

Préface

La forme de la mortalité

Hervé Le Bras

L'idée d'une mortalité-type est à l'origine même de la démographie avec la première table de mortalité publiée dans les *Observations naturelles et politiques* de John Graunt (1977[1661]). On peut en effet montrer que les survivants de cette table, de 6 à 66 ans, sont divisés par deux à chaque décennie (Le Bras, 2001). Par la suite, le niveau de mortalité selon l'âge a suscité de nombreuses modélisations sous la forme de lois de mortalité. Les mathématiciens de Moivre et Simpson ont proposé chacun une fonction mathématique au XVIII^e siècle, puis le rythme s'est accéléré. Dans leur ouvrage, Ewa Tabeau et ses co-auteurs (Tabeau, van der Berg et Heathcote, 2001) signalent cinq lois proposées au XIX^e siècle, dont celles, restées célèbres, de Gompertz et de Makeham, puis dix-sept au XX^e siècle, dont celles de Weibull et de Heligman-Pollard (1980, p. 49-80).

Ces tentatives unificatrices se sont toutes heurtées à la réalité de tables de mortalité de plus en plus nombreuses à être établies et de plus en plus précises, qui se refusaient à suivre l'une ou l'autre de ces lois dont la multiplicité dénotait déjà l'inadéquation aux données. Or deux problèmes majeurs se sont posés aux démographes. Le premier est celui de l'estimation de la mortalité par âge quand on dispose seulement d'informations parcellaires sur les décès dans certaines classes d'âges, mortalité infantile ou mortalité des adultes assez jeunes, par exemple. Le second problème est celui du choix de tables prospectives de mortalité pour la réalisation des projections de population par âge, qui se sont multipliées après les premières tentatives de la fin des années 1920, prenant de l'ampleur après la Seconde Guerre mondiale.

Pour faire face à cette double demande à laquelle les lois mathématiques de mortalité ne permettaient pas de répondre, des catalogues de tables possibles correspondant à différents niveaux généraux de mortalité sont apparus.

Le premier a été établi par Vassilio Valaoras à la Division de la population des Nations unies en 1956, sous la forme de « tables-types » (en anglais « *model life table* »; Nations unies, 1956). Le principe de leur construction était simple. À partir de 157 tables de mortalité connues, fabriquer un ensemble de tables correspondant à tous les niveaux de mortalité, qui les résume. Pour cela, Valaoras est

parti de la mortalité infantile. Il a estimé le quotient de mortalité de 1 à 5 ans par régression parabolique sur cette dernière, puis les quotients quinquennaux suivants de proche en proche par régression sur les précédents (et non directement sur les quotients observés). Cette procédure évitait des fluctuations dans les tables. L'usage de ces tables était simple. Lorsqu'on connaissait le quotient à un âge donné, on cherchait dans quelle table-type se trouvait le quotient le plus voisin du même âge et on sélectionnait cette table.

La simplicité d'usage de ces tables et le patronage des Nations unies leur ont assuré un grand succès. Inévitablement, leur méthode a été critiquée car elle entraînait des biais en chaîne, se transmettant des quotients d'un âge donné au suivant et ceci jusqu'au dernier. Deux démographes et statisticiens, Gabriel et Ronen, ont montré ces biais et proposé de nouvelles tables correctement calculées, mais reposant sur le même principe de partir de la mortalité infantile (Gabriel et Ronen, 1958). C'est peu après que Sully Ledermann s'est emparé de la question avec des outils statistiques plus puissants. Il avait été sensibilisé aux différences sociales et nationales de mortalité dès un travail publié dans la revue *Population* en 1946 sur la mortalité des adultes dans lequel il pointait une énorme différence entre les Français et les Hollandais dont la mortalité entre 20 et 50 ans était trois fois plus faible, écart qu'il imputait à l'alcoolisme de nos compatriotes (Ledermann, 1946). Pour le prouver, il réalisa ensuite une importante étude, *Alcool, Alcoolisme, alcoolisation*, publiée dans les Cahiers de l'Ined (Ledermann, 1956). Puisque les quotients de mortalité à certains âges dépendaient du mode de vie, il était clair qu'une loi mathématique de la mortalité était impossible à formuler étant donné la grande variété des facteurs influant sur la santé, donc sur la mortalité à chaque âge de la vie. Dès lors, il fallait abandonner la recherche d'une loi mathématique. Les méthodes de la statistique mathématique offraient le seul moyen de mettre en évidence des régularités dans les tables de mortalité.

