

Séance « Déterminants de la mortalité adulte en Afrique »  
5ème Conférence sur la Population Africaine  
Arusha, Tanzanie: 10-14 Décembre 2007

Analyser les différentiels de la mortalité adulte dans une  
perspective biographique: l'apport des enquêtes EDS

*Version provisoire*

Bruno Masquelier

Chercheur FNRS

Institut de démographie (UCL)

1, Place Montesquieu, Bte 17

1348 Louvain-la-Neuve

Tel: ++ 32 (0)10 - 479 191

<http://www.uclouvain.be/demo>

Résumé :

L'objectif de cette communication, d'ordre méthodologique, est de poser les jalons nécessaires à une analyse des différentiels de mortalité adulte à partir des enquêtes EDS. Nous cherchons donc à répondre à trois séries de questions :

(1) quelle est la *fiabilité* des informations fournies sur les frères et sœurs dans les modules EDS relatifs à la survie des collatéraux ?

(2) Quelle est la *validité* des estimations déduites de ces modules et comment corriger les principaux biais de sélection ?

(3) Quel est l'apport de ces modules EDS pour explorer les *différentiels* ?

Ces trois questions structurent la logique de la recherche, qui s'appuie à la fois sur les enquêtes EDS classiques pour l'Afrique sub-saharienne et sur une enquête EDS menée en deux phases en 1995 dans la région de Mwanza (Tanzanie). Dans cette enquête, les déclarations faites sur une *même* femme par *différentes* sœurs peuvent être appariées, et confrontées entre elles. Ce couplage permet d'évaluer la qualité des informations fournies sur les sœurs, comme il permet d'estimer l'ampleur des biais de sélection dans le calcul classique des taux de mortalité. Ce couplage permet également d'estimer l'hétérogénéité imputable au fait que les fratries peuvent être répétées dans le fichier, puisque différentes sœurs sont interrogées sur leur fratrie. Après avoir évalué la fiabilité des données et la validité des estimations, nous cherchons à mettre à jour des différentiels de mortalité adulte en optant pour un modèle biographique au niveau national, à partir d'équations d'estimations généralisées (GEE). L'analyse met en évidence de profondes inégalités inter-ethniques à travers l'Afrique sub-saharienne.

## 1. INTRODUCTION

La mortalité adulte africaine souffre d'un manque d'intérêt scientifique et politique flagrant (Tabutin, 1995 ; Timaeus, 1999). Durant les dernières décennies, elle n'a que rarement figuré parmi les priorités des programmes de santé, qui sont restés très focalisés sur la santé maternelle et infantile (Phillips & al., 1993). Puisque les systèmes d'état civil sont souvent limités aux grandes villes,<sup>1</sup> et que les données issues des établissements sanitaires restent fragmentaires et biaisées, l'analyse du phénomène repose encore beaucoup sur des enquêtes ponctuelles au cours desquelles les personnes interrogées témoignent du décès de leurs proches (parents, premiers conjoints, ou frères et sœurs).<sup>2</sup> Différentes méthodes indirectes ont dès lors été développées pour en déduire des estimations fiables (Brass & Hill, 1973 ; Timaeus, 1992 ; Hill & Trussel, 1977 ; Nations Unies, 1983 ; Timaeus & al., 2001). Ces méthodes sont qualifiées d'indirectes car la durée d'exposition au risque de décès n'est pas calculée au niveau individuel mais plutôt approximée par groupe d'âge des enquêtés (soit parce que les informations manquent sur la durée précise d'exposition, soit parce qu'elles sont jugées de qualité douteuse). Pour différentes durées d'exposition, des coefficients correcteurs peuvent ensuite être calculés pour convertir les proportions de proches survivants en probabilité de survie aux âges adultes.<sup>3</sup> Dans l'ensemble, ces méthodes indirectes procurent des estimations relativement fiables,<sup>4</sup> bien que les comparaisons avec des sources extérieures restent peu fréquentes (Feeney, 2001 ; Stanton & al., 1990 ; Timaeus & Jasseh, 2004.) Elles autorisent toutefois peu de développements lorsqu'il s'agit d'analyser les tendances et les déterminants de la mortalité, et pas uniquement son niveau. Il est par exemple difficile de tenir compte simultanément de différentes variables explicatives. En effet, les méthodes indirectes partent de proportions agrégées de proches survivants, et les nouveaux critères d'agrégation vont chaque fois réduire la taille des sous-groupes. Il est par ailleurs impossible d'analyser l'interaction entre ces variables explicatives et l'âge ou la période, puisque le calcul ne procure qu'une seule estimation pour une période donnée. Enfin, les coefficients qui servent à convertir les proportions de survivants en probabilité de survie sont obtenus à partir des modèles démographiques classiques. Or, lorsque les calculs sont menés sur des sous-populations spécifiques (par niveau d'instruction par exemple), l'hypothèse de stabilité qui caractérise ces modèles démographiques peut ne plus être tout à fait adéquate. L'exemple le plus simple en la matière concerne la méthode des orphelins, pour laquelle la durée de l'exposition est calculée à partir de l'âge moyen à la naissance des enfants. Ce calcul suppose que l'histoire génésique de la cohorte des femmes d'un groupe d'âge donné soit représentée de façon satisfaisante par l'état des femmes des groupes d'âges inférieurs. Or ce n'est pas le cas lorsque les calculs sont menés sur des sous-populations spécifiques, regroupées par exemple par niveau d'instruction (surtout lorsque ces sous-populations changent de composition). En effet, si le niveau d'instruction moyen augmente, les femmes éduquées de trente à trente-cinq ans sont proportionnellement moins nombreuses que les femmes éduquées des groupes d'âge inférieurs. Elles ont également connu un calendrier de fécondité qui n'est pas nécessairement comparable à celui des femmes des groupes d'âges inférieurs (Brass, 1979). Or nous allons supposer que l'âge moyen à la naissance des enfants est stable, par exemple, ce qui n'est pas le cas.

Durant longtemps, en Afrique sub-saharienne, la mortalité adulte a été estimée essentiellement à partir des méthodes indirectes. Leurs limites expliquent sans doute pour une bonne part l'insuffisance des recherches sur les différentiels de mortalité adulte. Il s'agit là d'un champ presque inexploré, surtout à l'échelle nationale. Aujourd'hui, nous disposons pourtant d'une série de données représentatives qui autorisent un calcul *direct* des taux (c-à-d. sans approximation de la période d'exposition). L'exploration des différentiels en est facilitée, puisque les calculs peuvent être menés au niveau individuel en conservant les informations fournies par les sœurs. Il s'agit des données issues des modules sur la survie des frères et sœurs inclus dans

<sup>1</sup> A l'exception de l'Afrique du Sud (Dorrington et al., 2004a) et du Zimbabwe (Feeney, 2001).

<sup>2</sup> Des estimations peuvent également être déduites du nombre de décès survenus dans le ménage au cours des douze derniers mois, mais les niveaux estimés par ce biais se sont souvent révélés peu plausibles, en raison d'une sous-déclaration importante ainsi que des problèmes de datation (reports des décès hors de la période de référence,...) D'autres informations précieuses peuvent être dégagées des observatoires de la population, mais la spécificité des sites interdit souvent de généraliser les estimations au niveau national.

<sup>3</sup> Ces méthodes indirectes s'accompagnent de procédés permettant de dater les estimations (Brass & Bamgboye, 1981), et d'en corriger les principaux biais. Par exemple, Timaeus et Nuun (1997) proposent de corriger les biais induits par la transmission du VIH des mères aux enfants (dans le cas de la méthode des orphelins) et Gakidou et King (2006) de corriger les biais induits par l'absence d'information sur les fratries dont personne n'a survécu.

<sup>4</sup> A l'exception peut-être de la méthode du premier conjoint, pour laquelle les mises à l'épreuve restent trop rares.

un grand nombre d'enquêtes EDS. Ces modules prennent la forme d'une histoire génésique complète des mères des enquêtées. Chacun des frères et sœurs des femmes enquêtées est listé par ordre de naissance et on enregistre pour chacun d'eux leur sexe, s'ils ont survécu ou non, de combien d'années date leur éventuel décès, ainsi que leur âge à l'enquête ou au décès.<sup>5</sup> Dans la mesure où les données prennent la forme d'une histoire génésique, elles peuvent être exploitées au niveau individuel dans une perspective biographique, ce qui simplifie grandement l'analyse et permet de conserver le maximum d'information. Par exemple, le nombre de sœurs survivantes est connu et peut servir de base à la correction des biais de sélection induits par la nature rétrospective des données. De même, il est possible de conserver les caractéristiques socioculturelles des femmes enquêtées pour approcher celle de leurs frères et sœurs. C'est aujourd'hui une des rares clés d'entrée pour analyser les différentiels de mortalité adulte. En effet, cette analyse se heurte encore souvent à la rareté des renseignements sur les personnes décédées aux âges adultes. Ceux-ci se limitent au mieux au détail des circonstances du décès et à celui des symptômes qui l'ont précédé, afin de mener des analyses par causes. Par contre, l'appartenance ethnique, le type d'équipement du logement, la région de résidence ou encore le niveau de scolarisation restent généralement inconnus. Face à cette absence d'informations *sur les personnes soumises au risque*, il peut être utile d'explorer les possibilités et les limites du recours aux informations disponibles *sur les fratries* afin d'obtenir une idée, même approximative, des inégalités de mortalité. Une première étape, qui sera adoptée ici, consiste à partir des caractéristiques dont on peut supposer qu'elles sont partagées par les frères et sœurs. C'est le cas de l'appartenance ethnique, par exemple.

L'exploitation directe des données et l'analyse des différentiels suppose toutefois deux choses : (1) les déclarations faites sur la survie des frères et sœurs doivent pouvoir être considérées comme fiables, (2) la méthode d'estimation doit pouvoir être considérée comme valide. L'objectif de cette communication, d'ordre méthodologique, est donc de répondre à ces deux questions pour poser les jalons nécessaires à l'analyse des inégalités de mortalité adulte à partir des enquêtes EDS. Elle débute sur une synthèse des indicateurs traditionnellement utilisés pour évaluer la qualité des informations fournies par les enquêtées sur leurs frères et sœurs. Elle se poursuit sur la correction des biais introduits par la spécificité des modules d'enquête. Pour ce faire, une enquête spécifique, menée dans le district de Kwimba en Tanzanie en 1995, est particulièrement utile, puisqu'elle autorise un appariement des fratries et le croisement des déclarations. Nous l'exploiterons successivement de façon « classique » puis en tenant compte de l'appariement, ce qui permet d'examiner sous un nouvel angle les biais qui caractérisent les modules d'enquêtes EDS classiques. Enfin, cette communication se conclut par l'amorce d'une analyse des disparités ethniques de mortalité adulte, au niveau national, pour une vingtaine de pays d'Afrique sub-saharienne.

## 2. QUELLE EST LA FIABILITE DES INFORMATIONS FOURNIES SUR LES FRERES ET SCEURS DANS LES MODULES EDS SUR LA SURVIE DES COLLATERAUX ?

Tout d'abord, estimer la mortalité adulte à partir des informations sur la survie des frères et sœurs suppose que l'on s'interroge sur la fiabilité de ces informations. En effet, celles-ci sont souvent suspectées de sous-estimer la mortalité, en raison d'une importante sous-déclaration des décès. Le décès d'un frère ou d'une sœur peut être ignoré, surtout s'il est survenu avant la naissance de l'enquêtée ou peu après celle-ci (Hill & Trussel, 1977).<sup>6</sup> Dans certains cas, il peut aussi être préférable de garder secret un décès, parce qu'il n'est pas socialement acceptable (pensons à la stigmatisation qui entoure les personnes infectées du VIH, par exemple). La qualité des informations fournies sur les frères et sœurs dépend également de la variabilité culturelle des définitions de la fratrie et de ses frontières : on parle ici des enfants biologiques, pas des enfants confiés, des cousins, ou des amis très proches... par exemple. Des problèmes de déclaration d'âge peuvent également se poser : quand ce frère est-il né, à quel âge est-il décédé ? Face à tous ces problèmes potentiels, cette section relative aux données vise à donner un ordre d'idée de la qualité des déclarations. Plusieurs indicateurs sont couramment utilisés, et nous en proposons ici une brève synthèse pour l'ensemble des données EDS disponibles. Il s'agit (1) du rapport de masculinité à la naissance des frères et sœurs déclarés, (2) de l'écart entre la médiane des années de naissance des

<sup>5</sup> Quelques questions sont également posées pour pouvoir identifier les décès maternels.

