

Recourir aux microsimulations pour étudier la mortalité de crise : illustration par la mortalité au Burundi en 1993

Studying crisis-related mortality using microsimulations : the case of Burundi in 1993

Christophe Bergouignan

Volume 40, numéro 2, automne 2011

Microsimulation en démographie

URI : <https://id.erudit.org/iderudit/1011543ar>

DOI : <https://doi.org/10.7202/1011543ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Association des démographes du Québec

ISSN

0380-1721 (imprimé)

1705-1495 (numérique)

[Découvrir la revue](#)

Citer cette note

Bergouignan, C. (2011). Recourir aux microsimulations pour étudier la mortalité de crise : illustration par la mortalité au Burundi en 1993. *Cahiers québécois de démographie*, 40(2), 299–329. <https://doi.org/10.7202/1011543ar>

Résumé de l'article

Souvent utilisées dans une optique prospective, les microsimulations permettent aussi d'appréhender les biais de sélection affectant certaines enquêtes rétrospectives, notamment lorsque la mesure de la mortalité s'appuie sur les déclarations des enquêtés sur la survie de leurs proches. Susceptibles d'être observés dans divers contextes, ces biais de sélection ont de grandes chances d'être accrus dans les situations d'épidémies ou de crise (catastrophes naturelles, guerres). La microsimulation permet de conduire une analyse de sensibilité de l'ampleur des biais en fonction des facteurs qui en sont à l'origine. Elle permet aussi de tenter de mesurer l'efficacité d'indicateurs imparfaits issus de l'enquête pour estimer la corrélation de la mortalité de crise entre parents biologiques et corriger les résultats de l'enquête. Appliquée aux données de l'enquête réalisée en 2002 au Burundi, cette correction conduit à réévaluer le nombre de personnes tuées par la crise burundaise au cours de l'année 1993.

Recourir aux microsimulations pour étudier la mortalité de crise : illustration par la mortalité au Burundi en 1993

CHRISTOPHE BERGOUIGNAN*

Souvent utilisées dans une optique prospective, les microsimulations permettent aussi d'appréhender les biais de sélection affectant certaines enquêtes rétrospectives, notamment lorsque la mesure de la mortalité s'appuie sur les déclarations des enquêtés sur la survie de leurs proches. Susceptibles d'être observés dans divers contextes, ces biais de sélection ont de grandes chances d'être accrus dans les situations d'épidémies ou de crise (catastrophes naturelles, guerres). La microsimulation permet de conduire une analyse de sensibilité de l'ampleur des biais en fonction des facteurs qui en sont à l'origine. Elle permet aussi de tenter de mesurer l'efficacité d'indicateurs imparfaits issus de l'enquête pour estimer la corrélation de la mortalité de crise entre parents biologiques et corriger les résultats de l'enquête. Appliquée aux données de l'enquête réalisée en 2002 au Burundi, cette correction conduit à réévaluer le nombre de personnes tuées par la crise burundaise au cours de l'année 1993.

English abstract, p. 329

INTRODUCTION

Fréquemment utilisées dans une optique prospective (Pennec, 1994 ; Bélanger et collab. 2005), les microsimulations peuvent aussi l'être pour appréhender les biais affectant certaines sources de données. Si cette dernière forme de recours aux microsimulations couvre un champ relativement large, elle s'avère particulièrement féconde pour redresser les biais de sélection affectant certaines enquêtes rétrospectives. C'est notamment le cas lorsque la mesure de la mortalité construite sur la base des déclarations des

* COMPTRESEC, Université de Bordeaux, France (christophe.bergouignan@u-bordeaux4.fr)

enquêtés sur la survie de leurs proches¹ se heurte à une corrélation des risques de décéder au sein des familles qui entraîne la sous-représentation, parmi les enquêtés, des personnes dont les proches ont été les plus touchés par la mortalité. Autrement dit, l'exploitation des déclarations des enquêtés sur la survie de leurs proches peut se traduire par un biais tendancielle-ment sous-estimateur de la mortalité². Susceptible de prévaloir dans de nombreux contextes anodins avec une intensité et des conséquences relativement modérées, cette corrélation a de grandes chances d'être augmentée lors d'épidémies ou dans les situations de crise (catastrophes naturelles, guerres). Pour certaines épidémies, il est possible, par une connaissance générale, indépendante des données collectées, d'obtenir une estimation du risque de transmission de la maladie entre parents biologiques³, et de l'impact de cette maladie sur le risque individuel de décéder. Cela permet, avec d'autres paramètres, plus contextuels, estimés éventuellement par l'enquête⁴ ou par d'autres sources caractérisant la situation épidémique, de construire une microsimulation reconstituant un niveau de mortalité corrigeant un biais structurellement sous-estimateur. En revanche, dans d'autres situations épidémiques, mais surtout en cas de catastrophes naturelles ou de guerres, il est souvent impossible de connaître *a priori* le niveau de corrélation entre parents biologiques en matière de mortalité associée. De fait, ce niveau dépend de facteurs contextuels propres à la société étudiée ou à la situation de crise elle-même, de sorte que la meilleure

-
1. Personnes ayant appartenu au même ménage que l'enquêté s'agissant de la mesure de la mortalité au cours de l'année précédant l'enquête, personnes ayant un lien biologique (père, mère, enfants, frères et sœurs) avec l'enquêté s'agissant de la mesure de la mortalité au cours des 15-25 années précédant l'enquête.
 2. Il existe aussi des facteurs non aléatoires de sur-estimation. C'est le cas lorsque, recourant aux déclarations de femmes enquêtées de 15 à 49 ans sur la survie de leurs enfants, on mesure la mortalité générale des moins de 5 ans pour une période assez ancienne (plus de 20 ans avant l'enquête) au cours de laquelle les femmes enquêtées avaient toutes moins de 25 ans, les enfants de mères très jeunes ayant très souvent un risque de décéder précocement supérieur à celui des enfants de mères d'âge médian. Ce biais sur-estimateur n'est toutefois pas toujours très net (Bergouignan, 2007) et il est probable qu'il cède souvent le pas aux biais sous-estimateurs liés aux corrélations familiales de mortalité.
 3. Comme c'est le cas pour le VIH/sida pour lequel le risque de transmission verticale entre une femme et ses enfants est documenté selon les situations en matière de thérapie anti-rétrovirale (absence de traitement et combinaisons mono ou multi-thérapeutiques).
 4. Pour le VIH/sida, il s'agira de la prévalence du VIH chez les femmes d'âge fécond, ce qui permet de reconstituer, sous hypothèses, la dynamique de l'épidémie dans cette tranche de population.

source pour en prendre la mesure est parfois l'enquête dont on soupçonne les résultats d'être biaisés par cette corrélation et donc de l'être aussi pour la quantifier. La microsimulation permet alors, sous certaines conditions, de mener une analyse de sensibilité de l'ampleur des biais en fonction des phénomènes susceptibles de les engendrer ou d'en faire varier l'ampleur. C'est l'objet de la première partie de ce texte. Elle permet aussi, dans une optique plus opératoire, de tenter de mesurer la façon dont l'ampleur du biais est associée aux indicateurs imparfaits mesurant, à partir de l'enquête, la corrélation entre parents biologiques en matière de mortalité de crise⁵, ce qui ouvre la voie à une correction des résultats de l'enquête. C'est l'objet de la deuxième partie de ce texte. La troisième et dernière partie de ce texte montre comment l'application de ce type de correction permet de réévaluer le nombre de personnes tuées par la crise burundaise au cours de l'année 1993⁶.

LA MICROSIMULATION OUTIL D'ANALYSE DE LA SENSIBILITÉ DES BIAIS DE MESURE DE LA MORTALITÉ DE CRISE

Dans l'absolu, la meilleure façon d'envisager une analyse de sensibilité est de construire une modélisation mathématique complète combinant les phénomènes intervenant dans la production de la quantité étudiée. Le calcul différentiel rend alors compte de l'influence des différents phénomènes dans les variations de cette quantité et identifie d'éventuelles zones d'évolution critique. Dans certaines situations, la combinaison des phénomènes concourant à engendrer la quantité étudiée revêt une telle complexité que seule une très forte simplification des hypothèses permet de construire ce modèle. Une telle simplification peut alors rendre triviales les conclusions de l'analyse ou, via ses hypothèses, déterminer arbitrairement des interférences entre phénomènes pourtant mal connues. Pour restituer la complexité de la combinaison de ces phénomènes, la microsimulation constitue une alternative intéressante car, ramenant à l'échelle de chaque individu la possibilité de subir chaque phénomène, elle permet de le faire en fonction de son histoire personnelle vis-à-vis des autres

5. Les Burundais préférant très souvent parler de crise plutôt que de guerre pour décrire les événements politiques les ayant touchés, on parlera ici de mortalité de crise et non de mortalité de guerre.

6. Ce texte se réfère à une situation de crise politique. Même si ce n'est pas ici exactement l'objet, il serait toutefois possible d'extrapoler ce principe à d'autres crises comme les catastrophes naturelles, voire certaines épidémies.

phénomènes. Cela inclut évidemment les liens biologiques (ou sociaux) avec d'autres individus de la population simulée, liens que ces phénomènes ont pu engendrer. Cela ouvre aussi la possibilité pour ces individus d'être sélectionnés au cours de la constitution d'un échantillon aléatoirement tiré dans la population simulée. Ces deux derniers aspects rendent la micro-simulation particulièrement adaptée :

- pour recréer les résultats que l'on obtiendrait, en matière de mortalité, par une enquête rétrospective interrogeant les personnes échantillonnées (donc survivantes) sur la survie de leurs proches,
- et pour les comparer avec la mortalité que l'on a effectivement fait subir à l'ensemble de la population.

Trop complexe pour faire l'objet d'une modélisation mathématique générale, cette comparaison par simulation peut être menée dans d'autres cas que l'étude de la mortalité de crise (Masquelier, 2008, 2009 ; Gakidou et King, 2006⁷). La simulation permet toutefois de prendre en compte facilement les fortes variations ponctuelles des phénomènes associées à la situation de crise, variations ponctuelles qui éloignent encore davantage la faisabilité d'une modélisation mathématique. On ne saurait néanmoins considérer qu'une simulation (qu'elle soit macro ou micro) se substitue parfaitement, et en toutes circonstances, à une modélisation mathématique complète disposant d'un grand pouvoir de généralisation. Pour s'approcher de cette capacité de généralisation, les simulations doivent se placer dans une très grande diversité de situations, notamment extrêmes, que ce soit sur le plan du niveau des phénomènes ou sur celui de leurs interactions et corrélations, y compris lorsqu'elles sont assez éloignées du cas étudié. Outre le fait que cela permet de préjuger au minimum des comportements que l'on étudie, cela offre la possibilité de situer le cas étudié dans un panorama plus général et, donc, de mieux comprendre sa spécificité éventuelle.