Sully Ledermann s'est attelé à cette tâche dans un article publié en 1959 avec Jean Bréas sur «Les dimensions de la mortalité», dans lequel il a développé l'appareil mathématique mettant à jour les caractéristiques principales de l'ensemble des 157 tables de mortalité utilisées par Valaoras (Ledermann et Bréas, 1952, p. 637-682). *A priori*, la méthode de l'analyse factorielle en facteurs communs et spécifiques était l'instrument *ad hoc*. Mais sa mise en œuvre est délicate et ses résultats sont sensibles aux erreurs. Ledermann a préféré l'analyse en composantes principales formulée par Karl Pearson (Morrisson, 1967). Elle consiste à estimer la combinaison linéaire des variables (ici les logarithmes des quotients par âge) dont la variance est la plus élevée. On peut se représenter l'opération comme la recherche d'un axe passant le mieux dans un nuage constitué par les tables de mortalité comme points dans un espace dont les 38 coordonnées correspondent aux 19 quotients par âge masculins et féminins. Une fois déterminée cette combinaison ou «composante», on répète l'opération pour les écarts qui subsistent et ainsi de suite. Les axes successifs ont la propriété d'être orthogonaux entre eux, donc indépendants (corrélation nulle).

Dans l'échantillon des Nations unies, la part de variance totale absorbée est de 81,5 % pour le premier axe, 7,4 % pour le deuxième et 3,7 % pour le troisième, soit globalement, 92,6 %. Les axes suivants ont donc une participation très faible à la variabilité des 157 tables. Les deux auteurs se sont alors concentrés sur les trois premiers axes, ce qui veut dire que chaque logarithme d'un quotient peut être représenté par une constante plus une combinaison linéaire de ses coordonnées sur les trois premiers axes avec une bonne approximation. Mais que signifient ces composantes ? Pour le savoir, les deux auteurs ont calculé leur corrélation avec divers aspects de la mortalité, l'espérance de vie, la mortalité sur telle ou telle plage d'âge, la mortalité infantile. Pour la première composante, l'espérance de vie à la naissance est la plus proche (corrélation $r = 0,974$), plus que la mortalité infantile ($r = 0,938$) utilisée par les Nations unies et par Gabriel et Ronen. Mais, l'interprétation des deux axes suivants n'est pas évidente si l'on se réfère à leurs coordonnées sur les 38 axes, donc les 38 quotients. Pour que l'on puisse les identifier à certains aspects majeurs de la mortalité, les auteurs effectuent des rotations du système des trois axes principaux. Dans un nouveau système, la part de variance de 92,6 % est conservée, mais les coordonnées des axes changent, ainsi que la part de variance dont chacun rend compte. Après plusieurs tentatives, trois nouveaux axes ont été adoptés : le premier correspond d'assez près aux quotients de mortalité des jeunes adultes, le deuxième à la fois à la mortalité de 1 à 4 ans et de 60 à 64 ans pour les hommes comme pour les femmes et le troisième aux âges élevés.

Le terrain est ainsi déblayé pour la construction rigoureuse de tables-types à laquelle Ledermann s'attaque les années suivantes. Il a constitué pour cela un échantillon de 154 tables. Son idée est simple : Maintenant que les principaux facteurs sont identifiés, il est possible de construire une bonne table-type avec d'autant plus de précision que l'on en utilisera un plus grand nombre. Pour la réalisation, il suffit d'effectuer des régressions simples si l'on n'utilise qu'un facteur, multiple si l'on en prend plusieurs, sur des proxys des facteurs ou « entrée », par exemple, l'espérance de vie pour le premier facteur auquel s'ajoute la mortalité des adultes avant 50 ans, pour le second facteur. Ledermann ne se limite pas aux meilleures entrées déterminées dans l'article de 1959, mais à celles dont on dispose facilement, ce qui donne des réseaux à une entrée : l'espérance de vie à la naissance, la mortalité infantile, le quotient de 1 à 5 ans, la mortalité jusqu'à 15 ans, celle de 20 à 30 ans, de 20 à 45 ans, la mortalité après 50 ans. Trois réseaux à double entrée sont aussi calculés, l'un à partir de la mortalité avant 5 ans et du quotient de 20 à 45 ans, le deuxième avec la mortalité avant 15 ans et celle, féminine, de 20 à 30 ans, le dernier avec la mortalité avant 15 ans et après 50 ans.