<sup>6</sup> Notons que dans le cas d'une exploitation directe, la sous-déclaration n'a d'importance que s'il s'agit de décès adultes, ce qui n'est pas le cas des estimations indirectes. Il reste que les décès de frères et sœurs adultes peuvent également être ignorés.

répondants et la médiane des années de naissance de ces frères et sœurs, et enfin (3) de la parité des mères des répondants selon leur âge. Dans un second temps, nous discuterons de la concordance des déclarations faites par les différentes enquêtées sur une même sœur, à partir de l'enquête EDS spécifique menée en Tanzanie qui a permis un couplage des déclarations. Au total, 47 enquêtes EDS ayant prévu un module sur la mortalité maternelle sont disponibles à ce jour pour l'Afrique sub-saharienne (tableau 1).

Tableau 1 : Ensemble des enquêtes africaines avec un module disponible sur la mortalité adulte (11/ 07)<sup>7</sup>



Pays	02-06	96-01	90-95	Pays	02-06	96-01	90-95
Afrique de l'Ouest				Afrique de l'Est			
Bénin		1996		Ethiopie	2005	2000	
Burkina Faso	2003	1998		Kenya	2003	1998	
Côte d'Ivoire	2005		1994	Madagascar	2004	1997	1992
Guinée	2005	1999		Malawi	2004	2000	1992
Mali		2001	1995	Mozambique	2003		
Niger	2006		1992	Ouganda	2006	2001	1995
Nigeria		1999		Rwanda		2000	
Sénégal	2005		1993	Soudan			1990
Togo		1998		Tanzanie	2004	1996	
				Zambie	2002	1996	
				Zimbabwe	2006	1999	1994
Afrique centrale				Afrique australe			
Cameroun	2004	1998		Afrique du Sud		1998	
Gabon		2000		Namibie		2000	1992
Centrafrique			1995				
Tchad	2004	1997					

Rappelons qu'en plus de ces enquêtes « classiques », une enquête EDS spécifique est particulièrement intéressante. Elle a été menée en 1995 dans la région de Mwanza, au nord-ouest de la Tanzanie, autour de la ville de Sumve (*Sumve Survey on Adult and Childhood Mortality - SACM 1995*). Cette enquête visait en fait à estimer l'ampleur des biais qui sont introduits par la surmortalité des orphelins dans l'estimation de la mortalité infanto-juvénile basée classiquement sur les déclarations faites par les mères. Cette collecte rétrospective induit un phénomène de sélection puisque la mortalité des enfants de mère décédée échappe à l'échantillonnage. Cela biaise les estimations à la baisse puisque les orphelins font souvent face à une surmortalité. L'enquête SACM a donc été conçue pour évaluer si l'histoire génésique des femmes peut être collectée de façon fiable en interrogeant leurs sœurs. Si c'était le cas, des informations sur la fécondité des mères décédées et la survie de leurs enfants pourraient être obtenues à partir des sœurs. Puisqu'il s'agit d'évaluer la qualité des déclarations données par les sœurs, cette enquête repose sur le croisement de différentes déclarations et s'est déroulée en deux phases. La première a été menée au sein de six villages dans le district de Kwimba. 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). Ensuite, pour chacune de ces sœurs, les femmes ont fourni un ensemble d'informations permettant de les identifier et de les retrouver.<sup>8</sup> Dans une seconde phase, une grande majorité des sœurs identifiées ont été interrogées à leur tour. A peu de choses près, les enquêteurs leur ont posé les mêmes questions que celles posées en phase I, afin de pouvoir confronter les déclarations et évaluer la qualité des données.

<sup>7</sup> Les enquêtes Mauritanie 2001, Guinée 1992, Erythrée 1995 et Lesotho 2004 ne sont pas encore dans le domaine public

<sup>8</sup> Nom du district où vit chacune d'elles, nom de leur village et de leur chef de village, nom de leur chef de ménage, nom de leur mère, etc.

## ENCADRÉ 1 : COUPLAGE DES DÉCLARATIONS FAITES PAR LES SŒURS LORS DE L'ENQUÊTE SACM

Bicego et al. (1997) ont apparié les déclarations qui portaient sur des sœurs ayant répondu à l'enquête lors de l'une des deux phases. Pour compléter ce couplage, nous avons apparié les déclarations qui portaient sur les sœurs vivant hors de la zone ou sur les sœurs décédées, à partir du travail fait par Bicego et al. (1997). Au total, 13700 déclarations ont été appariées. Au départ des nom et prénom de la mère des enquêtées, des fratries potentielles ont été assemblées et numérotées. Ensuite, un identifiant unique a été associé aux déclarations faisant référence à une même sœur, qu'il s'agisse d'une déclaration faite par la personne elle-même (*self-report*) ou par une de ces sœurs (*proxy-report*). En moyenne, chaque personne est mentionnée près de 2.5 fois (certaines d'entre elles peuvent l'être plus d'une dizaine de fois). En cas de doute sur des déclarations discordantes supposées faire référence à la même femme, afin de décider s'il s'agissait de personnes différentes issues de fratries distinctes, la structure de la fratrie déclarée a été examinée (combien de frères, combien de sœurs, dans quel ordre de naissances...). Dans le cas d'incohérences manifestes - des âges déclarés très éloignés, des parités très différentes, etc. - des femmes qui ont été considérées comme sœurs à première vue sont séparées dans des fratries distinctes. Dans un premier temps, l'appariement a été mené manuellement, à partir des critères suivants : l'âge et le rang de naissance de chaque femme, ses nom, prénom et patronyme, son histoire génésique<sup>9</sup>, l'ordonnement de ses naissances par sexe, le nombre d'enfants décédés (par sexe), ainsi que la survie de ses parents, et son âge à la première naissance. Le nombre des enfants survivants et la séquence de leurs naissances par sexe a été contrôlée de façon systématique. Cet appariement manuel est laborieux, mais il autorise une souplesse qui convient bien aux données imparfaites.<sup>10</sup>

Dans un second temps, des appariements probabilistes ont été utilisés afin de repérer encore quelques erreurs de couplage. Il s'agit en fait de calculer la probabilité statistique que deux déclarations fassent référence à la même personne à partir d'un ensemble de variables d'identification (noms, prénoms, noms de la mère, rangs de naissance et année de naissance).<sup>11</sup> Les couplages les plus probables qui contredisaient l'appariement manuel ont été vérifiés.<sup>12</sup> Cela dit, il reste difficile de juger de la qualité du couplage en l'absence de données fiables sur les liens de parenté.

### 2.1 Rapports de masculinité

Commençons par le test de qualité des déclarations le plus classique : l'examen des rapports de masculinité. Il a son importance pour notre propos, notamment dans la perspective d'une analyse des différentiels sexuels de mortalité adulte. La littérature documente généralement chez les adultes une surmortalité masculine légère et quasi-générale.<sup>13</sup> Ces différentiels sexuels ont sans doute été modifiés par la propagation du VIH. Les infections sont plus fréquentes chez les femmes, en raison d'une vulnérabilité biologique plus grande que chez les hommes, ainsi que d'une activité sexuelle plus précoce avec des partenaires plus âgés. Gregson et Garnett (2000) ont toutefois fait remarquer qu'une mortalité masculine plus élevée pourrait être un trait caractérisant les premiers temps de l'épidémie, notamment parce que la létalité est plus élevée chez les hommes infectés qui sont en moyenne plus âgés que les femmes. L'exploitation directe des enquêtes EDS pourrait permettre d'étudier les différentiels sexuels de mortalité adulte. Mais les informations données sur les sœurs sont-elles aussi fiables que les informations données sur les frères ? La question n'est pas simple. Un examen sommaire des rapports de masculinité peut en tous cas mettre en évidence une sur-représentation masculine dans l'échantillon des frères et sœurs (ou l'inverse), *sans refléter pour autant nécessairement un problème lié à la qualité des déclarations*. Il y a deux raisons à cela. Tout d'abord, les rapports de

<sup>9</sup> En autorisant des oublis en cas de naissance récente, très ancienne ou d'enfant décédé, et en vérifiant les naissances de jumeaux.

<sup>10</sup> Les erreurs d'âge faites par une répondante qui tend à vieillir toutes ses sœurs sont plus facilement acceptables que lorsqu'elles sont considérées isolément, par exemple. La phonétique des prénoms ou des noms de famille est également mieux examinée : Lorencia, Rolencia et Laurencia se réfèrent à la même personne, et Barthomeo, Batolomayo et Bathlomeo à une même famille.

<sup>11</sup> Avec Link Plus, un programme d'appariement développée par le US National Program of Cancer Registries pour les données médicales : <http://www.cdc.gov/cancer/npcr/tools/registryplus/lp.htm>

<sup>12</sup> Dans la grande majorité des cas, l'appariement manuel restait plus plausible, notamment parce qu'il tient compte de la cohérence d'ensemble des déclarations faites par les sœurs interrogées. Par exemple, si une femme intervertit les noms et prénoms de ses sœurs, l'examen des déclarations faites par ses autres sœurs permet facilement de détecter le problème, alors que le logiciel ne peut pas l'intégrer.

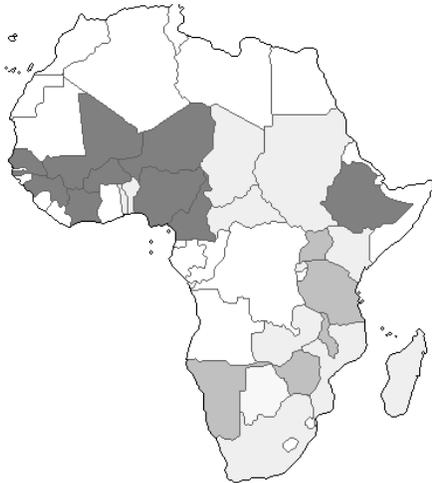
<sup>13</sup> A l'exception de l'Afrique australe où elle semble plus prononcée, ce qui peut s'expliquer par l'importance des migrations de travail vers l'Afrique du sud et l'incidence de la tuberculose dans les régions minières de la sous-région (Timaus, 1991).

masculinité qui sont présentés dans les rapports d'enquêtes sont légèrement biaisés à la baisse, parce qu'ils sont calculés sur l'ensemble des frères et sœurs. Ensuite, les rapports de masculinité ne sont pas constants en Afrique sub-saharienne, et l'hétérogénéité des rapports observés à partir des enquêtes pourrait refléter une hétérogénéité réelle des rapports, qui suit une géographie précise.

D'une part, calculer les rapports sur tous les frères et sœurs introduit un biais. Il faut en effet considérer que 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. En effet, différentes sœurs peuvent répondre à l'enquête, alors que les frères ne remplissent pas cette partie du questionnaire. Il n'est donc pas étonnant que les rapports de masculinité calculés au sein des fratries *mentionnées plusieurs fois*<sup>14</sup> se situent très en deçà de la valeur théorique, située à 103 environ en Afrique noire (Brass & al., 1968). Dans une trentaine d'enquêtes, les rapports de ces fratries répétées sont même inférieurs à 90.<sup>15</sup> Pour mieux évaluer la qualité des données, les calculs devraient donc plutôt être menés après avoir pondéré chaque observation. Je propose de les pondérer *par l'inverse du nombre de sœurs enquêtées*, ce qui a pour conséquence de relever quelque peu les rapports. Ainsi, celui obtenu à partir de l'enquête menée au Gabon (2000) passe de 100 à 104, celui obtenu en Afrique du Sud (1998) de 100 à 105. Une dizaine de rapports, qui, non-pondérés, semblaient témoigner d'une sous-déclaration des frères par rapport aux sœurs (au seuil 0.05), n'indiquent plus de biais significatif une fois pondérés (Kenya (1998), Afrique du sud (1998), Tanzanie (1996), Zambie (1996), etc.).<sup>16</sup>

Le tableau 2 répartit l'ensemble des modules disponibles sur la mortalité maternelle selon qu'ils présentent des rapports contenus dans l'intervalle de confiance (à 95%), en dessous de la borne inférieure (en fond gris clair) ou au-dessus à la borne supérieure (en fond gris foncé).

Tableau 2 : Rapports de masculinité des frères et sœurs déclarés dans les enquêtes EDS



Pays	02-06	96-01	90-95	Pays	02-06	96-01	90-95
Afrique de l'Ouest				Afrique de l'Est			
Bénin		104		Ethiopie	111	111	
Burkina Faso	107	109		Kenya	103	101	
Côte d'Ivoire	108		104	Madagascar	105	105	103
Guinée	108	105		Malawi	102	101	100
Mali		108	106	Mozambique	104		
Niger	108		105	Ouganda	102	102	100
Nigeria		113		Rwanda		105	
Sénégal	107		109	Soudan			-
Togo		103		Tanzanie	101	102	
				Zambie	103	102	
				Zimbabwe	100	99	100
Afrique centrale				Afrique australe			
Cameroun	108	105		Afrique du S.		105	
Gabon		104		Namibie		98	103
Centrafrique			101				
Tchad	107	108					

■ = Sous-déclaration des filles      ■ = Sous-déclaration des garçons

<sup>14</sup> A ce stade, les fratries sont sommairement identifiées à partir du cluster, du ménage, et de la taille de la fratrie de l'enquêtée. On fait donc l'hypothèse qu'il ne peut y avoir deux fratries différentes de même taille dans un même ménage déclarées par deux femmes de 15 à 49 ans. On fait également l'hypothèse d'une faible probabilité d'échantillonner une même fratrie dans des ménages différents.