Le contexte d'observation

L'objectif étant de contrôler puis d'essayer de corriger, sur la base des résultats de microsimulations, les indices de mortalité de crise issus des déclarations des enquêtés sur la survie de leurs proches dans l'enquête

7. Il s'agissait d'évaluer la capacité de différentes options d'exploitation des déclarations des enquêtés sur la survie de leurs frères et sœurs pour estimer la mortalité des années précédant l'enquête.

rétrospective réalisée au Burundi en 2002, soit 9 ans après le début de la crise, le 21 octobre 1993, qui a été marqué dans les jours suivants par d'importants massacres de masse (Nations Unies, 1996) on en fait ici une présentation rapide.

Lors des crises politiques (ici de 1993 à la date de l'enquête), la plupart des systèmes permanents d'observation démographique préexistants (ici l'état civil au taux de couverture déjà très loin de l'exhaustivité en 1993) cessent quasiment de fonctionner. Par ailleurs, les grandes opérations de collecte, comme le recensement, sont alors fréquemment reportées *sine die* (comme ce fut le cas au Burundi). Il existait certes des décomptes du nombre de personnes tuées réalisés par des associations de la société civile. Toutefois, ces derniers ne concernaient que quelques territoires, certes parmi les plus touchés (comme on a pu le vérifier avec les données de l'enquête, Bergouignan, 2004), tout en ne totalisant que 7,5 % (8 942 tués) des résultats de l'enquête 2002 (117 832 tués pour 1993), avant tout redressement corrigeant leur tendance sous-estimatrice. Au total, malgré ses imperfections, les effets de sélection dont ses résultats seraient affectés, et la présence d'environ 530 000 Burundais à l'étranger (principalement dans les camps de réfugiés le long de la frontière, en Tanzanie), la réalisation d'une enquête constituait encore la meilleure option pour prendre la mesure des conséquences de cette crise.

Organisée selon un sondage aréolaire stratifié par type d'espace (sous-collines rurales, camps de déplacés, zones urbaines de dénombrement) et par province⁸, l'enquête a été réalisée de mi-août à fin septembre 2002, dans un contexte de maintien d'une insécurité sporadique qui n'a que modérément perturbé la collecte. Au total, 7 119 ménages ont été enquêtés, soit 32 805 personnes. La mortalité y était estimée par la déclaration des enquêtés sur la survie de leurs enfants⁹ ainsi que sur la survie de leur père et de leur mère¹⁰. En cas de mention d'un décès pour ces questions, une

8. Le recensement de 1990 n'a pas été mobilisé pour constituer la base de sondage de l'enquête, dans la mesure où, en 2002, la répartition spatiale de la population était très différente de ce qu'elle était en 1990, en raison des migrations forcées liées à la crise. Le plan de sondage s'est appuyé sur les décomptes du nombre de ménages annuellement fournis par les responsables administratifs de chaque colline et de chaque zone urbaine de dénombrement, et sur un dénombrement préalable à l'enquête (organisé à dessein) dans les camps de déplacés.

9. Tous les enquêtés de 14 ans et plus (hommes et femmes) répondaient à cette question.

10. Tous les enquêtés répondaient à cette question, donc tous les membres des ménages échantillonnés, quel que soit leur âge, y compris les moins de 14 ans. Pour les très

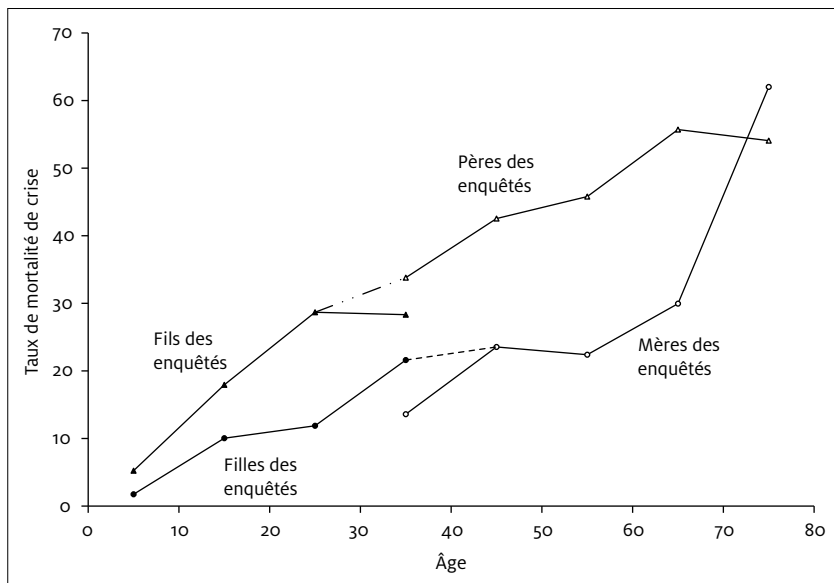
catégorisation sommaire des causes de ces décès était précisée, en distinguant les décès de crise (résultant des blessures infligées lors des violences survenues lors de massacres et des combats). De plus, le nombre total de frères et sœurs survivants et présents au Burundi à la date de l'enquête était demandé aux enquêtés de façon à redresser les indices de mortalité issus des déclarations sur la survie de leur père et de leur mère selon leurs probabilités d'inclusion dans les données de l'enquête.

S'il existe diverses méthodes pour estimer la mortalité à partir des déclarations faites par des enquêtés sur la survie de leurs proches (Brass et Hill, 1973 ; Hill et Trussell, 1977 ; United Nations, 1983 ; Timaeus, 1986, 1992 ; Blacker et Muzika-Gapere, 1988 ; Makinson, 1993), toutes s'appuient sur des régularités statistiques et/ou des relations démographiques fondamentales. Par sa nature exceptionnelle et ponctuelle, la mortalité de crise tend à échapper à ces régularités et se prête donc peu à ces méthodes, qui ne sont pas conçues pour isoler cette cause de l'ensemble de la mortalité. Dans le souci de disposer d'indices de mortalité assez détaillés selon l'âge, et devant ce recours impossible à ces méthodes, en partie indirectes, un

jeunes enquêtés (de moins de 8 ans), les proches présents répondaient à leur place et, s'agissant de très jeunes orphelins, les personnes les ayant recueillis répondaient à leur place. Le choix d'interroger les enquêtés sur la survie de leur père et de leur mère biologiques, plutôt que sur la survie de leurs frères et sœurs, répond au souhait de disposer d'un effectif suffisant de personnes âgées de plus de 60 ans en 1993 (9 ans avant l'enquête). Les analyses de la mortalité de crise au Kosovo, à partir de l'enquête conduite en 1999, avaient alors montré la plus forte exposition des personnes très âgées aux massacres (Bergouignan et Blayo, 2005). Si dater le décès du père et de la mère des enquêtés s'est toujours avéré relativement aisé, grâce à un axe historique permettant d'intercaler les événements vécus par les enquêtés entre les principales dates (nationales et provinciales) jalonnant l'histoire collective, la datation de leur naissance, nécessaire au calcul de la mortalité par âge, s'est avérée plus délicate. Il est clair que nombre de déclarations, concernant l'année de naissance du père et de la mère des enquêtés, sont très approximatives. Pour autant, si l'on excepte une double attractivité pour les années de naissance rondes (1920, 1930, 1940, 1950) et pour les années de naissance correspondant à un âge en « 0 » ou en « 5 » à l'enquête (1932, 1942, 1952, 1957 et 1962), attractivité des années de naissance correspondant à un âge en « 0 » ou en « 5 » à l'enquête aussi observée pour l'année de naissance des enquêtés eux-mêmes, la distribution des années de naissance du père et de la mère des enquêtés ne présente pas d'anomalie particulière, les non réponses restant rares (135 pour la mère, 185 pour le père, pour 32 805 enquêtés). Autrement dit, il y a toutes les raisons de penser que, hormis ces attractivités, les approximations des enquêtés se répartissent assez aléatoirement autour des véritables années de naissance de leur père et de leur mère. Le calcul de taux de mortalité par groupes décennaux d'âge permet ainsi en grande partie d'en limiter les conséquences.

calcul direct de taux de mortalité de crise par sexe et âge décennal a été privilégié¹¹. Les taux obtenus pour l'année 1993, année ayant connu les massacres les plus intenses et pour laquelle les indices de mortalité tirés de l'enquête sont au moins 6 fois supérieurs à ceux mesurés pour les autres années de crise (1994 à 2002 ; Bergouignan, 2004 ; Blayo et collab, 2004 ; Bergouignan et Blayo, 2005), sont présentés figure 1.

FIGURE 1 Taux de mortalité de crise 1993 (pour 1000) directement calculés à partir des données de l'enquête Burundi-2002, selon la nature des déclarations dont ils sont issus



Le calcul direct des taux à partir des données de l'enquête montre une mortalité de crise croissante avec l'âge (les femmes très âgées étant les plus touchées) et une surmortalité masculine, assez importante entre 20 et 70 ans. Ce profil par âge de la mortalité de crise, que l'on retrouve aussi au Kosovo en 1999 (Bergouignan et Blayo, 2005), correspond à une logique de violence systématique, où les groupes armés tuent toutes les personnes appartenant au(x) groupe(s) qu'ils considèrent devoir éliminer. Les personnes les plus exposées sont alors celles ne pouvant pas fuir ou se

11. À l'échelon individuel de chaque enquêté, les données concernant la survie de son père et de sa mère ont été pondérées en fonction de l'inverse du nombre de frères et sœurs qu'il avait déclaré survivant(s) et présent(s) au Burundi.

cache (ou être cachées par des populations non visées). Les personnes les plus âgées sont donc proportionnellement les plus touchées, sans doute parce que les adultes physiquement aptes à fuir rapidement choisissent plutôt d'emmener avec eux leurs enfants que leurs parents plus âgés. La meilleure survie des enfants par rapport aux jeunes adultes s'explique par une identification sociale plus faible, ce qui leur a permis de se cacher dans les espaces naturels environnants et/ou d'être recueillis par d'autres familles¹².

Par ailleurs, on observe des écarts entre les taux de mortalité de crise à 30-39 ans selon qu'ils sont issus des déclarations des enquêtés sur la survie de leurs enfants ou de leur père et de leur mère. Cette confrontation est rendue possible par des Burundais ayant très largement eu au moins un enfant à 30 ans, et par l'absence de limite d'âge pour recruter les enquêtés, qui permet à une part importante des enfants qu'ils déclarent d'avoir déjà pu atteindre entre 30 et 40 ans en 1993. Cette discordance modérée entre les taux, mais pas négligeable, révèle un biais dont il est difficile d'être certain qu'il résulte uniquement de corrélations familiales de la mortalité, mais qui est tendanciellement sous-estimateur. Aussi, le choix de la valeur la plus élevée parmi les deux taux de mortalité de crise semble raisonnable, d'autant qu'il donne aux deux séries un profil cohérent, néanmoins sans certitude qu'il s'agisse d'une correction suffisante.

