGURE A.45 – *Tanzanie : Estimations du quotient ${}_{45}q_{15}$ à partir de la survie des frères, des soeurs et des parents / estimations UN et OMS*

A.7.21 Togo

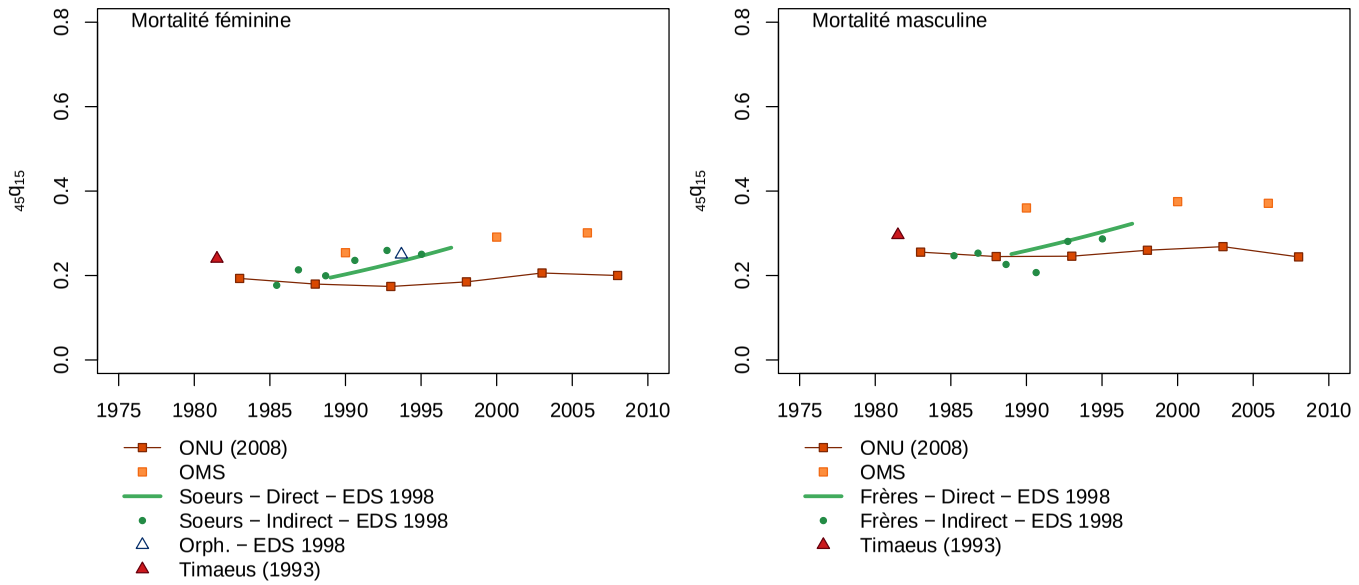

FIGURE A.46 – *Togo : Estimations du quotient $_{45}q_{15}$ à partir de la survie des frères, des soeurs et des parents / estimations UN et OMS*

Acronymes

ART	Thérapies antirétrovirales
EDS	Enquêtes démographiques et de santé
OMD	Objectifs du Millénaire pour le Développement
OMS	Organisation mondiale de la santé
PDM	proportion de décès maternels parmi les décès de femmes en âge de procréer
PNUD	Programme des Nations unies pour le développement
RMM	Rapport de mortalité maternelle (nombre de décès maternels enregistrés pour 100 000 naissances vivantes)
SIDA	Syndrôme d'immunodéficience acquise
VIH	Virus de l'immunodéficience humaine
WHS	Enquêtes mondiales de santé de l'OMS

DHS	Demographic and Health Surveys (EDS)
LSMS	Living Standard Measurement Study Survey
MICS	Multiple Indicator Cluster Survey
PMDF	Proportion maternal among deaths of females of reproductive age

Ce document a été préparé à l'aide de LaTeX 2ε à partir d'un modèle créé par Julien Chiquet (http ://stat.genopole.cnrs.fr/ jchiquet/).

Titre Estimation de la mortalité adulte en Afrique subsaharienne à partir de la survie des proches: apports de la microsimulation

Résumé La mortalité des adultes en Afrique subsaharienne reste peu étudiée. En l'absence de systèmes d'état civil efficaces, elle doit souvent être estimée à partir de données imparfaites. Parmi ces données, les informations fournies par les individus sur la survie de leurs proches (parents ainsi que frères et sœurs) constituent un patrimoine statistique important, qui mériterait d'être valorisé davantage, notamment pour mieux reconstituer les tendances de la mortalité adulte. En combinant les enquêtes démographiques et de santé (EDS), des données de recensements ainsi que des microsimulations, cette thèse revisite les techniques d'estimation de la mortalité des adultes à partir de la survie des proches. Elle présente également les principaux biais de sélection et les problèmes de qualité qui caractérisent ces données. Elle fournit ensuite des estimations empiriques récentes sur les risques de décès entre 15 et 60 ans. Dans l'ensemble, cette thèse documente des hausses importantes de mortalité adulte ainsi qu'une très grande diversité des situations nationales. Si la mortalité adulte africaine peut varier du simple au triple d'un pays à l'autre, c'est notamment parce que les différents pays subsahariens ont été frappés à des degrés très divers par l'épidémie du VIH-sida et la récente crise sanitaire.

Mots-clés Mortalité adulte, Afrique subsaharienne, orphelins, survie des frères et sœurs, microsimulation

Title Estimation of adult mortality in Sub-Saharan Africa from kin survivorship statistics : insights from microsimulation

Abstract Adult mortality in Sub-Saharan Africa remains a neglected issue in the field of population studies. Due to the lack of civil registration systems, adult deaths risks still need to be estimated from defective and deficient data. Among these data, the survival of parents and siblings, as collected in censuses and surveys, can be used to reconstruct past trends of mortality. This dissertation combines such data with microsimulations to revisit methodological aspects of the estimation of adult mortality from kin survivorship statistics. The main selection biases are then discussed, alongside problems of data quality. Recent estimates of the probability of dying between ages 15 and 60 are also presented. Overall, this dissertation highlights the considerable heterogeneity of Sub-Saharan Africa, where adult mortality can triple from one country to another, mainly because of the differential impact of the HIV-aids pandemic.

Keywords Adult mortality, Sub-Saharan Africa, orphanhood, sibling survival, microsimulation