<sup>15</sup> Par exemple, les rapports de masculinité calculés à partir des fratries mentionnées plusieurs fois s'élèvent seulement à 76 [110-100] en Namibie (EDS 2000), à 77 [109-100] en Afrique du sud (EDS 1998), à 79 [109 – 100] au Malawi (EDS 2000), à 80 [111-98] au Zimbabwe (EDS 1999), ou encore à 84 [109 – 100] au Gabon (EDS 2000).

<sup>16</sup> Par contre, une fois pondérés, les rapports calculés à partir des enquêtes menées en Côte d'Ivoire (2005), Tchad (1997) et Sénégal (2005 et 1992) semblent témoigner d'une sous-déclaration des sœurs par rapport aux frères.

Les calculs sont menés sur les cohortes de naissances survenues entre 15 et 50 ans avant l'enquête, après pondération avec une valeur théorique de 103. Malgré la légère correction apportée aux rapports tels qu'ils sont publiés, un peu moins de la moitié des enquêtes présentent toujours un rapport de masculinité qui s'écarte de la valeur attendue (au-delà de ce qu'autorisent les fluctuations aléatoires). Ce constat rapide semble indiquer que la qualité des déclarations est sujette à caution. Cela étant, deux éléments nous invitent à remettre cette conclusion en question.

Tout d'abord, des régressions logistiques menées par pays indiquent que la sous-déclaration d'un sexe par rapport à l'autre ne varie pas de façon significative avec le niveau d'instruction des sœurs enquêtées ni avec leur âge ou encore l'équipement de leur ménage. Cela est vrai même dans les pays où le rapport de masculinité s'écarte sensiblement de la valeur attendue (Malawi, Zimbabwe, etc.)

Ensuite, il y a une certaine logique géographique qui doit retenir notre attention. Il est possible de classer les enquêtes en trois groupes. Le premier regroupe les enquêtes menées au Nigeria et en Ethiopie, où la moyenne des rapports avoisine 112, ce qui est très élevé. Le second regroupe les autres enquêtes menées en Afrique de l'Ouest et en Afrique centrale, avec une moyenne qui s'élève 106, et où se concentrent la plupart des enquêtes où le rapport est apparemment trop élevé<sup>17</sup>. Dans le dernier groupe, on retrouve les enquêtes menées en Afrique australe et en Afrique de l'Est, avec une moyenne de 102. La plupart des rapports de ce groupe qui s'écartent de la valeur théorique semblent témoigner d'une tendance à la sous-déclaration des frères. Il est difficile de soutenir que des biais systématiques pourraient être à l'origine de ce regroupement régional. Certes, celui-ci pourrait résulter de comportements migratoires masculins plus importants en Afrique de l'Est et en Afrique australe, qui amèneraient les déclarantes à ignorer davantage les frères éloignés. Elles pourraient également refléter une géographie des inégalités sexuelles de mortalité, qui se traduiraient par une sur-mortalité féminine en Afrique de l'Ouest et une sur-mortalité masculine en Afrique australe et en Afrique de l'Est. Mais ces deux hypothèses supposent que les différentiels sexuels de mortalité soient très prononcés ou que les comportements migratoires soient très sexués. Et elles n'expliquent pas la forte sur-représentation des garçons en Ethiopie et au Nigeria. Une troisième hypothèse peut donc être formulée : cette géographie reflète moins la variation dans les erreurs de déclarations, que la *variation dans les rapports de masculinité à la naissance*. Le découpage régional observé ici recoupe en effet partiellement celui observé par Garenne (2002b). Celui-ci distingue également trois sous-groupes de pays au départ des rapports de masculinité à la naissance (calculés à partir des enfants dans les enquêtes EDS) : un sous-groupe aux rapports peu élevés, essentiellement dans les populations bantoues de l'Afrique australe et de l'Afrique de l'Est (101), un second sous-groupe 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. Comme les travaux de Garenne (2002), l'exploitation des modules d'enquête EDS sur les frères et sœurs semble donc indiquer que les rapports de masculinité à la naissance suivent une géographie opposant les populations à dominance bantoue aux autres populations d'Afrique sub-saharienne. Par contre, même si la pondération utilisée ici doit sans doute être affinée, la sous-déclaration d'un sexe par rapport à l'autre n'est pas évidente. Une sous-déclaration des sœurs n'est toutefois pas à exclure, au vu des rapports élevés observés en Afrique de l'Ouest et en Afrique centrale.

## 2.2 Distribution des dates de naissances

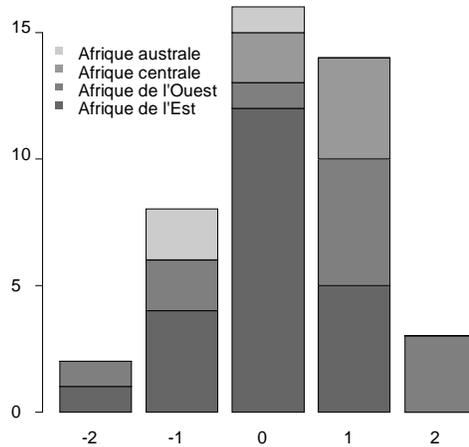
Nous continuons notre évaluation de la qualité des données sur les frères et sœurs en examinant les dates de naissances déclarées. Tandis que l'examen du rapport de masculinité vise à déceler une éventuelle sous-déclaration plus importante d'un sexe par rapport à l'autre, la distribution des dates de naissances permet de déceler une éventuelle sous-déclaration plus importante des frères et sœurs *nés avant le répondant*. Théoriquement, chacun a en moyenne autant de petits frères que de grands frères.<sup>18</sup> On peut donc considérer que l'année de naissance d'un répondant est très proche de la médiane des années de naissance

<sup>17</sup> C'est-à-dire où l'existence de certaines sœurs semble avoir été davantage sous-déclarée que celle des frères.

<sup>18</sup> Plus précisément, dans une population stationnaire, le ratio du nombre moyen de frères plus jeunes sur le nombre moyen de frères plus âgés avoisine l'unité (pour les individus dont la mère a terminé son cycle reproductif). Par contre, dans une population stable *non stationnaire*, pour un ensemble d'individus d'âge  $a$ , plus les mères sont jeunes, plus le nombre moyen de frères et sœurs plus jeunes sera élevé par rapport au nombre moyen de frères et sœurs plus âgés (Timaus et al. 2001, Goldman, 1978).

de ses frères et sœurs. Ici, nous examinons donc la *médiane des dates de naissance des frères et sœurs* et nous la comparons à la *médiane des dates de naissance des répondants*. Si les individus nés avant l'enquête(e) sont sous-déclarés, la médiane des dates de naissances des frères et sœurs va être trop élevée et elle va dépasser de plusieurs années la médiane des années de naissance des enquête(e)s. Dans l'ensemble, cela ne semble heureusement pas être trop le cas. En effet, comme le montre le graphe 1, la distribution des écarts (en années) entre ces deux médianes est concentrée autour de zéro. Puisque ces écarts restent minimes et ne dépassent jamais deux ans, on peut considérer que l'omission des générations plus anciennes est peu prononcée.

Graphe 1 : Différences (en années) entre la médiane des dates de naissances des frères et sœurs déclarés et la médiane des dates de naissance des enquêtées.



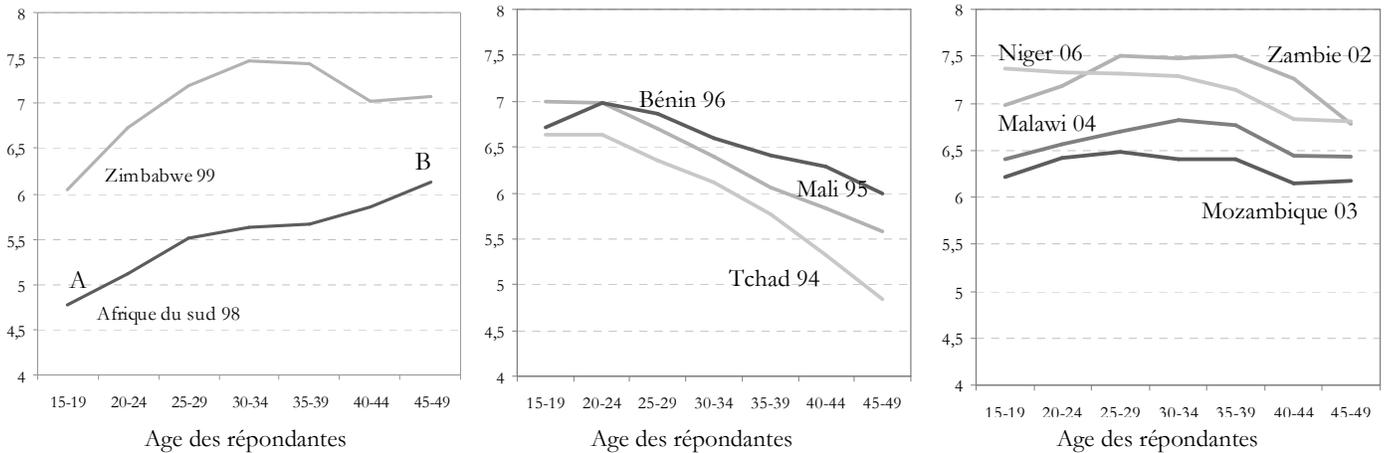
### 2.3 Evolution de la parité des mères

Troisième type de test : le nombre de frères et sœurs déclarés par les répondantes. Lorsqu'un membre de la fratrie (la personne enquêtée) est ajouté, nous obtenons la parité des mères des enquêtées, que l'on peut comparer à la fécondité de la génération antérieure. Par exemple, en Afrique du sud, les enquêtées de 15 à 19 ans déclarent avoir eu 3,7 frères et sœurs en moyenne. La génération de leurs mères auraient donc donné naissance à environ 4,7 enfants en moyenne (point A sur le graphe suivant). Est-ce que cela est cohérent avec les estimations dont nous disposons ? Un rapide calcul indique que l'âge moyen des mères à la naissance de leurs enfants s'élevait à 26,8 ans en 1998, en Afrique du sud. Si cet âge n'a pas évolué dans le temps, les mères de ces répondantes âgées en moyenne de 17,5 ans ont aujourd'hui approximativement quarante-quatre ans (17.5+26.8). Or les femmes de quarante à quarante cinq ans déclarent avoir eu 4,2 enfants. Le nombre de frères et sœurs déclarée par les enquêtées les plus jeunes est donc cohérent avec la parité déclarée dans la même enquête par les répondantes les plus âgées. Et nous devrions retomber sur les tendances de la fécondité (de la génération antérieure) en examinant le nombre de frères et sœurs déclarés par des enquêtées plus âgées (Stanton & al., 2000). Le graphe 2 présente donc les évolutions du nombre de frères et sœurs *par les répondantes de différents âges* pour différentes enquêtes. En conservant toujours l'exemple de l'Afrique du sud, on constate que selon les répondantes âgées de 45 et 49 ans, la génération de leurs mères a donné naissance à un peu plus de six enfants en moyenne (point B du graphe 2). Comme ces naissances remontent à une période antérieure aux naissances des mères des répondantes plus jeunes, cela signifie que la fécondité a baissé, ce qui est cohérent. Les mêmes constats pourraient être faits pour l'enquête menée au Zimbabwe en 1999. Plus on avance dans l'âge des enquêtées, plus on remonte dans le temps.