Les paramètres de microsimulation

La première étape de la microsimulation, et la plus importante, est l'engendrement de la population dans laquelle l'échantillon reproduisant les mesures de mortalité de crise que l'on obtiendrait avec une enquête comparable à Burundi-2002 puisse être tiré. Cette enquête se référant à des caractéristiques du père, de la mère et des enfants de l'enquêté, et interrogeant des personnes sans limite supérieure d'âge, il est nécessaire, pour la microsimulation, d'engendrer des personnes nées jusqu'à 40-45 ans avant la personne enquêtée la plus âgée (103 ans). Autrement dit, cela exige un

12. La population burundaise étant très jeune, ce profil par âge des taux de mortalité de crise est « compensé » par la structure par âge de la population, et la répartition par âge des décès masque ainsi fortement l'inégale intensité de la mortalité de crise selon les âges. Si l'on considère la mortalité, non plus de crise (résultant directement des violences) mais liée à la détérioration des conditions de vie, les groupes physiquement les plus fragiles (moins de 5 ans, et, à un moindre degré, 60 ans et plus) sont ceux dont les risques de décéder par une autre cause que la crise ont le plus augmenté avec la crise (Bergouignan, 2007).

recul d'environ 150 ans avant l'enquête. La dynamique de la population microsimulée de 1852 à 1992 est celle d'une population stable définie par un niveau de fécondité et de mortalité¹³. En 1993, la mortalité de crise vient s'ajouter à la mortalité générale associée à cette population stable. De 1994 à 2002, la population microsimulée est soumise à une mortalité de crise beaucoup plus modeste et moins susceptible d'être familialement corrélée. Quoi qu'il en soit, les résultats ici présentés se concentrent sur l'année 1993 pour laquelle la mortalité de crise a été la plus forte mais aussi la plus sensible aux biais résultants de corrélations familiales de la mortalité liées à des massacres à caractère ethnique très marqué.

Les paramètres de microsimulation sont donc :

- avant 1993, la fécondité (que l'on fait varier de 1,4 à 7,1 enfants par femme — 4 niveaux testés) et la mortalité par une autre cause que la crise (que l'on fait varier d'une série de quotients correspondant à une espérance de vie à la naissance de 50 ans à une série de quotients correspondant à une espérance de vie à la naissance de 80 ans¹⁴ — uniquement 2 niveaux testés),
- à partir de 1993, on ajoute la mortalité de crise (dont on fait varier le taux standardisé¹⁵ de 20 pour 1000 à 480 pour 1000 — 4 niveaux testés) et la corrélation de cette mortalité de crise entre les membres d'une même famille (R variant de 0,01 à 0,85¹⁶ — 13 niveaux testés).

13. Bien que cela ne change rien aux résultats, la structure de la population de départ (150 ans avant l'enquête) correspond, par souci de cohérence, à la population stable définie par cette fécondité et cette mortalité.

14. Ces séries de quotients ont été construites à partir des relations correspondant aux tables types de Ledermann (1969).

15. Cette standardisation utilise la structure par âge moyenne des 416 populations microsimulées, donnant un poids dominant aux moins de 30 ans. Les fluctuations inhérentes à la microsimulation font, pour les différentes populations microsimulées, légèrement dévier de leur valeur ciblée l'ensemble des paramètres, notamment celui-ci.

16. Dans l'absolu, cette corrélation de la mortalité de crise entre membres d'une même famille varie de 0 à 1. En fait, la seule inégale distribution géographique des phénomènes associés aux crises rend inutile l'intégration d'une corrélation nulle dans les microsimulations. Inversement, l'introduction d'une corrélation totale (égale à 1), au delà de son résultat évident (opposition exclusive entre des familles entièrement tuées et des familles entièrement épargnées par la mortalité de crise), dépasse les situations les plus critiques jamais rencontrées. L'exemple du Rwanda est particulièrement évocateur, puisque, malgré l'extrême violence des massacres, leur durée beaucoup plus importante et leur caractère ouvertement génocidaire, il subsiste dans l'Enquête Démographique et de Santé de 2000 de nombreuses personnes pour déclarer le décès de leurs proches en 1994, ce qui permet d'y mesurer une surmorta-

Après essais, la diversité des niveaux testés pour chaque paramètre a été établie selon son influence sur l'ampleur des biais. Introduite en raison des interférences entre causes de mortalité, la mortalité par une autre cause que la crise joue un rôle négligeable sur ces biais, d'où la réduction de son champ de variation à deux niveaux encadrant la très grande majorité des situations du monde d'aujourd'hui. À contrario, la corrélation de mortalité de crise entre les membres d'une même famille joue logiquement un rôle primordial, d'où les 13 niveaux testés. Cette corrélation est programmée de façon à ce que dès qu'une personne simulée est tuée par la crise, le risque d'être tué par la crise des autres membres de sa famille biologique augmente (en fonction du niveau de corrélation), une augmentation de ce même facteur s'observant à chaque fois qu'un nouveau membre de ce réseau familial est tué par la crise. La fécondité, on le verra, joue aussi un rôle important, d'où son large éventail de valeurs (de 1,4 à 7,1 enfants par femme) qui, parfois, diffèrent très fortement du contexte burundais, que l'on peut grossièrement associer, à l'orée de la crise de 1993, aux populations microsimulées avec 7,1 enfants par femme et 50 ans d'espérance de vie à la naissance. Le fait d'introduire d'autres contextes macro-démographiques, y compris fort éloignés, permet de mettre en évidence leur rôle sur l'ampleur des biais.

On engendre ainsi 416 populations différentes, dont les effectifs survivants, donc susceptibles d'être enquêtés, sont très variables, du fait de la dynamique de population stable associée aux paramètres de mortalité et de fécondité. Pour limiter cette disparité d'effectifs de personnes survivantes pouvant virtuellement être échantillonnées à la date de l'enquête, et de façon à disposer pour toutes les populations d'échantillons de taille comparable (autour de 15 000 personnes¹⁷) sans avoir des taux de sondage variant d'un facteur supérieur à 1000 entre les populations, les effectifs initiaux (150 ans avant l'enquête) ont été modulés selon les populations.

lité spectaculaire. On peut, sans le moindre redressement, estimer à partir des taux tirés de l'enquête 2000, à 760 000 (Bergouignan, 2007), le nombre de Rwandais tués en 1994, soit presque autant que les estimations « officielles » (800 000 personnes tuées) basées sur d'autres sources qui, sans être parfaites (elles sous-évaluent probablement le nombre de décès de crise), fournissent un repère. Autrement dit, y compris dans les situations de génocide visant l'assassinat de familles entières, on observe une forte corrélation familiale de la mortalité de crise sans qu'elle soit égale à 1. Dans le cas burundais, son estimation approximative indirecte (voir infra) conduit à un coefficient de corrélation de l'ordre de 0,5.

17. Ordre de grandeur des effectifs de la plupart des enquêtes démographiques réalisées dans le monde en développement.

On dispose ainsi, en 2002, de 416 populations composées de vivants et de morts, dont les personnes initialement présentes à la base de la microsimulation (en 1852, soit 150 ans avant l'enquête) sont exclues car ne pouvant pas être reliées à leurs pères et mères. Au sein de chacune de ces 416 populations de vivants et de morts, chaque personne est caractérisée par : son secteur de vie¹⁸, son sexe, son année de naissance, son année de décès éventuel, son éventuelle cause de décès (crise ou autre cause). Elle peut être appariée par un identifiant à un fichier où sont renseignés : l'année de naissance de son père et de sa mère, leur année de décès éventuel et leur éventuelle cause de décès (crise ou autre cause) et le nombre de ses frères et sœurs survivants à la date de l'enquête¹⁹. Par ce même identifiant, elle peut être appariée à la liste de ses éventuels enfants avec leur année de naissance, leur année de décès éventuel, leur éventuelle cause de décès (crise ou autre cause).

La deuxième étape de la microsimulation est la reproduction de conditions de tirage de l'échantillon parmi les survivants de chacune des 416 populations à la date de l'enquête. Il s'agit notamment ici de reproduire le caractère aréolaire du sondage dans la mesure où, dans un contexte de mortalité de crise très variable géographiquement, les effets de grappe ont une forte incidence sur la précision des résultats obtenus. Les autres aspects de l'échantillonnage n'ont guère d'importance dans la problématique ici étudiée. Au final, on dispose donc d'un échantillon d'environ 15 000 personnes pour chacune des 416 populations microsimulées.

La troisième et dernière étape de la microsimulation est le calcul de la mortalité de crise de l'année 1993. D'une part, elle est réalisée à partir des données concernant le père, la mère et les enfants des enquêtés, de façon à reconstituer les résultats que l'on obtiendrait via une enquête rétrospective post-crise interrogeant les enquêtés sur la survie de leurs proches (ici selon les mêmes questions que l'enquête Burundi-2002). D'autre part, elle est

18. La mobilité résidentielle est ici considérée comme nulle, ce qui est très éloigné de la réalité (exode rural et mobilités forcées intérieures et extérieures au Burundi). Toutefois, elle ne relève pas, à proprement parler, de la question des biais de mesure de la mortalité de crise. Si les corrélations entre mobilité résidentielle des déplacés et mortalité de crise créent des concentrations de personnes plus exposées à la mortalité de crise dans les camps de déplacés intérieurs au Burundi, cette question est ici gérée via la simulation du caractère aréolaire de l'échantillonnage.

19. Comme dans l'exploitation de l'enquête Burundi-2002, les taux de mortalité de crise issus des déclarations des enquêtés microsimulés sur la survie de leur père et de leur mère doivent tenir compte de la taille des fratries survivantes à l'enquête, pour corriger l'effet des différences de probabilité d'inclusion dans l'enquête des personnes pouvant parler d'un même père et/ou d'une même mère.

réalisée à partir des données concernant toutes les personnes ayant vécu dans la population considérée (apparentées ou non à une personne survivante à l'enquête), soit la véritable mortalité de crise de l'ensemble de la population simulée. Dans les deux cas, les indices de mortalité de crise obtenus sont des taux par groupe décennal d'âges et par sexe²⁰. La comparaison des indices obtenus via ces deux démarches de mesure permet de rendre compte, aux fluctuations aléatoires près (échantillonnage, et, à un moindre degré, microsimulation), de l'ampleur des biais résultant des corrélations familiales de la mortalité. Les biais sont ensuite synthétisés pour l'ensemble des moins de 40 ans et pour l'ensemble des 40-79 ans²¹, afin de résumer les analyses correspondant à une même catégorie de données exploitées²².