Les problèmes surviennent lorsque les femmes plus âgées déclarent avoir eu moins de frères et sœurs que les femmes les plus jeunes, comme c'est le cas au Bénin (1996). La génération antérieure aurait donc connu une

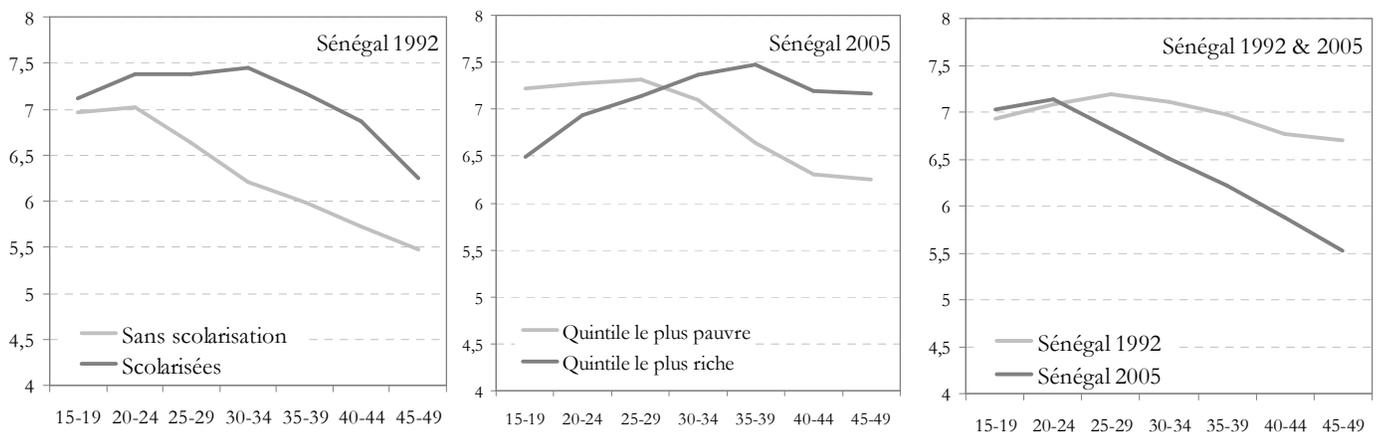
fécondité en hausse, ce qui est peu probable<sup>19</sup>. Qu'en conclure ? La propension à ne pas déclarer tous ses frères et sœurs augmente avec l'âge des répondants. Cette tendance se manifeste dans de nombreux pays, quel que soit le niveau de fécondité. Elle est relativement forte au Mali (1995) et au Tchad (1994), par exemple, où elle surpasse les effets de la baisse de la fécondité. Elle compense la baisse de la fécondité dans d'autres pays (Niger 2006, etc.).

Graphe 2 : Parité des mères des répondantes selon l'âge de ces dernières - EDS Afrique du Sud 1998, Bénin 1996, Mozambique 2003, Nigeria 1999, Sénégal 1992 et 2005, Zambie 2002 et Zimbabwe 1999



Comme l'indique le graphe suivant, à travers l'exemple du Sénégal, la sous-déclaration des frères et sœurs par les enquêtées âgées peut se manifester de façon inégale selon les niveaux d'instruction, ainsi que selon les niveaux d'équipement des ménages. Par exemple, les femmes du quintile le plus pauvre déclarent de moins en moins de frères et sœurs à mesure qu'elles avancent en âge, alors que les femmes du quintile le plus riche en déclarent de plus en plus (ce qui reflète mieux les tendances de fécondité de la génération antérieure). Une analyse des inégalités de mortalité adulte qui reposerait sur les caractéristiques des répondantes (qui comparerait par exemple la mortalité des frères et sœurs de femmes non instruites à la mortalité des frères et sœurs de femmes instruites) serait donc biaisée : les femmes sans scolarisation omettent davantage de frères et sœurs (et sans doute d'autant plus qu'il s'agit de frères et sœurs décédés) que les femmes scolarisées, etc.

Graphe 3 : Parité des mères des répondantes selon l'âge de ces dernières - EDS Sénégal 1992 et 2005 : par niveau d'instruction, équipement du ménage et par année d'enquête



Ce troisième test indique donc que la fiabilité des déclarations diminue avec l'âge des répondants, mais augmente avec leur niveau d'éducation et avec leur condition sociale. En définitive, la sous-déclaration semble moins

<sup>19</sup> Même en admettant que la fécondité ait légèrement remonté dans un certain nombre de pays africains durant les années septante et quatre-vingt.

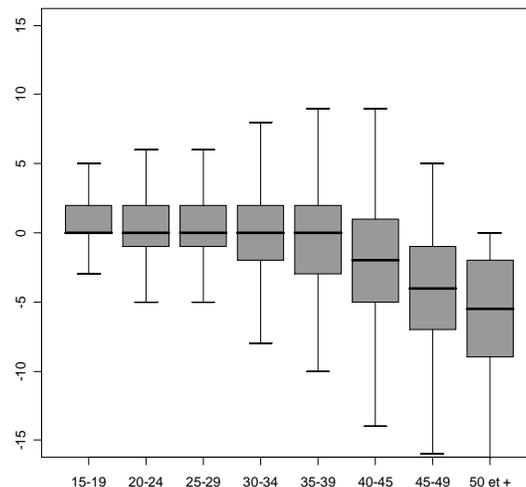
caractériser les déclarations anciennes que les déclarations faites par des femmes âgées, puisque les dates de naissances médianes sont proches mais que la taille des fratries déclarées par les femmes plus âgées est douteuse.

## 2.4 Concordance des déclarations entre sœurs

Un quatrième et dernier type de test peut être mené à partir de l'enquête conduite en Tanzanie en 1995. On l'a souligné, cette enquête permet de croiser les informations fournies par les répondantes sur leurs sœurs (c.-à-d. les informations *proxy-reported*) avec celles fournies par ces mêmes sœurs lorsqu'elles ont pu être réinterrogées (c.-à-d. les informations *own-reported*). Il est donc possible de calculer pour chaque femme 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*. En reprenant toutes les déclarations relatives à une sœur qui a pu être enquêtée (soit 44% des déclarations<sup>20</sup>), la moyenne de ces écarts absolus s'élève à deux ans et demi, et leur écart-type s'élève à trois ans. Ce constat indique que les erreurs d'âges déclarés pour les autres membres de la fratrie restent relativement faibles (bien que la mesure de ces écarts interfère avec les choix faits lors de l'appariement). 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 « moyenne » (avec un coefficient kappa évalué à 0.493)<sup>21</sup>.

Le graphe 4 détaille 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-mêmes le déclarent (*own-report*). Il montre que les problèmes de discordance se situent plus particulièrement au niveau des âges déclarés pour les femmes plus âgées (à partir de quarante ans) ; ces âges sont moins fiables et caractérisés par un biais à la baisse. Une tendance au rajeunissement des femmes par leurs sœurs se dessine donc à mesure que ces femmes avancent en âge.

Graphe 4 : Evolution des écarts d'âge en fonction de l'âge des femmes tel qu'elles-mêmes le déclarent (Enquête SACM - Tanzanie 1995).



Puisque nous disposons des caractéristiques des femmes enquêtées, nous pouvons tenter de mieux appréhender la variation des écarts entre l'âge qu'elles déclarent pour leurs sœurs (*proxy-report*) et l'âge que ces mêmes sœurs déclarent elles-mêmes (*own-report*). En conservant toujours toutes les déclarations relatives à une sœur qui a pu être enquêtée, une régression linéaire met simplement en relation ces écarts d'âge avec une série de variables indépendantes. Les coefficients obtenus et leur niveau de signification sont présentés dans le tableau 3. Comme l'a montré le graphe précédent, les erreurs d'âges augmentent de façon significative avec l'âge de la personne *sur lesquelles elles portent*. Comme l'a montré la section précédente, les

<sup>20</sup> 27 % des déclarations sont des déclarations sur soi-même et 29% des déclarations portent sur des sœurs qui n'ont pas pu être enquêtées.

<sup>21</sup> Le coefficient Kappa exprime la différence entre la proportion d'accord observé et la proportion d'accord aléatoire (c.-à-d. la valeur espérée en cas d'indépendance des déclarations), divisée par la proportion d'accord disponible au-delà de l'accord aléatoire :  $(P_{obs} - P_{al}) / (1 - P_{al})$ . On tient compte de la sorte de la quantité d'accord qui serait observé sous l'effet du hasard.

enquêtées sont également moins fiables à mesure qu'elles avancent en âge.<sup>22</sup> Le fait que la mère des enquêtées soit toujours en vie va faire diminuer les erreurs d'âge. De même, la scolarité atteinte par les répondants diminue significativement ces erreurs, tout comme certaines caractéristiques socio-économiques, telles le fait de posséder un sol de ciment ou une radio.

Dans l'ensemble, cette section met en évidence une qualité variable des déclarations, qui peut toutefois être considérée comme satisfaisante dans un contexte de quasi-absence de données sur la survie aux âges adultes. Puisque ces données prennent la forme d'une histoire génésique, elles sont analysées de façon la plus complète en optant pour une perspective biographique. Dans ce cas, l'analyse ne porte pas exclusivement sur l'occurrence d'un décès mais considère plutôt le temps écoulé depuis l'entrée dans la période d'exposition au risque jusqu'au décès ou jusqu'à la date de l'enquête menée auprès d'une sœur. Le modèle biographique adopté est détaillé dans la section suivante.

Tableau 3 : Régression (linéaire) sur les caractéristiques des répondants des écarts (en mois) entre âges déclarés par les femmes (de 15 à 50 ans) et âges déclarés pour ces mêmes femmes par leurs sœurs.

	Coeff. non standardisés (b)	Coeff. standardisés (B)	t	Signif.	
Constante	14,214		4,215	,000	***
Age 'réel' déclaré par la sœur [15-49]	0,830	,204	14,380	,000	***
Age de l'enquêtée	0,401	,104	6,827	,000	***
Ecart dans les rangs de naissances	0,709	,040	3,092	,002	***
Survie de la mère	-5,023	-,049	-3,722	,000	***
Fraction de la scolarité obligatoire att. à 15 ans	-9,144	-,108	-7,603	,000	***
Kiswahali (vs. Kisukuma)	3,128	,041	3,156	,002	***
Sol de ciment	-9,218	-,055	-4,157	,000	***
Possession d'une radio	-1,588	-,023	-1,692	,091	*
Nombre de chèvres	-0,112	-,017	-1,201	,230	°
Nombre de vaches	-0,013	-,006	-,330	,741	°
Possession d'une charrette	3,029	,021	1,529	,126	°
Possession d'une charrue	1,573	,022	1,427	,154	°

### 3. ESTIMATION DES TAUX DANS UNE PERSPECTIVE BIOGRAPHIQUE

Les modèles biographiques vont modéliser le risque de décès en l'exprimant comme une fonction de la durée d'exposition ( $t$ ) et d'un vecteur de caractéristiques  $x_t$ , individuelles ou contextuelles.<sup>23</sup> Ils prennent souvent la forme de la régression, et une option parmi d'autres consiste à recourir à la régression de Poisson. Elle est particulièrement appropriée pour les événements rares survenant au sein de larges échantillons. Largement utilisée par les épidémiologistes et les économistes (Winkelmann & Zimmermann, 1994, Cameron et Triverdi, 2005), la régression de Poisson a été utilisée dans un cadre plus démographique par Rodriguez & Cleland (1988) ainsi que par Schoumaker (2004) pour modéliser la fécondité. Timaeus & Jasseh (2004) l'utilisent pour analyser les tendances de la mortalité en Afrique sub-saharienne et évaluer l'impact de la propagation du VIH-sida. Ces auteurs insistent sur la diversification des régimes de mortalité provoquée par la propagation de l'épidémie. Ils ajustent le nombre de décès par année et par âge en

<sup>22</sup> Mais à la différence des méthodes indirectes, l'exploitation directe des données, qui est détaillée dans la section suivante, peut permettre d'éliminer les enquêtées âgées lorsque celles-ci procurent des informations significativement différentes des autres enquêtées.

<sup>23</sup> Qui peuvent être fixes (sexe, lieu de naissance, appartenance ethnique) ou varier au cours du temps (lieu de résidence, structure du ménage...).

intégrant tous les pays dans le même modèle. Nous proposons de travailler plutôt au niveau individuel. Cela permet de conserver les informations relatives à chacune des sœurs enquêtées, et donc de corriger les biais de sélection induit par la nature rétrospective des données. Cela permet également d'étendre le modèle à l'analyse multi-niveaux, afin de tenir compte de la corrélation des risques de décès au sein des fratries et des contextes. Enfin, cela autorise une exploration des différentiels à partir des caractéristiques des sœurs enquêtées.

Revenons d'abord à la régression de Poisson. Celle-ci compte parmi différentes variantes du modèle linéaire généralisé.<sup>24</sup> Elle se caractérise par trois éléments (McCullagh & Nelder, 1989) : (1) d'abord par l'hypothèse selon laquelle la distribution des événements suit une loi de Poisson dont la moyenne est notée  $\mu_i$ , (2) ensuite par une « fonction de lien » de type logarithmique telle que  $\ln(\mu) = \eta$ , (3) et enfin par un « prédicteur linéaire ». La composante *aléatoire* est modélisée à partir de la fonction de densité de la loi de Poisson, qui correspond à la probabilité d'occurrence du nombre d'événements  $y_i$ . Lorsque la durée d'exposition ( $t$ ) varie, l'expression de la distribution est la suivante:

$$P(Y_i = y_i | \mu_i, t) = e^{-t\mu_i} \frac{t\mu_i^{y_i}}{y_i!}$$

Comme  $\mu_i$  est nécessairement positif, la moyenne est exprimée sous une forme exponentielle telle que  $\mu = e^\eta$ . La variation du phénomène est ensuite modélisée sous une forme log-linéaire avec un vecteur de caractéristiques  $x_k$  et un vecteur de coefficients  $\beta_k$  :

$$\mu_i = \exp \sum_{k=1}^{k=K} \beta_k x_{ki}$$

La fonction de lien logarithmique implique que les variables ont des effets multiplicatifs sur les taux, et non plus additifs.<sup>25</sup> Cela simplifie l'interprétation des coefficients de régression puisque leur exponentielle constitue des rapports de taux. Cette expression renvoie à la composante *systématique*, qui dépend des caractéristiques introduites, tandis que la composante aléatoire suit une loi de Poisson. La probabilité de chaque occurrence est supposée indépendante du temps écoulé depuis la dernière occurrence (c'est l'hypothèse d'*indépendance*).<sup>26</sup> 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 (*équidispersion*). Lorsque la variance est supérieure à la moyenne, les données présentent un phénomène appelé « sur-dispersion », ce qui va compromettre les inférences au même titre que l'hétéroscédasticité présente dans un modèle linéaire. Comme il y a plus de variation que ne l'autorise le modèle, les erreurs standards risquent d'être sous-estimées.<sup>27</sup> Par prudence, nous adoptons un modèle de quasi-Poisson, qui est moins contraignant, et permet de tenir compte de la surdispersion. Un paramètre de dispersion variable est alors introduit sous une forme multiplicative ( $Var(Y) = \phi\mu$ ), et sera estimé au cours de l'ajustement du modèle. Les coefficients de régression restent identiques, seule leur erreur standard va varier.

Dans sa version simplifiée, le modèle implique également que les événements surviennent à un taux invariable dans le temps (c'est l'hypothèse d'*homogénéité*). En d'autres termes, le nombre d'occurrences au sein d'un intervalle de temps est supposé proportionnel à la longueur de cet intervalle. Or, les durées d'exposition sont variables, et une fonction de la durée d'exposition de chaque observation doit donc être introduite comme variable indépendante. Comme la fonction de lien est logarithmique, c'est le logarithme de la durée d'exposition qui est intégré. Pour conserver un risque proportionnel à la durée, son coefficient est toujours fixé à l'unité. Cette variable, appelée offset, ne sera pas testée dans le modèle.

$$\mu_i = t_i \exp \sum_{k=1}^{k=K} \beta_k x_{ki} \Leftrightarrow \ln \mu_i = \ln t_i + \sum_{k=1}^{k=K} \beta_k x_{ki}$$

<sup>24</sup> Le GLM étend le modèle linéaire au-delà de la linéarité et introduit des distributions non-normales.

<sup>25</sup> Puisque l'exponentielle d'une somme correspond au produit des exponentielles des différents termes de cette somme.

<sup>26</sup> La régression de Poisson repose donc sur une distribution déterminée par un seul paramètre, sa moyenne.

<sup>27</sup> A l'inverse, si la variance réelle est inférieure à la moyenne, les erreurs standards peuvent être sur-estimées. A partir des données appariées de l'enquête SACM, regroupées par fratrie et groupes d'âge, cet indice s'élève à 1.22, ce qui traduit une surdispersion, mais il chute à 1.01 lorsque les personnes-années sont conservées telles quelles. Cfr. note méthodologique 1.

La condition d'homogénéité suppose que les variables explicatives conservent le même effet durant toute la durée, ce qui réduit la richesse des informations. Une solution consiste à transformer le fichier « individu » en fichier « personnes-périodes » dans lequel chaque année qui a été vécue (partiellement ou totalement) par un frère ou une sœur mentionné dans l'enquête constitue une observation. Les parcours individuels sont ainsi découpés en différentes séquences, en précisant la durée d'exposition correspondante. Une variable qui fait référence au groupe d'âge quinquennal est conservée pour chaque observation (Schoumaker, 2004).<sup>28</sup>

Avant d'ajuster ce modèle sur les données d'enquêtes, rappelons que plusieurs sœurs peuvent répondre aux enquêtes EDS et certaines fratries sont comptabilisées plusieurs fois. Combinées à la nature rétrospective des données, cette répétition des fratries introduit un biais dans les niveaux estimés et enfreint l'hypothèse d'indépendance des observations. Ces deux types de biais nous amène à notre seconde question : quelle est la validité des estimations obtenues des modules EDS classiques ?

#### 4. QUELLE EST LA VALIDITE DES ESTIMATIONS OBTENUES DES MODULES EDS CLASSIQUES ?

Les données sur la survie des frères et sœurs collectées dans les enquêtes EDS présentent plusieurs avantages par rapport à celles qui concernent la survie des parents ou du premier conjoint. Comme la survie d'une enquêtée adulte est relativement indépendante de la survie de ses frères et sœurs, les biais de sélection restent limités, ce qui est un avantage certain dans les contextes perturbés par le VIH-sida où les risques d'infection sont importants (entre conjoints ainsi qu'entre mères et enfants). Les erreurs d'échantillonnage sont également réduites car chaque répondant témoigne de la survie d'autant de personnes soumises au risque qu'il a de frères et sœurs. Néanmoins, en raison de la nature rétrospective des données et de la possibilité d'interroger plusieurs sœurs dans la même enquête, des biais de sélection sont introduits. Les diverses formes de dépendance entre la survie d'un individu et la mortalité de ses sœurs (ou encore la taille de sa fratrie) vont intervenir et biaiser les coefficients de régression, et par là, les *niveaux* estimés. Les *écarts-types* de ces coefficients de régression seront également entachés d'erreurs du fait de l'hétérogénéité non contrôlée induite par la répétition des fratries. Nous traitons ces deux problèmes successivement à partir de l'enquête SACM.

##### 4.1. Biais sur les niveaux de mortalité

Les données EDS sont caractérisées par une sous-représentation des fratries à mortalité élevée : une fratrie de cinq sœurs dont trois survivent a plus de chances d'être comptabilisée qu'une fratrie de cinq sœurs dont deux survivent. Un biais est également introduit du fait qu'aucune information n'est donnée pour les fratries dont personne ne survit. Enfin, lorsque les répondants sont exclus du fichier, l'information donnée par le fait qu'ils sont nécessairement survivants est négligée, ce qui va biaiser les estimations à la hausse. On peut se demander comment, dans ces conditions, il est encore possible d'obtenir des estimations valides. En fait, Trussel & Rodriguez (1990) ont prouvé que ces trois biais s'annulaient mathématiquement lorsque la mortalité était indépendante de la taille de la fratrie: "unrestricted random sampling from a population of sibships, which requires interviewing all sisters who fall in the sample, guarantees that the three potential sources of error cancel each other exactly, provided that the mortality experiences of siblings are drawn independently from the same distribution."<sup>29</sup> A partir de là, seule une mineure correction est nécessaire. En effet, l'indépendance entre la taille de la fratrie et la mortalité est rarement rencontrée dans la pratique. Cette association entre taille de la fratrie et mortalité introduit un effet de sélection, qui peut entraîner tant une sous-estimation qu'une sur-estimation de la mortalité, selon le signe de cette association. Si la mortalité et la fécondité sont positivement associées, les estimations seront biaisées à la hausse, et vice-versa. Dans les pays

<sup>28</sup> Le fait de travailler par classe d'âge revient à supposer que les risques de mortalité sont identiques au sein d'un même groupe d'âge mais varient d'un groupe d'âge à l'autre (modèle exponentiel constant sur intervalles réduits). Une fois le fichier scindé en personnes-années, il faut donc subdiviser en deux parties les années calendrier lorsque les frères et sœurs passent d'un groupe d'âge quinquennal à l'autre, en fonction de leur date d'anniversaire. Comme la somme de deux variables qui suivent une distribution de Poisson suit également une distribution de Poisson, et que le paramètre de la loi résultante est égal à la somme des deux paramètres des lois qui la composent, décomposer les unités d'observations en années ou ajuster le modèle sur des données groupées procurera les mêmes coefficients de régression, pour peu que la durée d'exposition soit contrôlée.

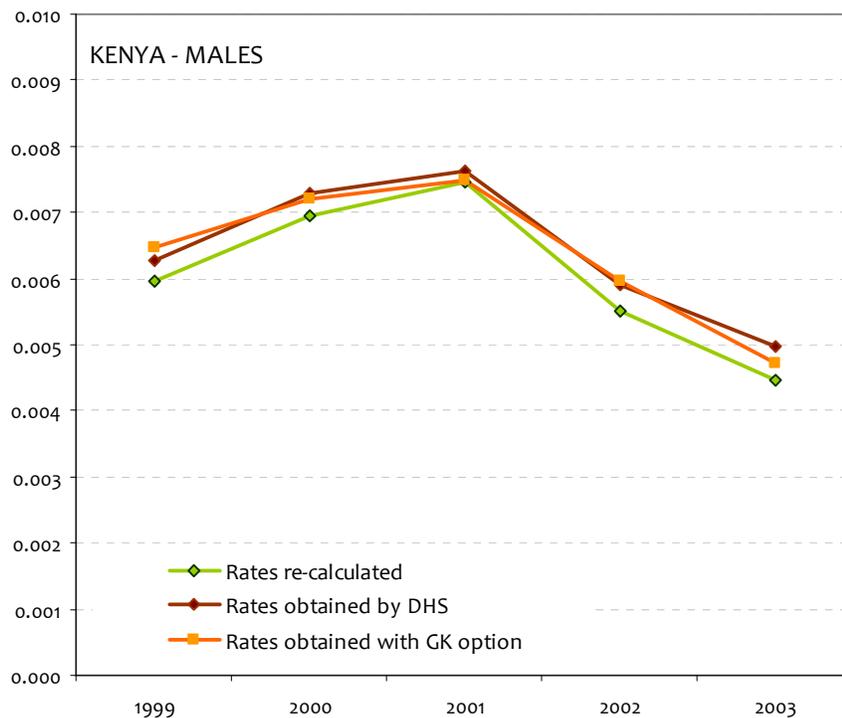
<sup>29</sup> Mais c'est le cas uniquement si elles peuvent toutes témoigner de la survie des autres membres de la fratrie. La mortalité est sur-estimée lorsqu'une seule personne est autorisée à répondre au module (comme ce fut le cas pour l'enquête EDS menée en Bolivie en 1990).

en développement, les risques de mortalité d'un enfant sont positivement associés au nombre de ses frères et sœurs (Zaba & David, 1996)<sup>30</sup>. Plus simplement, les fécondités élevées sont souvent le cas de familles plus pauvres aux risques de mortalité plus élevés.

Gakidou & King (2006) ont récemment repris cette question d'une façon très stimulante en pointant le fait que les estimations sont proches du niveau réel lorsque les effets de cette association entre mortalité et taille de la fratrie peuvent être neutralisés. Une première correction consiste à établir des poids en fonction du nombre de sœurs *présentes en début de période* et du nombre de sœurs  *survivantes à l'enquête*. Tous les membres de la fratrie (répondants y compris) doivent être intégrés dans les calculs.<sup>31</sup> A l'exception des fratries où aucune sœur n'a survécu, ils montrent par simulation que ce recours à la pondération résout le problème du biais de sélection, même en levant l'hypothèse d'indépendance entre taille de la fratrie et mortalité. En cas de calcul direct, cela revient à pondérer les observations par l'inverse du nombre de sœurs qui auraient pu répondre à l'enquête, en prenant soin d'intégrer les enquêtées dans le calcul.<sup>32</sup>

Les résultats des micro-simulations indiquent toutefois que cette estimation tend à sous-évaluer la mortalité féminine.

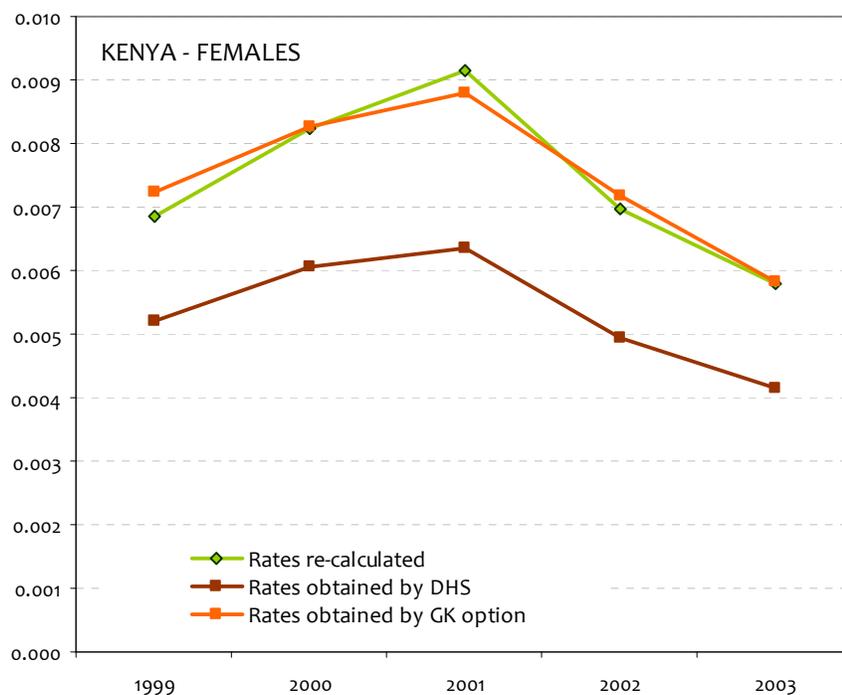
(A commenter)



<sup>30</sup> Ce phénomène peut s'expliquer par une plus grande compétition entre frères, notamment pour les ressources nutritives, ainsi que par un certain épuisement des femmes, ou encore par une transmission facilitée des maladies infectieuses. Ces mécanismes peuvent avoir des effets qui se prolongent aux âges adultes.