L'ampleur des biais résultant des corrélations familiales de la mortalité

Avant de mener cette comparaison synthétique, une comparaison détaillée (par groupe décennal d'âges, sexes réunis, les résultats étant très proches pour les hommes et les femmes) permet d'identifier les principaux facteurs influençant l'ampleur des biais. Les figures 2a et 2b illustrent, pour quelques populations microsimulées, la nature des biais affectant l'exploitation des déclarations des enquêtés sur la survie de leur parenté biologique.

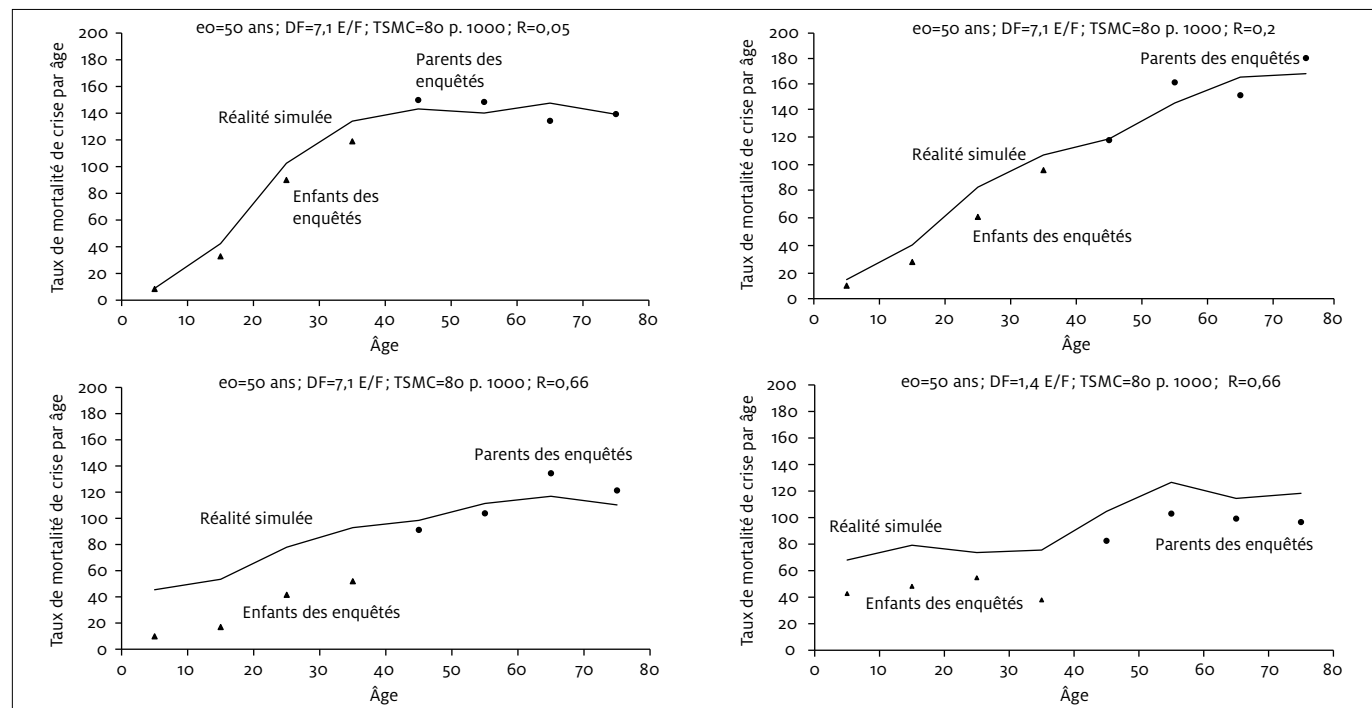
Lorsque la corrélation de la mortalité de crise entre les membres d'une même famille biologique est faible ($R < 0,1$), les indices tirés de l'exploitation des déclarations des enquêtés sur la survie de leurs enfants comme sur

20. Par souci de comparabilité, les taux de mortalité de crise issus des déclarations des enquêtés microsimulés sur la survie de leur père et de leur mère ne couvrent que 4 groupes décennaux d'âges (de 40-49 ans à 70-79 ans). Si pour deux régimes microsimulés de fécondité (7,1 et 4,4 enfants par femme) il est possible de construire un taux de mortalité de crise des 30-39 ans du fait d'une entrée dans la maternité et la paternité précédant ces âges pour l'immense majorité des personnes, ce n'est pas le cas des deux autres régimes microsimulés de fécondité (1,4 et 2,1 enfants par femme). Les taux de mortalité de crise issus des déclarations des enquêtés microsimulés sur la survie de leurs enfants couvrent 4 groupes décennaux d'âges (de 0-9 ans à 30-39 ans).

21. Pour neutraliser les différences de composition par âge internes à ces deux grands groupes d'âges, les indices synthétiques d'ampleur des biais résultent d'une standardisation directe par sexe et âge, la structure type étant celle des survivants de la moyenne des populations microsimulées au 1^{er} janvier 1993.

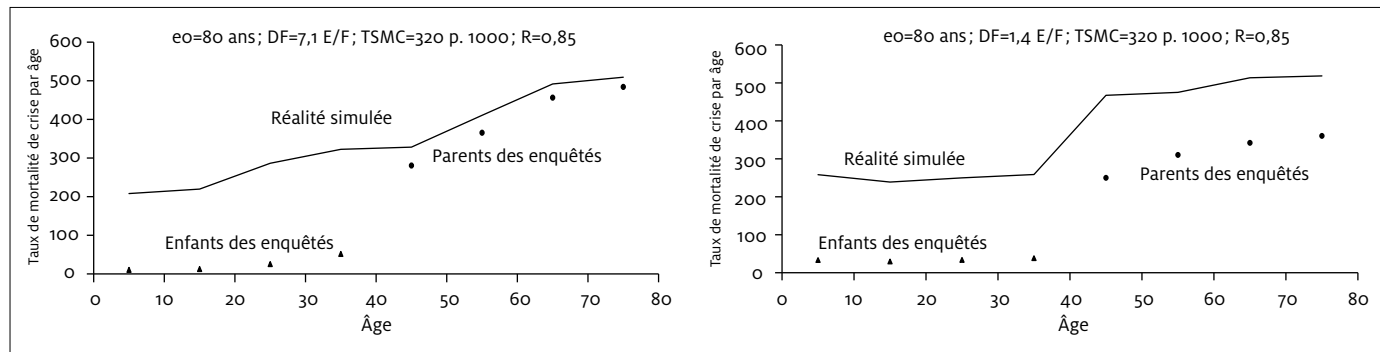
22. Déclarations des enquêtés sur la survie de leurs enfants versus déclarations des enquêtés sur la survie de leur père et de leur mère.

FIGURE 2A Taux de mortalité de crise par groupe décennal d'âges selon son origine (réalité microsimulée, déclaration des enquêtés microsimulés sur la survie de leurs proches) et selon les caractéristiques de la population microsimulée



Abréviations : DF, descendance finale ; TSMC, taux standardisé de mortalité de crise (structure type moyenne des populations microsimulées) ; eo, espérance de vie à la naissance en l'absence de crise ; R, coefficient de corrélation de la mortalité de crise entre membres d'une famille.

FIGURE 2B Taux de mortalité de crise par groupe décennal d'âge selon son origine (réalité microsimulée, déclaration des enquêtés microsimulés sur la survie de leurs proches) et selon les caractéristiques de la population microsimulée



Abréviations : DF, descendance finale; TSMC, taux standardisé de mortalité de crise (structure type moyenne des populations microsimulées); eo, espérance de vie à la naissance en l'absence de crise; R, coefficient de corrélation de la mortalité de crise entre membres d'une famille.

celles de leurs parents semblent conduire à des taux de mortalité de crise très proches de ceux correspondant à la réalité microsimulée. Ainsi, avec une faible corrélation de la mortalité de crise entre les membres d'une même famille biologique on ne constate pas de biais systématique lié au recours à ces déclarations des enquêtés, mais seulement des écarts reflétant les fluctuations d'échantillonnage. Lorsque cette corrélation de la mortalité de crise entre les membres des familles biologiques devient plus forte ($R > 0,4$) :

- les « véritables » taux de mortalité de crise des moins de 20 ans sont, en valeur relative, plus proches de ceux des âges ultérieurs, cette corrélation reflétant une volonté de tuer des familles entières indépendamment de l'âge des membres,
- on observe un biais systématiquement sous-estimateur pour les taux de mortalité de crise des moins de 40 ans construits à partir de l'exploitation des déclarations des enquêtés sur la survie de leurs enfants.

En revanche, lorsque la fécondité est élevée (7,1 enfants par femme), ce biais sous-estimateur n'est pas observé pour les taux de mortalité de crise des 40-79 ans issus de l'exploitation des déclarations des enquêtés sur la survie de leurs parents. On constate ainsi, pour une même corrélation familiale de la mortalité de crise (ici $R = 0,66$), que le biais sous-estimateur affectant ces taux s'accroît avec une fécondité plus faible. Modeste pour une fécondité de 4,4 enfants par femme (environ 8 % de la valeur des « véritables » taux), il atteint un niveau nettement plus important pour une fécondité de 1,4 enfant par femme (près du quart de la valeur des « véritables » taux).

De fait, même avec une forte corrélation familiale de mortalité de crise, une forte descendance réduit la probabilité qu'aucun enfant des familles les plus touchées ne survive à la crise pour rapporter le décès de ses parents. Avec une faible descendance, cette probabilité devient beaucoup plus élevée et les biais sous-estimateurs se manifestent pleinement. Dans l'absolu, s'agissant de personnes survivantes, la mortalité par une autre cause que la crise interfère avec la fécondité, son rôle étant toutefois compensé par les interférences avec la mortalité de crise.