<sup>31</sup> Les auteurs suggèrent de pondérer les proportions de décédées calculées au sein des fratries ( $M_i$ ) par le rapport ( $W_i$ ) entre le nombre de survivants ( $S_i$ ) et le nombre personnes soumises au risque au départ ( $B_i$ ) :  $\sum M_i W_i / \sum W_i$ .

<sup>32</sup> Puisque la formule  $\sum M_i W_i / \sum W_i$  se simplifie en  $\sum (D_i / S_i) / \sum (B_i / S_i)$ . Les poids correspondent donc à l'inverse du nombre de femmes survivantes et âgées à l'enquête entre 15 et 49 ans. Pour conserver la taille de l'échantillon, ces poids sont multipliés par la taille de l'échantillon et divisés par la somme des poids. Notons que le système de poids devrait idéalement être légèrement revu pour la mortalité masculine.



#### 4.2. Biais sur les écarts-types des coefficients de régression

L'appariement des déclarations est surtout utile pour aborder une seconde source de biais, qui est particulièrement problématique pour l'analyse des différentiels. En effet, le fait que les fratries peuvent être comptabilisées plusieurs fois dans le fichier contredit l'hypothèse d'indépendance des observations, et invalide le recours à des procédés statistiques classiques. Par rapport à ce qui est prédit sous l'hypothèse d'indépendance, il y a davantage de fratries avec peu de survivants et davantage de fratries sans décédés. Les erreurs standards vont être sous-estimées, et les intervalles de confiance réduits. Un modèle à variance composée permet d'estimer cette variance imputable au niveau des fratries, en intégrant un effet  $u_i$  commun à tous les membres d'une même fratrie. Comme la valeur de cet effet est inconnue, elle est fixée aléatoirement avec pour seule contrainte d'être issue d'une distribution normale de moyenne nulle.

Il nous faut toutefois distinguer cette hétérogénéité imputable à la structure des données et l'hétérogénéité qui serait imputable à une corrélation « réelle » des risques de décès au sein des fratries. Il y a en effet des raisons plus substantielles qui peuvent expliquer la corrélation des risques de décès, tel le fait qu'une certaine vulnérabilité biologique peut être partagée par les différents frères et sœurs. En d'autres termes, le capital santé initial peut ne pas être réparti équitablement parmi les fratries. En absence de mobilité sociale élevée, le fait que la condition sociale se reproduise d'une génération à l'autre va également entraîner une similarité des profils entre frères et sœurs, et donc une concentration des risques au sein de certaines fratries.<sup>33</sup> Les situations vécues dans l'enfance, partagées par les frères et sœurs, peuvent à leur tour avoir des impacts sur les situations vécues à l'âge adulte.<sup>34</sup>

Quelle est cette variance « réelle » relative au niveau de la fratrie ? Quelle est son importance par rapport à la variance induite par la structure des modules EDS ? Faut-il en tenir compte dans l'analyse des différentiels ou non ? Pour l'estimer et la comparer à la variance observée dans les fichiers EDS classiques, il est à nouveau utile d'exploiter l'enquête SACM de deux façons : tantôt de façon classique (avec répétitions des fratries, comme

<sup>33</sup> C'est l'hypothèse du *pathway model* utilisée pour expliquer les inégalités sociales de santé

<sup>34</sup> Soit directement (carences alimentaires...), soit indirectement, si l'on considère par exemple que le niveau scolaire ou le statut socio-professionnel atteints résultent d'un enchaînement d'opportunités offertes ou non aux enfants. C'est plutôt l'hypothèse appelée *life-course model*.

les autres enquêtes EDS), et tantôt de façon à tenir compte de l'appariement. Nous ajustons donc un modèle à variance composée minimal (n'intégrant que l'âge et la durée d'exposition) sur trois types de fichiers : d'abord sur les données non appariées et non pondérées de l'enquête SACM (1), ensuite sur les données appariées (2), et enfin sur les données pondérées (3).<sup>35</sup> Les résultats sont présentés dans le tableau 4 (pour la période 1986-95).

Lorsque le modèle est appliqué au fichier exploité de façon classique, la variance contextuelle s'avère élevée (1.496) et significative au seuil 0.01. Par contre, à partir du fichier apparié, la variation contextuelle estimée est bien plus faible (0.256). Elle pourrait refléter la concentration des désavantages au sein de certaines fratries, mais elle n'apparaît pas significative. Les données ne font donc pas ressortir d'inégalités des chances en matière de santé qui seraient transmises d'une génération à l'autre et partagées au sein des fratries. La variation élevée au niveau des fratries obtenue lors d'une exploitation classique semble donc essentiellement induite par la structure du fichier EDS. Elle est également assez élevée dans le cas des données pondérées (elle est significative au seuil 0.15).<sup>36</sup> Cela devrait encore être confronté aux résultats des micro-simulations.

Tableau 4 : Modèle à variance composée (avec effet aléatoire par fratrie) sur les données de l'enquête SACM sans pondération ni appariement (1), après appariement (2) et après pondération (3).

	Exploitation classique sans pondération		Données appariées		Données pondérées (avec répondants)	
	Coef.	E.S	Coef.	E.S	Coef.	E.S
Constante	-6.56	0.30	-6.76	0.29	-6.98	0.32
20-24	0.86	0.36	0.77	0.35	1.18	0.39
25-29	0.94	0.36	1.15	0.34	1.29	0.41
30-34	1.25	0.37	1.36	0.35	1.64	0.40
35-39	1.77	0.36	1.52	0.36	2.05	0.39
40-44	1.86	0.40	1.68	0.48	1.81	1.43
45-49	0.99	0.59	1.07	0.64	1.31	0.66
$\sigma_{u0}^2$	1.496	0.384	0.256	0.356	2.182	1.344

## 5. QUEL EST L'APPORT DES MODULES EDS POUR EXPLORER LES DIFFÉRENTIELS ?

Nous arrivons à notre dernière question : celle de l'apport des enquêtes EDS pour l'analyse des différentiels de mortalité adulte. Selon Brass (1979): « the use of retrospective reports on mortality beyond childhood for the study of differentials is almost unexplored. This is unfortunate since the prospects of useful returns are by no means unpromising although there are obvious difficulties.»<sup>37</sup> Parmi ces difficultés, on peut compter le fait que les techniques indirectes, qui ont longtemps été la référence pour la mortalité adulte, sont peu appropriées pour analyser les différentiels. Mais même en optant pour l'exploitation directe, il subsiste un problème sérieux : puisque les caractéristiques des personnes soumises au risque sont inconnues, elles doivent être approximées par celles de leurs proches (sœurs, enfants, etc.).<sup>38</sup> Certaines caractéristiques sont toutefois suffisamment stables et partagées dans l'espace familial pour pouvoir être considérées comme identiques au sein des fratries. C'est le cas de l'appartenance ethnique, puisque les enfants héritent le plus souvent de l'ethnie de leurs parents. L'analyse des disparités interethniques de mortalité adulte peut donc être menée à l'échelle nationale à partir des enquêtes EDS. A ce stade, l'analyse ne peut nécessairement que rester sommaire, puisque seuls l'âge, le sexe et l'évolution de la mortalité dans le temps peuvent être

<sup>35</sup> Les estimations sont faites à partir du logiciel MLWIN.

<sup>36</sup> La question des pondérations en analyse multi-niveaux est assez récente et nécessite d'interpréter les résultats avec prudence. Comme les poids ne sont pas indépendants de la variation aléatoire, un biais est introduit. A ce stade, nous n'entrons pas dans de tels détails.

<sup>37</sup> Brass, 1979, *op. cit.*, p. 34

<sup>38</sup> « The socioeconomic conditions of the respondent will be altered from those of the parent and the differentials found will be a somewhat removed reflection of the influencing factors. » *Ibid.*

intégrés, en l'absence d'autres caractéristiques connues des personnes soumises au risque. Mais il peut être utile de documenter des inégalités inter-ethniques significatives, qui pourront être expliquées par la suite. Pour ce faire, un modèle de régression est ajusté sur les données de quasiment tous les pays pour lesquels des informations sur l'appartenance ethnique sont disponibles<sup>39</sup> (en combinant les enquêtes quand la taille des échantillons le permet). Seules les onze années précédant les enquêtes sont retenues pour limiter la sous-déclaration.<sup>40</sup> Pour tenir compte des biais de sélection, la pondération reposant sur l'inverse du nombre de sœurs survivantes (15-49) est introduite. Pour contrôler la répétition des fratries, les coefficients et les erreurs standards sont estimés à partir d'équations d'estimation généralisées. Un modèle de quasi-Poisson est choisi pour tenir compte de la surdispersion. Pour limiter la durée du calcul, les années sont regroupées par paires et les tendances sont modélisées sous forme polynomiale sur tout l'intervalle, en intégrant un effet d'interaction entre les années et la structure par âge. Les calculs sont menés sur les adultes de 15 à 50 ans. Les ethnies de référence sont systématiquement les ethnies majoritaires (selon les enquêtes EDS). Les inférences se font à partir des tests de Wald. Les intervalles de confiance (au seuil 0.05) sont repris entre crochets.

L'analyse ne met pas en évidence d'inégalités inter-ethniques de mortalité adulte significatives au Bénin (DHS-96 1985-95), ni en Centrafrique (DHS-94/95 1983-94), ou au Togo (DHS-98 1988-98). Dans les autres pays, des inégalités apparaissent, et parfois de façon très prononcée, comme au Kenya, en Ethiopie ou au Nigeria. De façon générale, l'ethnie majoritaire est favorisée, bien que certains pays se distinguent par la présence de quelques ethnies minoritaires avantagées (Afrique du sud, Gabon, Tchad).

Les résultats observés recourent certains constats relatifs à la mortalité des enfants. Au Kenya, Tabutin & Akoto (1989) notent que les risques de décès avant deux ans étaient deux fois plus élevés chez les Luos que chez les Kikuyus. Les Luos (ainsi que les Luhyas) apparaissent ici également très désavantagés aux âges adultes, avec des taux plus de deux fois supérieurs aux Kalenjins ou aux Kikuyus (toutes choses égales par ailleurs). Dans une étude plus ancienne, Hill & Randall (1984) montraient que la mortalité des enfants Peuls de moins de deux ans était de 10% supérieure à celle des Bambaras, et l'écart observé ici est encore plus creusé.<sup>41</sup> Brokerhoff et Hewett (2000) documentent quant à eux un avantage des Bembas en Zambie, que l'on retrouve ici. D'autres groupes ethniques, tels les Touaregs semblent devoir faire face à une mortalité significativement plus élevée dans différents pays, comme au Burkina Faso ainsi qu'au Niger. Les Peuls sont quant à eux désavantagés au Mali et au Sénégal (mais pas au Tchad, ni au Niger, en Guinée ou au Burkina).

Les limites de cette démarche sont évidentes. L'identité ethnique doit être reconnue dans sa dimension de construit social, et les catégories utilisées dans les enquêtes figent des catégories culturelles disparates et mouvantes, en y apposant des étiquettes ethniques globalisantes. D'un point de vue méthodologique, l'analyse devrait également prendre en compte le rôle des structures sociales qui viennent médier et donner sens à la relation entre ethnie et santé. Dans bien des cas, elle doit en outre tenir compte du fait que l'appartenance ethnique est étroitement liée à la région de résidence. Les découpages régionaux constituent souvent des ensembles ethniques relativement homogènes, et les inégalités mises en avant ici peuvent donc refléter des disparités régionales de mortalité.

Tableau 5 : Inégalités interethniques de mortalité adulte à partir des enquêtes EDS

Pays	Ethnies de référence	Autres ethnies	Période de référence
	Ethnies significativement favorisées ou défavorisées		
Afrique du sud DHS- 1998	<i>Ethnie de référence</i> Noirs-africains  <i>Ethnies sign. favorisées</i>	« Colored », Asiatiques	1987-97

<sup>39</sup> Et où les groupes ethniques sont en nombre réduit – cfr. note méthodologique 3.