En revanche, avec une forte corrélation familiale de la mortalité de crise, une forte fécondité ne réduit pas la forte probabilité que dans les familles les plus touchées, père et mère soient tués par la crise et donc ne puissent déclarer la mortalité de crise de leurs enfants dans une enquête rétrospective. Il est donc logique, y compris avec une forte fécondité, que

les taux de mortalité de crise des moins de 40 ans issus des déclarations des enquêtés sur la mortalité de leurs enfants soient affectés d'un biais sous-estimateur dès qu'il existe une forte corrélation familiale de la mortalité de crise. En fait, mais c'est peu perceptible dans les figures 2a et 2b, lorsqu'il existe une corrélation familiale de la mortalité de crise, une forte fécondité accroît légèrement les sous-estimations de la mortalité de crise des moins de 40 ans issue des déclarations des enquêtés sur la survie de leurs enfants. Cet effet mineur, mais non négligeable, s'explique par un effet d'échantillonnage, avec un moindre éventail de possibilités en matière de nombre d'enfants tués par la crise pour les petites descendances. Ainsi, en cas de corrélation familiale substantielle de la mortalité de crise, dans les familles fortement touchées, dont les deux parents ont été tués, les aléas de la mortalité de crise des enfants vont engendrer :

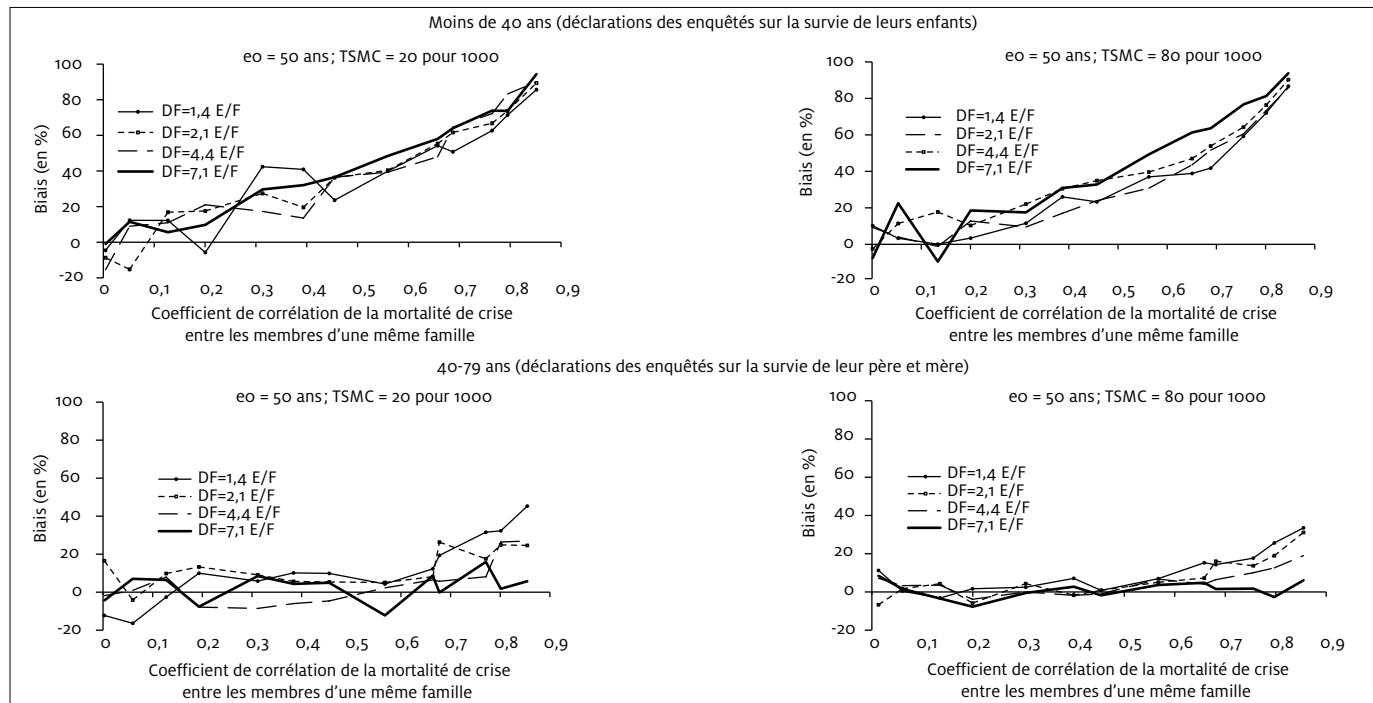
- pour les petites descendances (notamment 1 enfant, mais aussi 2 voire 3), une plus grande polarisation du nombre d'enfants tués rendant plus fréquentes les situations où aucun enfant n'a été tué,
- pour les descendances plus vastes, une distribution plus homogène du nombre d'enfants tués, impliquant une grande rareté des situations où aucun enfant n'a été tué.

Les différences de mortalité de crise entre les fratries d'enfants dont le père et la mère ont été tués et ne seront donc pas enquêtés et celles dont le père ou la mère ont survécu et pourront être enquêtés sont alors plus faibles en cas de faible fécondité. Autrement dit, via un plus grand aléa de la mortalité de crise affectant les grappes d'enfants, une faible fécondité traduit un niveau théorique de corrélation de la mortalité de crise entre parents et enfants en un niveau empirique plus modéré de corrélation. Naturellement, lorsque ce niveau théorique de corrélation familiale de la mortalité devient très élevé ($R > 0,75$), cet effet atténuateur des faibles fécondités se réduit puisque, dans ce cas, lorsque le père et la mère ont été tués, les fratries d'enfants indemnes deviennent très rares même si elles sont petites.

Cette inversion asymétrique de l'effet de la fécondité sur les biais affectant les taux de mortalité de crise selon la forme de déclaration des enquêtés sur la survie de leurs proches est souvent renforcée dans les situations de guerre civile par la plus grande facilité des moins de 15 ans à échapper aux massacres (voir supra).

La figure 3 illustre clairement ces mécanismes. Elle montre aussi le faible rôle du niveau de mortalité de crise dès lors que l'écart entre les « véritables » taux de mortalité de crise et ceux issus de l'enquête s'exprime

FIGURE 3 Écarts relatifs synthétiques (ou biais, en %) entre les «véritables» taux de mortalité de crise simulés et ceux issus de l'enquête simulée, selon le grand groupe d'âge (donc le mode d'exploitation de l'enquête)



Abréviations: DF, descendance finale; TSMC, taux standardisé de mortalité de crise (structure type moyenne des populations microsimulées); eo, espérance de vie à la naissance en l'absence de crise.

en proportion des « véritables » taux de mortalité de crise. Ainsi, ce rôle du niveau de mortalité de crise sur l'ampleur relative des biais se limite essentiellement à la réduction des fluctuations aléatoires affectant les écarts relatifs entre les « véritables » taux de mortalité de crise et ceux issus de l'enquête. Assez logiquement, ces fluctuations sont plus modérées lorsque la mortalité de crise est élevée puisque le nombre de décès de crise devient plus important pour une même taille d'échantillon. Pour environ 15 000 personnes enquêtées, les fluctuations affectant les écarts relatifs s'atténuent avec une mortalité de crise dont le taux standardisé atteint 80 pour 1000.

Au total, mesurés en proportion de la valeur devant être estimée, les biais affectant les taux de mortalité de crise des moins de 40 ans issus des déclarations des enquêtés sur la survie de leurs enfants dépendent :

- du niveau de corrélation de la mortalité de crise entre les membres d'une même famille biologique ; en deçà d'un coefficient de corrélation de 0,1 il n'y a pas de biais sous-estimateur mais plutôt des écarts aléatoires ; au-delà d'un coefficient de corrélation de 0,4 les sous-estimations dépassent très fréquemment 20 % de la « véritable » valeur du taux de mortalité de crise ; globalement le niveau de corrélation familiale de la mortalité de crise explique environ 77 % de la valeur de l'écart relatif²³ ;
- du niveau de la fécondité ; les biais sous-estimateurs sont accrus dans les populations à forte fécondité, notamment pour une corrélation familiale substantielle mais limitée ($0,4 < R < 0,75$) de la mortalité de crise ; globalement le niveau de la fécondité explique environ 16 % de la valeur de l'écart relatif.

Mesurés en proportion de la valeur devant être estimée, les biais affectant les taux de mortalité de crise des 40-79 ans issus des déclarations des enquêtés sur la survie de leurs parents dépendent :

- du niveau de corrélation familiale de la mortalité de crise ; en deçà d'un coefficient de corrélation de 0,4 il n'y a pas de biais sous-estimateur mais des écarts aléatoires ; au-delà de ce niveau de corrélation, les sous-estimations se manifestent lorsque la fécondité est

23. Ces pourcentages d'explication sont tirés de régressions linéaires emboîtées ajustées sur les résultats construits pour les 416 populations, ces régressions tentent d'expliquer les écarts relatifs (entre les « véritables » taux de mortalité de crise et ceux issus des déclarations des enquêtés sur la survie de leurs proches — père, mère et enfants), en fonction du niveau de corrélation familiale de la mortalité de crise, du niveau de la mortalité de crise, du niveau de la mortalité par une autre cause que la crise, du niveau de la fécondité et enfin des interactions multiplicatives entre ces facteurs.

faible, ils ne dépassent alors 20 % de la « véritable » valeur du taux de mortalité de crise que pour des corrélations assez fortes ($R > 0,7$) ; globalement le niveau de corrélation familiale de la mortalité de crise explique environ 47 % de la valeur de l'écart relatif ;

- du niveau de la fécondité ; les biais sous-estimateurs ne se manifestant pas avant une très forte corrélation dans les populations à forte fécondité (> 6 enfants par femme) ; globalement le niveau de la fécondité explique environ 9 % de la valeur de l'écart relatif ; en fait l'interaction entre corrélation familiale de la mortalité de crise et fécondité explique davantage (17 %) la valeur de l'écart relatif.

DE L'IDENTIFICATION DES FACTEURS DE BIAIS À UN POSSIBLE OUTIL DE CORRECTION EMPIRIQUE

L'analyse, pour les 416 populations microsimulées, des écarts entre les « véritables » taux de mortalité de crise et ceux construits à partir des déclarations des enquêtés sur la survie de leurs parents et enfants permet d'identifier les situations susceptibles d'engendrer d'importants biais de sélection. Qu'il s'agisse de déclarations des enquêtés sur la survie de leurs parents ou sur la survie de leurs enfants, ces situations de biais sont essentiellement définies à partir du niveau de fécondité et du niveau de corrélation familiale de la mortalité de crise. Ces deux facteurs à eux seuls et leurs interactions expliquent 93 % de l'ampleur des écarts relatifs synthétiques entre les « véritables » taux de mortalité de crise et ceux construits à partir des déclarations des enquêtés sur la survie de leurs enfants et 73 % de l'ampleur des écarts relatifs synthétiques entre les « véritables » taux de mortalité de crise et ceux construits à partir des déclarations des enquêtés sur la survie de leurs parents. Aussi, la maîtrise de ces facteurs, dans une situation de collecte donnée obéissant au même protocole (questionnaire, échantillonnage) que l'enquête Burundi-2002, peut permettre d'estimer l'ampleur de l'écart entre les véritables taux de mortalité de crise et ceux tirés de l'enquête. Autrement dit, le contrôle de ces facteurs peut conduire à une correction des taux de mortalité de crise tirés de l'exploitation des déclarations des enquêtés sur la survie de leurs parents et enfants, et donc à une réduction du biais éventuel affectant la mortalité de crise.