<sup>40</sup> Sauf pour le Niger où ce seuil est élevé à 15 ans afin de pouvoir combiner les enquêtes de 1992 et de 2006. Comme les fonctions splines sont assez sensibles aux outliers, nous éliminons la première ou la dernière année lorsque le nombre d'observations est très inférieur à la moyenne des autres années.

<sup>41</sup> Par contre, ils documentaient un avantage des Peuls dans la région de Sine-Saloum, au Sénégal, alors que la mortalité de ces derniers semble ici significativement plus élevée que celle de la majorité wolof.

	Blancs	0.59 [0.42 - 0.83]		
Burkina Faso DHS-03 DHS-98/99	<i>Ethnie de référence</i> Mossis		Bobos, Dioulas, Fulfuldes/Peuls, Gourmantchés, Gourounsis, Sénoufos, Lobis, autres ethnies	1987-03
	<i>Ethnies sign. défavorisées</i> Touaregs	1.67 [1.07 - 2.60]		
Ethiopie DHS-04	<i>Ethnie de référence</i> Amara		Tigraway, Oromo, Afar, autres ethnies minoritaires	1993-04
	<i>Ethnies sign. favorisées</i> Somali Sidama Gurage	0.78 [0.63 - 0.96] 0.71 [0.53 - 0.95] 0.67 [0.54 - 0.84]		
	<i>Ethnies sign. défavorisées</i> Anuak Weleyta	2.20 [1.71 - 2.83] 1.51 [1.17 - 1.95]		
Gabon DHS-00	<i>Ethnie de référence</i> Fang		Mbede-Teke, Myene, Nzabi- Duma, Okande-Tsogho	1989-00
	<i>Ethnies sign. favorisées</i> Kota-Kele Shira-Punu/Vili Pygmées et autres minorités	0.60 [0.40 - 0.88] 0.76 [0.60 - 0.97] 0.75 [0.58 - 0.96]		
Guinée DHS-99 DHS-05	<i>Ethnie de référence</i> Peulh		Malinké, Toma, Guerze, Kissi, autres ethnies minoritaires	1988-05
	<i>Ethnies sign. défavorisées</i> Soussou	1.20 [1.04 - 1.39]		
Kenya DHS-98 DHS-03	<i>Ethnie de référence</i> Kalenjin		Kamba, Kikuyu, Meru/Embu, Somali, Taita/Tavate, autres ethnies	1987-03
	<i>Ethnies sign. favorisées</i> Masai	0.66 [0.49 - 0.91]		
	<i>Ethnies sign. défavorisées</i> Kisii Luhya Luo Mijikenda/Sw.	1.27 [1.15 - 1.40] 2.56 [2.17 - 3.02] 2.30 [2.09 - 2.54] 1.37 [1.12 - 1.68]		
Malawi DHS-04	<i>Ethnie de référence</i> Chewa		Sena, Nkonde, Ngoni	1993-04
	<i>Ethnies sign. favorisées</i> Tumbuka	0.69 [0.58 - 0.81]		
	<i>Ethnies sign. défavorisées</i> Lomwe Tonga Yao Autres ethnies minoritaires	1.51 [1.36 - 1.51] 1.72 [1.37 - 2.17] 1.23 [1.09 - 1.38] 1.23 [1.06 - 1.43]		
Mali DHS-95 DHS-01	<i>Ethnie de référence</i> Bambaras		Malinkés, Sarakolés/Soninkés, Dogons, Bobos, autres ethnies minoritaires	1984-04
	<i>Ethnies sign. favorisées</i> Senouf./Minianka	0.83 [0.68 - 0.99]		
	<i>Ethnies sign. défavorisées</i> Peulhs Sonrai	1.24 [1.08 - 1.42] 1.55 [1.33 - 1.81]		

	Tamashek	1.53 [1.25 - 1.88]		
Niger DHS-06 DHS-92	<i>Ethnie de référence</i> Haoussas		Arabes, Kanouris, Toubous, Peuls	1980-05
	<i>Ethnies sign. défavorisées</i>			
	Djerma/Songhai	1.36 [1.17 - 1.58]		
	Gourmantché	1.47 [1.19 - 1.82]		
	Touaregs	1.31 [1.09 - 1.58]		
	Autres ethnies minoritaires	1.61 [1.22 - 2.13]		
Nigeria DHS-99	<i>Ethnie de référence</i> Haoussas		Yoruba, Tiv, Nupe, Igbala, Ibara, Ijaw, Ogoni, Autres ethnies du Centre/Ouest, bibio, Autres ethnies du Nord, autres ethnies du Sud & Est, autres ethnies minoritaires	1988-98
	<i>Ethnies sign. défavorisées</i>			
	Fulani	2.85 [1.83 - 4.44]		
	Igbo	1.47 [1.01 - 2.13]		
	Edo	2.09 [1.14 - 3.88]		
	Annang	2.71 [1.36 - 5.38]		
Sénégal DHS-92 DHS-05	<i>Ethnie de référence</i> Wolofs		Serers, Sarakolé/Soninkés, Mandingues, Diolas	1981-04
	<i>Ethnies sign. défavorisées</i>			
	Peulhs	1.15 [1.02 - 1.31]		
	Etrangers	1.49 [1.07 - 2.06]		
Tchad DHS-04 DHS-96/97	<i>Ethnie de référence</i> Sara		Gorane, Kanem-Bornou, Arabe, Baguirmien, Hadjarai, Fitri-Batha, Tandjile, Lac Iro, Peulh, autres ethnies minoritaires	1985-04
	<i>Ethnies sign. favorisées</i>			
	Ouaddai	0.75 [0.62 - 0.91]		
	Kebbi	0.62 [0.51 - 0.77]		
Zambie DHS-02	<i>Ethnie de référence</i> Bembas		Tonga, Lunda, Kaonde, Lunda, Luvale, Ushi, Mbunda, Namwanga, Lenje, Bisa, Ngumbo, Tabwa, Ila, Chishinga, Senga, Mambwe, Chewa, autres ethnies minoritaires	1990-01
	<i>Ethnies sign. favorisées</i>			
	Tumbuka	0.74 [0.56 - 0.95]		
	Luchazi	0.32 [0.17 - 0.62]		
	<i>Ethnies sign. défavorisées</i>			
	Lala	1.49 [1.21 - 1.84]		
	Lamba	1.43 [1.11 - 1.84]		
	Lozi	1.31 [1.10 - 1.55]		
	Nsenga	1.31 [1.09 - 1.57]		
	Ngoni	1.23 [1.01 - 1.50]		

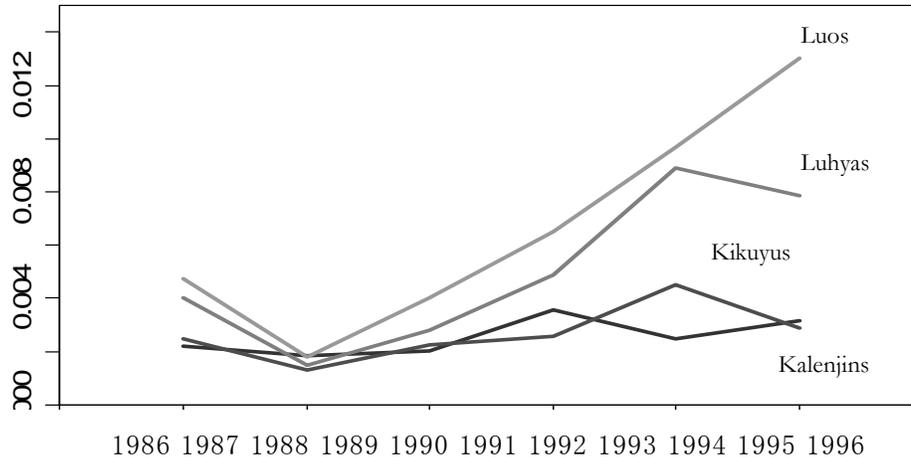
Une démarche d'appariement des fratries devrait être développée afin de mieux estimer les erreurs standards.

En outre, une méthodologie permettant d'expliquer ces inégalités interethniques en combinant des caractéristiques individuelles et des caractéristiques contextuelles devrait être développée. Cela suppose de faire des hypothèses sur le caractère différencié de la mobilité sociale et géographique au sein des fratries (puisque les frères et sœurs n'habitent pas nécessairement dans la même zone, par exemple). Cela suppose également de recourir à des modèles multi-niveaux, puisqu'au niveau de la fratrie doit venir s'ajouter un niveau relatif au contexte de vie. A partir de là, la concentration des risques de décès au sein de certains contextes peut être estimée (en neutralisant la corrélation au sein des fratries). Le recours aux modèles multi-niveaux devient en quelque sorte une porte d'entrée vers l'analyse des différentiels géographiques de mortalité, qui permet de contourner partiellement les problèmes associés à l'absence d'informations sur les personnes décédées. Au Kenya, par exemple, où apparaissent les inégalités interethniques de mortalité parmi les plus prononcées, les régions ont été déterminées par l'administration coloniale britannique en recoupant la géographie des groupes ethniques. Si la minorité Masai semble avantagée, tout comme les Kalenijns, il

faut sans doute y voir un avantage associé à la région du Rift Valley, où ces deux ethnies sont dominantes. L'analyse multi-niveaux devrait permettre de mettre ce type d'hypothèses à l'épreuve.

De même, il conviendrait également de s'interroger sur l'impact de la crise sanitaire de ces dernières années sur les inégalités inter-ethniques de mortalité. Dans quelles ethnies les reprises de mortalité adulte ont-elles été les plus fortes ? Comment la crise sanitaire a-t-elle modifiée la structure des écarts ? En reprenant encore l'exemple du Kenya, on constate que les inégalités inter-ethniques ont sensiblement augmenté durant la décennie 1990 (graphe 7).

Graphe 5 : Evolution dans le temps des taux de mortalité (15-50 ans) des Kalenjins, Kikuyus, Luhyas et Luos, entre 1987 et 1998 (EDS 1998-2003)



## 6. CONCLUSION

Plusieurs raisons plaident pour qu'une plus grande attention soit accordée à la mortalité adulte en Afrique sub-saharienne. Tout d'abord, comme l'illustrent les taux du Malawi et de la Zambie présentés plus haut, la mortalité adulte enregistre par endroits des reprises importantes, parfois sans précédent. Elle est par ailleurs encore largement imputable à un nombre réduit de pathologies évitables. Enfin, elle grève de façon substantielle les ressources budgétaires consacrées aux soins de santé. A ces trois raisons bien connues s'en ajoute une quatrième ; la mortalité adulte est traversée d'inégalités très creusées qui sont encore très mal documentées. En prenant l'exemple des structures ethniques, nous avons montré qu'il était possible de mettre en avant des différentiels significatifs de mortalité adulte. L'objectif de cette communication, d'ordre plus méthodologique, n'était toutefois pas d'analyser en profondeur ces inégalités inter-ethniques. Il s'agissait essentiellement de montrer qu'il y a dans les modules d'enquêtes EDS une richesse à exploiter pour l'analyse des différentiels. Nous avons montré que la qualité des données relatives à la survie des frères et sœurs était satisfaisante pour mener une telle analyse, tout en pointant les principales sources d'erreurs. Nous avons rappelé que la combinaison de la nature rétrospective de l'enquête et du fait que les fratries peuvent être comptabilisées plusieurs fois provoquait une sous-représentation des fratries à mortalité élevée. L'hétérogénéité au niveau des fratries induite par leur répétition dans le fichier doit en outre être contrôlée en étendant le modèle dans un cadre multi-niveaux ou en optant pour des équations d'estimation généralisées. A partir de là, il est possible de mener une analyse des inégalités de mortalité adulte. Il reste toutefois à développer une méthodologie permettant de faire intervenir des caractéristiques individuelles et contextuelles à partir des informations relatives aux enquêtées. L'analyse multi-niveaux constitue sans doute une piste de recherche à creuser en ce sens.

## 7. BIBLIOGRAPHIE

Akoto E., 1993, *Déterminants socio-culturels de la mortalité des enfants en Afrique noire*, Louvain-la-neuve, Academia-Bruylant, 240 p.

- Akoto E. & D. Tabutin, 1989, « Les inégalités socio-économiques et culturelles devant la mort », in Pison G., van de Walle, E. Sala-Diakanda M. (ed.), *Mortalité et santé en Afrique au sud du Sahara*, Paris : INED/PUF, pp. 35-63
- Basen J. & Moradi A., 2005, « Inequality in sub-saharan Africa ; new data and new insights from anthropometric estimates », *World Development*, vol. 33, n°8, pp. 1233-1265
- Bicego G., 1997, “Estimating adult mortality rates in the context of the AIDS epidemic in sub-Saharan Africa: analysis of DHS sibling histories”, *Health Transition Review*, 7, 2, pp. 7-22
- Bicego G., Curtis S., Raggars H., Kapiga S. & Ngallaba S., 1997, *Summe 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, Demographic and Health Surveys, Calverton, Maryland, 92 pp.
- Blacker J. & al., 1987, “Mortality differentials in Kenya”, *International Union for the scientific study of population seminar on Mortality and society in Sub-Saharan Africa*, Yaoundé
- Brass W., 1979, “Some methodological problems of the analysis of mortality differentials”, in *Proceedings of the Meeting on socio-economic determinants and consequences of Mortality, Mexico*, New York: World Health Organization
- Brass W. & K. Hill, 1973, “Estimating adult mortality from orphanhood”, in *International Population Conference, Liège, 1973*, 3, Liège, International Union of the scientific study of Population, pp.111-123.
- Brass W. & E. Bamgboye, 1981, “The time location of reports of survivorship: estimates for maternal and paternal orphanhood and the ever-widowed”, *Working paper n°81-1*, London: Centre for Population Studies, London School of Hygiene et Tropical Medecine
- Brockerhoff M. & P. Hewett, 2000, “Inequality of child mortality among ethnic groups in sub-Saharan Africa”, *Bulletin of the World Health Organization*, 78, 1, pp. 30-41
- Cameron A. & Trivedi K., 1998, *Regression Analysis of Count Data*, Econometric Society Monograph, No.30, Cambridge University Press, 432 p.
- Cameron A. & Trivedi K., 2005, *Microeconometrics: Methods and Applications*, Cambridge University Press
- Dorrington R., & al., 2004, « Estimates of provincial mortality and fertility in South Africa 1985-1996 », *Southern African Journal of Demography*, 9, 2, pp. 25-57
- Easterly W. & Levine R., 1997, “Africa’s growth tragedy: policies and ethnic divisions”, *The Quarterly Journal of Economics*, 112, 4, pp. 1203-1250
- Feachem R., & al. (ed.), 1992, *The Health of adults in the developing world*, New York, Oxford University Press
- Feeney G., 2001, “The impact of HIV/AIDS on adult mortality in Zimbabwe”, *Population and Development Review*, 27, 4, pp. 771-780
- Fellegi I. & A. Sunter, 1969, “A Theory for Record Linkage,” *Journal of the American Statistical Association*, 64, pp. 1183-1210
- Gakidou E., M. Hogan & A. Lopez, 2004, “Adult mortality: time for a reappraisal,” *International Journal of Epidemiology*, 33, 4, pp. 710–717
- Gakidou E. & King G., 2006, “Death by survey, Estimating adult mortality without selection bias”, *Demography*, 43, 3, pp. 569-585
- Garenne M., 2002a, « Différences entre filles et garçons : une revue des indicateurs sanitaires en Afrique subsaharienne. », *Les dossiers du Ceped*, 70, 37 p.
- Garenne M., 2002b, « Sex ratios at birth in African populations : a review of survey data », *Human Biology*, 74.
- Graham W., W. Brass & R. Snow, 1989, « Estimating maternal mortality : the sisterhood method », *Studies in Family Planning*, 20, 3, pp. 125-135
- Hill A. & Randall S., 1984, « Différences géographiques et sociales dans la mortalité infantile et juvénile au Mali », *Population*, 39, 6, pp. 921-946
- Hill K., 1977, “Estimating adult mortality levels from data on widowhood”, *Population Studies*, 31, 1, pp. 75-84
- Hill K. & J. Trussel, 1977, “Further developments in indirect mortality estimation”, *Population Studies*, 31, 2, pp. 313-334
- Loomis D., D. Richardson, L. Elliot, 2005, “Poisson regression analysis of ungrouped data”, *Occupational and environmental medicine*, 62, p. 625-629
- Liang, K. & Zeger, S., 1986, “Longitudinal data analysis using generalized linear models”, *Biometrika*, 73, 13-22.
- Powers D. & Xie Y., 2000, *Statistical methods for categorical data analysis*, Academic Press, 305 p.
- Prentice R. & Zhao L., 1991, “Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses”, *Biometrics*, 47, p. 825-839
- McCullagh P. & Nelder J. A., 1989, *Generalized Linear Models*, London, Chapman & Hall/CRC
- Nations Unies, 1983, *Manuel x. Indirect techniques for demographic estimation*, New York : Nations Unies, 304 p.
- Palloni A., 1981, “Design problems and data collection strategies in studies of mortality differentials: developing countries”, in Vallin J., & al. (ed.), *Methodologies for the collection and analysis of mortality data*, Liège: Ordina editions

- Phillips M., Feachem R., Murray C., Over M. & Kjellstrom T., 1993, "Adult health: a legitimate concern for developing countries", *American Journal of Public Health*, 83, 11, pp. 1527-1530
- Rodriguez G. & Cleland J., 1988, « Modeling marital fertility by age and duration: an empirical appraisal of the Page model », *Population Studies*, 42, 2, pp. 241-257
- Rutenber, N. & Sullivan J., 1991, "Direct and indirect estimation of maternal mortality from the sisterhood Method", in *Proceedings of the Demographic and Health Surveys World Conference, Washington, D.C., August 5-7, 1991*, Vol. 3. Columbia, Maryland: IRD/Macro International, pp.1669-1695
- Schoumaker B., 2004, « Une approche personnes-périodes pour l'analyse des histoires génésiques », *Population*, 59, 5, pp. 783-796
- Scott L. Zeger; Kung-Yee Liang, "Longitudinal Data Analysis for Discrete and Continuous Outcomes", *Biometrics*, Vol. 42, No. 1. (Mar., 1986), pp. 121-130
- Stanton C., A. Nouredine & K. Hill, 2000, "DHS Maternal Mortality Indicators: an assessment of data quality and implications for Data Use", *Studies in Family planning*, 31, 2, pp.111-123
- Tabutin D., 1995, « Transitions et théories de mortalité », in Gérard H. & V. Piché (ed.), *La sociologie des populations*, Montréal : Presses de l'Université de Montréal, pp. 257-288
- Timaues I., 1991, "Estimation of Mortality from Orphanhood in Adulthood," *Demography*, 28, 2, pp. 213-227
- Timaues I., 1992, "Estimation of adult mortality from paternal orphanhood: a reassessment and a new approach", *Population Bulletin of the United Nations*, n°33, pp. 47-63
- Timaues I. & A. Nuun, 1997, "Measurement of adult mortality in populations affected by AIDS: an assessment of the orphanhood method", *Health transition Review*, 7, 2, pp.23-43
- Timaues I., 1999, « Mortality in Sub-Saharan Africa », in Chamie J. & R. Cliquet (ed.), *Health and mortality, issues of global concern, Proceedings of the symposium on health and mortality*, New York: United Nations and Population and Family Study Centre, pp. 110-131
- Timaues I., B. Zaba & M. Ali, 2001, "Estimation of adult mortality from data on adult siblings", in Zaba B. & J. Blacker (dir.), *Brass tacks: essays in medical demography*, Londres : Athlone Press, pp. 43-66
- Timaues I. & Jasseh M., 2004, "Adult mortality in sub-saharan africa: evidence from demographic and health survey", *Demography*, 41, 4, pp. 757-772
- Trussel J. & Roriguez G., 1990, « A note on the sisterhood estimator of maternal mortality », *Studies in Family Planning*, 21, 6, pp.344-346
- Wedderburn R., 1974, « Quasi-likelihood functions, Generalized Linear Models, and the Gauss-Newton Method », *Biometrika*, 61, 3, pp. 439-447
- Winkelmann R. & Zimmermann K., 1994, "Count data models for demographic data", *Mathematical Population Studies*, 4, 3, pp. 205-221
- Zaba B. et David P., 1996, « Fertility and the distribution of child mortality risk among women: an illustrative analysis », *Population studies*, 50, 2, pp. 263-278

## 8. NOTES MÉTHODOLOGIQUES

### Note méthodologique 1 : indice de dispersion

Afin d'établir si l'équidispersion est rencontrée dans la pratique, un indice de dispersion peut être estimé à partir de la somme des carrés des résidus de Pearson. Comme cette statistique de Pearson est supposée suivre une distribution de  $\chi^2$ , l'estimation  $\hat{\varphi}$  s'obtient en divisant cette statistique par le nombre de degrés de libertés (c'est-à-dire le nombre d'observations moins le nombre de coefficients dans le modèle). Idéalement, ce rapport doit approcher l'unité. Il existe une estimation alternative de  $\varphi$  basée sur le rapport entre la déviance. Ces deux estimateurs sont asymptotiquement égaux, mais dans la pratique, ils procurent souvent des résultats distincts, surtout sur des données éparses (voir Venables & Ripley, 2002, *Modern Applied Statistics with S. Fourth Edition*, Springer., Boyle P., & al., 1997, "Evaluating the goodness of fit in models of sparse medical data: a simulation approach", *International Journal of Epidemiology*, 26, pp. 651–656, ou Heinzl H., Mittlböck M., 2003, "Pseudo R-squared measures for Poisson regression models with over- or underdispersion", *Computational Statistics & Data Analysis*, 44, 1-2, pp. 253-271)

### Note méthodologique 2 : modèle de Poisson p. 15

L'interaction entre l'âge et la durée est introduite sur toute la période analysée, à la différence du modèle de Timaeus & Jasseh (2004). Dans leur modèle, Timaeus & Jasseh (2004) autorisent les niveaux de mortalité par groupe d'âge ( $x$ ) et sexe ( $g$ ) à varier à la fois en fonction du pays et à des rythmes distincts en fonction du temps. Cela dit, les modifications dans la structure par âge de la mortalité sont identiques dans tous les pays concernés. L'ampleur des écarts par rapport au standard varie au sein des pays avec la durée de l'épidémie, mais pas d'un pays à l'autre, ce qui revient à supposer que la hausse de mortalité modifie le schéma par âge de façon identique dans les pays touchés mais que cette hausse se produit à une vitesse variable. Dans ce cas-ci, les changements dans la structure par âge sont spécifiques à chaque pays.

Les fonctions splines cubiques consistent à introduire un certain nombre de nœuds sur l'axe temporel et à sélectionner un modèle qui remplit les conditions suivantes : entre les deux nœuds situés aux extrémités, l'évolution des taux suit un polynôme de degré  $n$  entre deux nœuds adjacents, et les premières et deuxièmes dérivées de cette tendance sont continues à chaque nœud - ce qui garantit le caractère lissé de la fonction. Le problème des splines cubiques réside toutefois dans le choix de la localisation des nœuds, dont dépend la qualité de l'ajustement. Avec les splines pénalisées, Eilers & Marx (1996) ont proposé une solution simple qui consiste à introduire un nombre élevé de nœuds, et à supprimer ceux qui n'apportent pas une information significative. Pour ce faire, ils ajoutent une pénalité sur la différence entre les coefficients de deux nœuds adjacents. Lorsque cette pénalité est très faible ( $\lambda \rightarrow 0$ ), la courbe est très accidentée mais régulière. Elle ressemble au résultat de l'interpolation entre les points, tandis qu'on s'approche d'un modèle polynomial, voire d'une régression linéaire lorsque la pénalité s'élève ( $\lambda \rightarrow \infty$ ). La pénalité joue donc le rôle de paramètre de lissage, et permet de trouver un équilibre entre la fidélité aux données (*goodness of fit*) d'une part, et l'ampleur des fluctuations de la courbe d'autre part (*wiggleness*). Le nombre de degré de liberté reflète ce compromis entre ajustement et complexité, et on peut donc choisir le nombre de degré de liberté plutôt que le paramètre  $\lambda$ . Plus il est élevé, plus la courbe sera peu lissée. Comme la qualité des données est sujette à caution, nous optons pour une solution contraignante à deux degrés de liberté seulement.

Les taux s'obtiennent en prenant, pour chaque groupe d'âge, la somme de l'exponentielle des valeurs prédites, puis en divisant cette somme par la somme des durées exposition correspondantes (pour tenir compte de l'offset). Puisque la structure par âge de la population à risque n'est pas identique, les taux de 15 à 50 ans sont standardisés au départ de la structure de population adulte estimée par les Nations Unies pour l'ensemble de l'Afrique sub-saharienne. Par commodité, nous avons fait le choix de ne pas recourir à des structures de populations nationales. Les Nations Unies fournissent ces données pour des intervalles d'années quinquennaux (1980, 1985, 1990...). Comme la structure par âge sur cette plage d'âges adultes est peu modifiée entre 1980 et 2005, une moyenne est prise pour les vingt-cinq années.

Les calculs sont menés avec le logiciel R, ainsi qu'en Mlwin. Le logiciel R est un logiciel libre d'analyse statistique et de fonctions graphiques largement reconnu pour sa souplesse et sa rapidité de calcul (<http://www.r-project.org/>).