Si l'appréhension du niveau de fécondité à partir de l'enquête ne pose pas de problème majeur, il n'en va pas de même pour la mesure de la corrélation de la mortalité de crise entre les membres d'une même famille. En effet, la qualité de la mesure empirique de cette corrélation à partir des

données de l'enquête dépend de son niveau : plus elle est importante, moins il sera aisé de l'évaluer. Ce serait même impossible avec une corrélation égale à 1, ce qui ne correspond néanmoins à aucune réalité connue (voir note 16). Autrement dit, sans prétendre produire, à partir de l'enquête, une mesure parfaite et unique de la corrélation familiale de la mortalité de crise, il est possible, par au moins deux indicateurs, de tenter d'approcher le niveau de cette corrélation sur la base des déclarations des enquêtés. Ils peuvent être construits à partir des informations relatives aux enquêtés issus des 416 populations microsimulées comme à partir des données de n'importe quelle enquête conçue comme l'enquête Burundi-2002.

On propose ici deux indicateurs théoriquement imparfaits, mais empiriquement intéressants, pour appréhender la corrélation familiale de la mortalité de crise.

Le premier d'entre eux estime la corrélation de la mortalité de crise au sein d'une fratrie d'enfants déclarés par un même enquêté. Il s'agit de mesurer, pour chaque taille de fratrie exprimée en nombre de survivants juste avant la crise (à la condition qu'elles en comptent au moins 2), l'écart entre :

- la distribution observée de ces fratries selon leur nombre de membres décédés de la crise et
- la distribution théorique de ces fratries selon leur nombre de membres décédés de la crise. Cette distribution théorique correspond à une distribution binomiale impliquant, par définition, l'indépendance des risques de décéder de la crise entre les membres de la fratrie. La probabilité d'être tué par la crise y est considérée identique pour l'ensemble des membres des fratries de cette taille, quel que soit le nombre d'autres membres décédés de la crise au sein d'une fratrie.

En pratique, il s'agit de l'écart absolu moyen entre chaque composante des distributions observée et théorique correspondant à chaque taille de fratrie en nombre de survivants juste avant la crise.

Le second d'entre eux estime la corrélation de la mortalité de crise entre le père et la mère déclarés par chaque enquêté. Il s'agit de mesurer, pour les couples de parents des enquêtés dont les deux membres étaient encore vivants juste avant la crise, l'écart entre :

- la distribution observée de ces couples selon leur nombre de membres décédés de la crise (0, 1 ou 2) et

- la distribution théorique de ces couples selon leur nombre de membres décédés de la crise. Cette distribution théorique correspond à une distribution binomiale impliquant, par définition, l'indépendance des risques de décéder de la crise entre les membres du couple. La probabilité d'être tué par la crise y est considérée identique pour les deux membres des couples, que l'autre membre du couple soit décédé de la crise ou non.

En pratique, il s'agit de l'écart absolu moyen entre chaque composante des distributions observée et théorique des couples de parents des enquêtés dont les deux membres étaient encore vivants juste avant la crise.

Le rôle de la fécondité dans l'ampleur du biais de mesure de la mortalité de crise transitant par la taille des fratries issues de personnes n'ayant pas obligatoirement achevé leur vie féconde, on peut recourir à la descendance atteinte par les enquêtées de 15 ans et plus pour rendre compte de l'impact de la fécondité. Pour éliminer les effets sur cet indicateur de la structure par âge interne au groupe des femmes enquêtées de 15 ans et plus, cette descendance atteinte doit être standardisée (avec la structure par âge moyenne des femmes de 15 ans et plus des 416 populations microsimulées).

Il est alors possible de construire (figure 4²⁴), à partir de l'ensemble statistique formé des 416 populations microsimulées, un ajustement linéaire multiple de l'ampleur synthétique relative des biais affectant la mortalité de crise (des moins de 40 ans d'une part et des 40-79 ans d'autre part), avec :

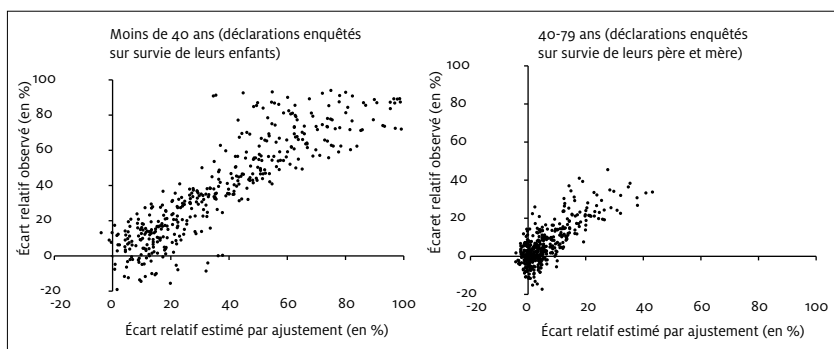
- les deux indicateurs imparfaits mesurant la corrélation familiale de la mortalité de crise,
- la descendance atteinte standardisée,
- deux termes exprimant les interactions multiplicatives entre les indicateurs de corrélation et la descendance atteinte.

Les coefficients de détermination de ces ajustements linéaires multiples sont assez élevés ($R^2 = 77,3$ % pour les biais affectant la mortalité de crise des moins de 40 ans issue des déclarations des enquêtés sur la survie de leurs enfants ; $R^2 = 57,6$ % pour les biais affectant la mortalité de crise des 40-79 ans issue des déclarations des enquêtés sur la survie de leurs parents). Dans les deux cas, ils correspondent à une perte d'explication

24. Par souci de lisibilité on présente ici les écarts synthétiques pour les moins de 40 ans et les 40-79 ans, dans la pratique on construit un ajustement linéaire multiple pour chaque groupe décennal d'âges. Si le niveau de détermination de ces ajustements varie assez peu autour des valeurs synthétiques ici obtenues, les coefficients associés à chaque indicateur et leurs interactions sont assez différents. Ils conduisent, pour chaque grand groupe d'âges (moins de 40 ans et 40-79 ans) à des corrections diminuant avec l'âge.

d'environ 16 points de pourcentage (77 % contre 93 % ; 57 % contre 73 %) par rapport aux ajustements réalisés entre l'ampleur synthétique relative des biais et les paramètres de base de la microsimulation en matière de fécondité et de corrélation familiale de la mortalité de crise. Cette perte de détermination est toutefois loin d'éliminer toute relation, et l'on peut essayer de l'utiliser pour approcher l'ampleur de biais qui, lorsque l'on exploite une enquête rétrospective, ne peuvent être anticipés. Il ne s'agit pas pour autant de prétendre proposer une série d'abaques correctifs basés sur les coefficients de l'ajustement, alors considérés comme parfaitement transposables. Ainsi, sans modifier la qualité globale des ajustements (mesurée par le R^2), les changements dans les paramètres de microsimulation influent sur la valeur des coefficients du modèle. C'est notamment le cas des coefficients de la descendance atteinte et des termes d'interaction qui sont assez sensibles à la variabilité des niveaux de fécondité entre populations microsimulées.

FIGURE 4 Représentation graphique des 416 populations microsimulées selon l'écart synthétique relatif (en %)



Note: écart observé entre les « véritables » taux mortalité de crise et ceux issus des déclarations des enquêtés sur la survie de leurs enfants (taux des moins de 40 ans) et de leurs parents (taux des 40-79 ans), et estimé par ajustement.

ESSAI DE CORRECTION DE LA MORTALITÉ DE CRISE AU BURUNDI EN 1993

La corrélation familiale de la mortalité de crise au Burundi en 1993

À partir de l'enquête Burundi-2002, il est possible de comparer distributions « théoriques » et observées selon le nombre de membres tués par la crise en 1993 :

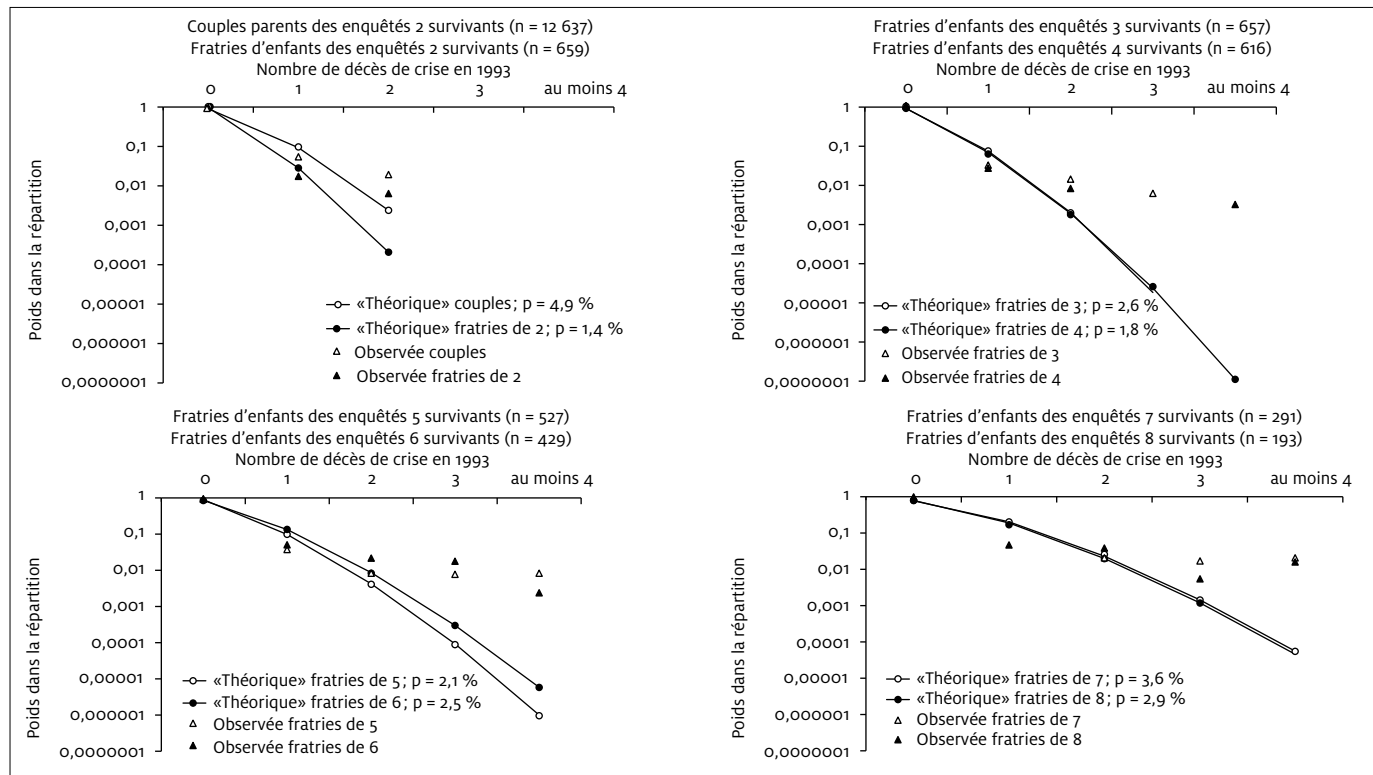
- des fratries d'enfants des enquêtés de chaque taille (en nombre de survivants début 1993, si celui-ci est supérieur à 1),
- des couples de père et mère des enquêtés comptant deux survivants début 1993.

Les distributions « théoriques »²⁵ sont construites en appliquant une loi binomiale à chaque taille de fratrie de survivants en 1993 et aux couples de deux parents survivants en 1993, la probabilité individuelle d'être tué par la crise étant estimée distinctement pour chaque taille de fratrie de survivants en 1993 et pour les couples de deux parents survivants en 1993. Les résultats de cette comparaison sont présentés figure 5. Elle se caractérise, pour les couples et pour les fratries comptant au moins deux survivants début 1993, par une fréquence de grappes ne comptant aucun tué par la crise ou en comptant au moins deux, plus importante dans les distributions observées que dans les distributions « théoriques ». *A contrario*, les grappes comptant un seul tué sont plus fréquentes dans les distributions « théoriques ». Autrement dit, les données de l'enquête Burundi-2002 montrent une corrélation de la mortalité de crise au sein des fratries d'enfants et au sein des couples de parents déclarés par les enquêtés. Elle se traduit par une opposition entre des fratries et des couples parfaitement indemnes et des fratries et des couples fortement touchés concentrant les décès de crise. En l'absence de toute corrélation familiale de la mortalité de crise, 87 % des tués seraient le seul membre de leur fratrie à l'être²⁶, alors

25. Le terme de distribution « théorique » est ici à relativiser, d'où les guillemets, dans la mesure où celle-ci se base non sur la véritable estimation de la probabilité individuelle d'être tué par la crise en 1993 (inconnue), pour les survivants début 1993 des fratries d'une taille donnée, mais seulement sur l'estimation de cette probabilité pour les survivants début 1993 des fratries d'une taille donnée dont au moins un parent a survécu jusqu'à la date de l'enquête pour en rapporter la survie. Les distributions observées sont aussi celles prévalant pour les fratries d'une taille donnée dont au moins un parent a survécu jusqu'à la date de l'enquête pour en rapporter la survie. De même, les distributions « théoriques » et observées de couples de parents survivants début 1993 correspondent aussi à celles des couples de parents dont au moins un enfant a survécu jusqu'à l'enquête. Ces restrictions limitent la signification concrète de ces probabilités et distributions, mais n'empêchent pas leur usage correctif, les résultats microsimulés utilisés pour estimer les paramètres du modèle de correction répondant aux mêmes restrictions.

26. Cette proportion théorique a été calculée en tenant compte de la structure des fratries d'enfants déclarés par les enquêtés selon leur nombre de survivants début 1993 ainsi que du léger différentiel des probabilités individuelles de décéder de la crise en 1993, selon la taille de la fratrie (de 1,4 % pour les fratries de 2 à 3,6 % pour les fratries de 7).

FIGURE 5 Distributions «théoriques» et observées des fratries d'enfants des enquêtés de chaque taille (en survivants début 1993) et des couples de parents des enquêtés de deux survivants début 1993, selon leur nombre de membres tués par la crise en 1993 (ordonnées: échelle Log)



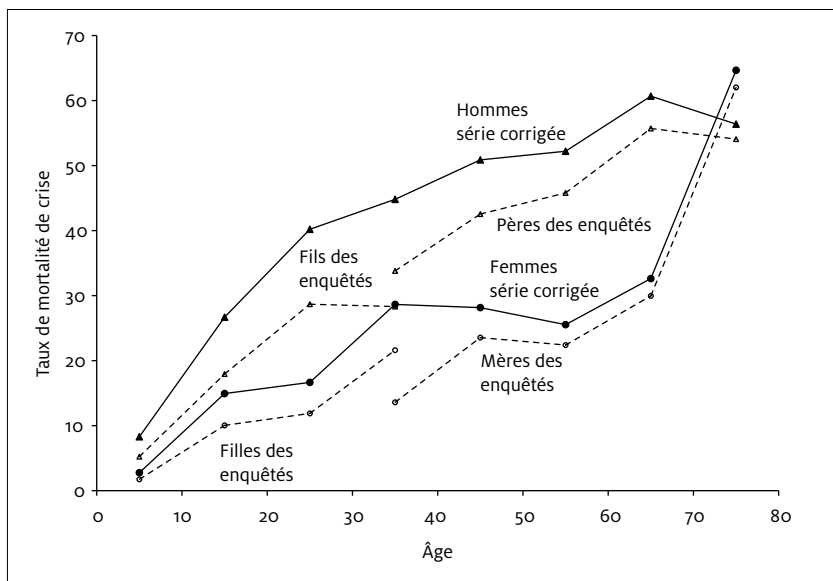
qu'ils ne sont que 32 % dans ce cas, les 68 % restants ayant été tués avec au moins un de leurs frères ou sœurs, 44 % l'ayant été avec au moins deux de leurs frères ou sœurs (la proportion « théorique » correspondante étant de 1 %) et 25 % d'entre eux l'ayant été avec au moins trois de leurs frères ou sœurs (proportion « théorique » d'environ 0,05 %).

Au total, cette corrélation se traduit par un indicateur de l'ordre de 2 % pour les fratries d'enfants déclarés par les enquêtés et de l'ordre de 2,5 % pour les couples de père et de mère déclarés par les enquêtés. Si l'on se réfère à l'échelle approximative définie par les 416 populations micro-simulées, on obtient de telles valeurs pour ces indicateurs lorsque la corrélation familiale de mortalité de crise est d'environ 0,5. Combinés, via les deux modèles d'ajustement précédemment déterminés, avec une descendance atteinte standardisée de 3,16 enfants par femme de 15 ans et plus de l'enquête Burundi-2002, ces indicateurs imparfaits de corrélation de la mortalité de crise au sein des fratries d'enfants et des couples de parents déclarés par les enquêtés permettent d'estimer la sous-estimation affectant les taux de mortalité de crise par groupe décennal d'âges. Très importante pour les 0-9 ans (augmentation de 58 % de la valeur du taux), cette correction est plus modeste pour les groupes plus centraux, et presque négligeable pour les 70-79 ans (augmentation de 4,5 % de la valeur du taux). La figure 6 présente les résultats de ces corrections. On constate que le choix de la valeur maximale observée, lorsqu'on peut calculer pour un même groupe d'âge un taux à partir des déclarations des enquêtés sur la survie de leurs enfants et un autre à partir des déclarations des enquêtés sur la survie de leurs parents, est une précaution nécessaire mais non suffisante.

En appliquant les taux par âge aux résultats de la projection 1990-1993 de la population recensée en 1990, on peut estimer le nombre de personnes tuées par la crise en 1993. En utilisant, les taux de mortalité de crise non corrigés (figure 1), on obtient environ 118 000 personnes tuées au Burundi en 1993 (Bergouignan, 2004), en recourant aux taux corrigés cette estimation atteint les 140 000 décès. Ce résultat dépasse le redressement conduisant à 126 000 décès de crise en 1993 (Bergouignan, 2004). Fondé sur les mêmes analyses de corrélation de la mortalité de crise au sein des fratries d'enfants et des couples de parents déclarés par les enquêtés, ce redressement nécessitait des hypothèses lourdes en termes d'interactions entre phénomènes et structures familiales. Le traitement des informations à l'échelle individuelle par le processus de microsimulations permet de gérer ces interactions de façon plus précise sans recourir à ces hypothèses. Il est donc vraisemblable de considérer ces 140 000 décès de crise de 1993 comme

une meilleure estimation que les 126 000 décès de crise précédemment proposés pour la même année 1993.

FIGURE 6 Taux de mortalité de crise 1993 (pour 1000), séries corrigées et initialement tirées de l'exploitation de l'enquête Burundi-2002.



En se limitant à l'année 1993, les résultats obtenus, non corrigés (117 832 personnes tuées par la crise) et a fortiori corrigés (140 000 personnes tuées par la crise), sont plutôt supérieurs aux chiffres avancés par des sources dont la neutralité semble difficile à établir et se situant entre 50 000 et 100 000 décès de crise pour 1993. Source plus fiable et plus neutre, le rapport d'enquête des Nations Unies (Nations Unies, 1996), qui comprend de nombreux témoignages, ne fournit pas d'estimation du nombre de personnes tuées en 1993. Il est donc difficile de confronter les résultats ici obtenus à autre chose qu'à la cohérence de leur processus d'élaboration.

DISCUSSIONS ET EXTENSIONS

Cette expérience visant à corriger des estimations que l'on suppose courir un fort risque de biais montre tout le parti que l'on peut tirer des micro-simulations, notamment pour comprendre comment se combinent les facteurs de ces biais. Traitant les destins individuels, elles constituent un outil privilégié pour rendre compte des interactions entre les phénomènes

démographiques et les compositions familiales, objet ici essentiel puisque les estimations de la mortalité par enquête rétrospective s'appuient sur ces compositions familiales. Cette compréhension des facteurs de biais et de leurs contextes de manifestation aiguë permet d'appréhender les situations dans lesquelles l'ampleur des sous-estimations devrait être importante. Cette anticipation du degré de sous-estimation de la mortalité de crise mesurée à partir d'une enquête rétrospective sur la base des déclarations des enquêtés sur la survie de leur parenté biologique peut notamment s'appuyer sur des indicateurs imparfaits mais néanmoins mesurables à partir de l'enquête et corrélés avec les biais. Cette démarche d'estimation de l'ampleur des biais à partir d'indicateurs imparfaits tirés de l'enquête peut même être systématisée en construisant des ajustements statistiques, fondés sur la diversité des résultats d'un grand nombre de microsimulations dont on fait varier les paramètres. Cela conduit alors à disposer d'un outil de correction des indices susceptibles d'être biaisés. Autrement dit, cela peut se traduire par la possibilité de contourner les effets de sélection des enquêtes rétrospectives basant l'étude de la mortalité sur les déclarations des enquêtés sur la survie de leur parenté biologique. Liés aux corrélations familiales de la mortalité, ces effets de sélection sont particulièrement fréquents lors des crises politiques, en particulier si on y observe des massacres de masse voire des génocides, situations où l'estimation de la mortalité revêt précisément une grande importance.

Un tel recours aux microsimulations pose toutefois un certain nombre d'interrogations.

Tout d'abord, et cela vaut pour l'ensemble des outils de microsimulation, on peut se demander à quel point on ne retrouve pas dans les résultats d'une telle démarche ce qu'on y a mis, c'est-à-dire les paramètres structurant le processus. Cela revient aussi à se demander s'il n'est pas inévitable de ne pas retrouver dans les résultats les hypothèses qui n'ont pas servi à programmer la microsimulation. Assez légitime dans sa globalité, cette interrogation ne saurait conduire à considérer les microsimulations comme une démarche purement tautologique. S'agissant de l'étude des biais de sélection associés à l'étude de la mortalité par enquête rétrospective, une grande partie du problème est prédéterminé :

- par la définition de l'opération collecte (enquêter, selon un processus d'échantillonnage particulier, certaines personnes nécessairement vivantes à la date de l'enquête et leur poser des questions précises) et
- par les structures familiales correspondant aux catégories de parents dont l'enquêté va rapporter la survie.

Autant d'éléments dynamiques et statiques essentiels que l'auteur de la microsimulation doit prendre en compte sans pouvoir les moduler selon ses choix, sauf en matière d'outil de programmation, ce qui a une influence modérée sur les résultats. La microsimulation proposant de reproduire une réalité en général très simplifiée, il est toutefois possible d'omettre certains facteurs influençant le phénomène étudié. Dans le cas de l'étude de la mortalité de crise au Burundi en 1993, la mortalité par une autre cause que la crise est corrélée à la mortalité de crise (Bergouignan, 2004), les jeunes orphelins (nombre d'entre eux l'étant par la crise) connaissant des conditions de vie plus difficiles et décédant dans de plus fortes proportions que les autres enfants du même âge. Pourtant, cette corrélation n'est pas introduite dans les paramètres de microsimulation, puisque la mortalité de crise y apparaît comme uniquement corrélée avec elle-même, via les parentés biologiques. L'introduction de cette hypothèse dans les programmes initiaux de microsimulation a toutefois conduit aux mêmes conclusions que celles obtenues sans cette hypothèse (le biais dépend des corrélations familiales de la mortalité et du niveau de fécondité), si ce n'est que la corrélation intra-familiale de la mortalité se mesurait en deux indices au lieu d'un et perdait en lisibilité²⁷. Les microsimulations servent à comprendre les combinaisons de facteurs engendrant des biais, et non à retracer les vies des Burundais dans toute leur complexité. Le programme de microsimulation pourrait connaître d'autres améliorations pour approcher de façon plus fine les facteurs de biais, notamment en considérant plus finement l'âge dans la construction des indicateurs « imparfaits » de corrélation familiale de la mortalité. Il serait de même possible d'ajouter à ces indicateurs intra-générationnels un indicateur inter-générationnel de corrélation familiale de la mortalité, de construction plus complexe, mais tout aussi « imparfait ». Le coefficient de détermination du modèle de correction des biais construit à partir des populations microsimulées pourrait alors augmenter un peu, sachant que les marges de progrès sont assez limitées (environ 16 points de pourcentage, voir supra).

L'impossibilité technique et l'apport heuristique modéré d'une microsimulation reproduisant le plus fidèlement possible la complexité des comportements passés entraînent un second questionnement. Est-il légitime de se baser sur une série de réalités simplifiées pour construire un

27. Dans ces microsimulations, cette corrélation entre mortalité de crise et mortalité par une autre cause a d'ailleurs tendance à moins affecter les taux de mortalité de crise que les taux de mortalité par une autre cause, en raison de l'importance relative de la mortalité de crise dans la mortalité globale.

ajustement statistique, s'appuyant ainsi sur un échantillon « virtuel » ? Si les microsimulations ne sauraient parvenir à reconstituer toute la complexité des destinées humaines, deux orientations, ici expérimentées dans les microsimulations des 416 populations, recréent une importante variabilité dans les résultats. D'une part, il est facile de programmer un plus large spectre de variation des comportements entre les individus d'une même population microsimulée. D'autre part, on peut faire varier sur une très large étendue de valeurs (recouvrant celles observées dans les populations humaines réelles), les paramètres de microsimulation des différentes populations. Il est alors possible, pour les phénomènes correspondant à ces paramètres, de créer, par croisement systématique, une diversité de populations virtuelles supérieure à la diversité des populations réelles. Il est néanmoins difficile de dire si ces options d'augmentation de la diversité au sein de chaque population et entre les populations permettent de recréer une variété semblable à celle d'un échantillon de populations réelles. Si cet échantillon de populations réelles présentait, toutes choses égales par ailleurs, une variance plus faible pour chaque phénomène, on pourrait néanmoins y observer une variabilité conditionnelle plus complexe née de synergies entre phénomènes aux échelles micro et macro, ce dont l'échantillon de populations virtuelles ne peut rendre complètement compte.

Enfin, cette analyse des biais de sélection via les microsimulations perdrait toute pertinence si la corrélation familiale de la mortalité de crise s'approchait de 1, c'est-à-dire si elle tendait à n'opposer que des familles indemnes et des familles entièrement tuées. Néanmoins, les microsimulations montrent ici que même lorsque la corrélation est très élevée ($R = 0,85$), les taux calculés conservent une signification et l'analyse des facteurs de biais reste très cohérente avec celles conduites pour des niveaux plus faibles de corrélation. Autrement dit, l'impossibilité d'analyser les résultats des microsimulations débute pour des niveaux de corrélation familiale de mortalité de crise encore supérieurs. C'est la raison pour laquelle on a pu étendre les corrections issues de ces microsimulations aux enquêtes démographiques et de santé 2000 et 2005 du Rwanda, bien que les corrélations familiales de la mortalité y soient plus élevées en raison de la durée des massacres de masse (d'avril à juin 1994). Les différences de modalités de l'enquête (femmes âgées de 15 à 49 ans déclarant la survie de leurs frères et sœurs et non celle de leurs parents, sans distinguer les décès de crise) modifient peu le programme de microsimulation.

Au final, malgré ses imperfections et celles des données d'enquête permettant d'étudier la mortalité de crise, la microsimulation est un outil

intéressant d'analyse voire de correction des biais de sélection. La fragilité de ces corrections est à la mesure de l'incertitude qu'elles visent à lever : très forte au Rwanda, assez forte au Burundi, moins dans d'autres contextes. C'est le cas de la mortalité des moins de 5 ans lorsque la forte prévalence du VIH et la transmission mère-enfant du rétrovirus viennent en biaiser l'estimation par des enquêtes rétrospectives où les femmes déclarent la survie de leurs enfants (Ward et Zaba, 2008 ; Hallett et collab, 2010). Les corrections fondées sur les microsimulations permettent alors de retrouver les résultats de corrections réalisées avec d'autres méthodes.

BIBLIOGRAPHIE

- BÉLANGER, A., E. CARON MALENFANT et L. MARTEL. 2005. *Projections de la population des groupes de minorités visibles, Canada, provinces et régions, 2001-2017*. Statistique Canada, 80 p.
- BERGOUIGNAN, C. 2004. *L'usage des confrontations en analyse démographique*. Rapport pour l'habilitation à diriger des recherches, Université Montesquieu-Bordeaux IV, 440 p.
- BERGOUIGNAN, C. 2007. « Les recours aux enquêtes rétrospectives dans l'analyse démographique des conflits », *Taking stock of data needs and monitoring for the millennium development goals in developing countries*, IUSSP seminar, 25-27 janvier.
- BERGOUIGNAN, C. et C. BLAYO. 2005. « Conséquences démographiques des conflits : méthodes et résultats de deux enquêtes (Kosovo 1999, Burundi 2002) », IUSSP international population conference, Tours (France), 18-24 juillet.
- BLACKER, J. et J. MUKIZA-GAPERRE. 1988. « The indirect measurement of adult mortality in Africa : results and prospects », African Population Conference, Dakar (Sénégal), 7-12 nov.
- BLAYO, C., C. BERGOUIGNAN, A. ETCHÉLECOU, V. CORDAZZO, M. LLOPART et N. OKOBO. 2004. *Situation démographique et sociale du Burundi*. Rapport remis au FNUAP Burundi, 405 p.
- BRASS, W. et K. HILL. 1973. « Estimating Adult Mortality in Africa From Orphanhood », dans *Proceedings of the International Population Conference*, Liège (Belgique) : 11-23.
- GAKIDOU, E. et G. KING. 2006. « Death by survey, Estimating adult mortality without selection bias », *Demography*, 43, 3 : 569-585.
- HALLETT, T. B., S. GREGSON., F. KURWA, G. P. GARNETT, S. DUBE, G. CHAWIRA, P. R. MASON et C. A. NYAMUKAPA. 2010. « Measuring and correcting biased child mortality statistics in countries with generalized epidemics of HIV infection », *Bulletin of the World Health Organization*, 88 : 761-768.
- HILL, K. et J. TRUSSELL. 1977. « Further Developments in Indirect Mortality Estimation », *Population Studies*, 31 : 313-334.

- LEDERMANN, S. 1969. *Nouvelles tables types de mortalité*, INED, 260 p.
- MAKINSON, C. 1993. « Estimates of adult mortality in Burundi », *Journal of Biosocial Science*, 25, 2 : 169-186.
- MASQUELIER, B. 2008. « Minorités ethniques et mortalité adulte en Afrique subsaharienne », dans *Démographie et cultures* (actes du colloque de l'AIDELF), Québec : 835-851.
- MASQUELIER, B. 2009. « Revisiting indirect methods based on orphanhood data with microsimulations », IUSSP international population conference, Marrakech (Maroc), 27 septembre-2 octobre.
- NATIONS UNIES. 1996. *Rapport sur le génocide de 1993 au Burundi*. 80 p.
- PENNEC, S. 1994. *Applications démo-économiques de la méthode des microsimulations*. Thèse de doctorat, IEP Paris, 474 p.
- TIMAEUS, I. 1986. « An Assessment of Methods for Estimating Adult Mortality From Two Sets of Data on Maternal Orphanhood », *Demography*, 23 : 435-50.
- TIMAEUS, I. 1992. « Estimation of adult mortality from paternal orphanhood : a reassessment and a new approach », *Population Bulletin of the United Nations*, 33 : 47-63.
- WARD, P. et B. ZABA. 2008. « The effect of hiv on the estimation of child mortality using the children surviving/children ever born technique », *South African Journal of Demography*, 11, 1 : 39-73.
- UNITED NATIONS. 1983. *Manual X : Indirect Techniques for Demographic Estimation*. 246 p.

ABSTRACT

Christophe BERGOUIGNAN

Studying crisis-related mortality using microsimulations: the case of Burundi in 1993

Frequently used for demographic projections, microsimulations can also appraise the selection bias of retrospective surveys, especially when mortality rate estimates are based on reported deaths of the relatives of those surveyed. Although these selection biases are commonplace, they are liable to be much more severe during disease epidemics or in crisis situations (natural disasters or conflicts). Microsimulations enable sensitivity analysis of the scale of these biases according to the phenomena that generate them. They are also a way to appraise the effectiveness of approximated correlation indexes based on survey data in estimating crisis-related mortality correlation among families and in correcting selection biases. Applied to crisis-related mortality rates estimated from the Burundi 2002 household survey, this correcting process leads to a reappraisal of the numbers killed in the crisis of 1993 in Burundi.