Étant donné les capacités de calcul actuelles, on ne publierait plus l'ensemble des réseaux (tables sexes réunis, hommes et femmes pour 16 valeurs de l'espérance de vie sexes réunis) qui occupe 175 pages de l'ancienne édition, mais à l'époque, surtout dans les pays du Sud, les ordinateurs étaient rares et peu savaient s'en

servir. Pour ceux qui disposaient d'un ordinateur, Ledermann avait fourni les coefficients des différents réseaux, ce qui permettait de calculer directement une table quand on connaissait la valeur d'un ou plusieurs entrées. Plus important, il a aussi donné pour chaque réseau les écarts-types de chaque quotient, qui mesurent l'imprécision, ce que les tables-types précédentes n'offraient pas.

Si l'ouvrage s'était borné à la construction rigoureuse des tables, cela aurait déjà constitué un net progrès, mais il s'accompagne d'une critique serrée des tables-types précédentes et d'un examen soigneux de la répartition des écarts des quotients autour de leurs estimations.

Un point épineux retient particulièrement l'attention, la différence des résultats selon que l'on travaille sur les quotients, les logarithmes des quotients ou les survivants (comme le fera William Brass ultérieurement) et, peut-on ajouter, les logits des quotients utilisés par Larry Heligman dans la dernière version des tables-types des Nations unies. Ainsi, les régressions sur les logarithmes des quotients sont sans biais sur les logarithmes des quotients, mais en ont un sur les quotients et les survivants. L'un des désaccords les plus spectaculaires concerne les espérances de vie. Dans le réseau « 101 » bâti à partir de l'espérance de vie, l'espérance de vie calculée à partir de la table ne correspond pas à l'espérance de vie en entrée. C'est logique, mais un peu déroutant. L'écart est cependant relativement modeste, de l'ordre de moins d'une année entre 35 et 70 ans, mais plus large pour les espérances de vie très basses ou très hautes (figure 5, chapitre 2). Comme le fait remarquer Ledermann, le problème ne se poserait pas si les tables-types avaient été construites à partir des survivants car l'espérance de vie en est une combinaison linéaire.

La question était-elle réglée? La seule amélioration que l'on semblait pouvoir apporter aurait été d'augmenter l'échantillon initial de tables en y incorporant de nouvelles, ce qui devenait de plus en plus réalisable avec les progrès de la statistique dans la plupart des pays du monde. Une autre méthode est cependant apparue. Elle a été élaborée à Princeton par Ansley Coale et Paul Demeny qui ont publié leurs *Regional model life tables* en 1966, quelques mois avant que Ledermann n'achève son manuscrit (Coale et Demeny, 1966). La base de données utilisée était constituée par 326 tables de mortalité masculines et autant de féminines. La construction des tables-types était à la fois plus pragmatique et plus compliquée, disons moins justifiée statistiquement. Les deux auteurs ont d'abord rangé par ordre croissant les 326 quotients d'un âge donné, puis fabriqué des tables standard en prenant tous les quotients d'un même rang. Ensuite, ils ont comparé les écarts des 326 tables réelles à ces standards. Ils ont alors conclu que l'on pouvait ranger les profils des écarts dans trois groupes, « Nord », « Sud » et « Est », qui correspondent *grosso modo* à des zones géographiques, et un certain nombre de tables qui s'écartaient des trois groupes, dans un quatrième groupe, « Ouest ». Chaque table a été calculée dans chaque groupe par une régression sur l'espérance de vie à 10 ans, e_{10} tantôt sur les quotients, tantôt sur leur logarithme, tantôt sur la moyenne des deux selon la valeur de e_{10} , procédure que les

auteurs eux-mêmes ont jugé « manquant d'élégance » mais qui permet de maintenir l'homoscédasticité, autrement dit, la variance des quotients estimés, à peu près constante.

Avec le recul du temps, il est tentant de comparer la précision des tables de Coale et Demeny à celle des tables de Ledermann. Auparavant, cependant, une analyse plus fine des quatre groupes « géographiques » est utile. Pour cela, on a recherché de quel groupe chacune des tables assignées par Coale et Demeny à un groupe, était la plus proche, selon quatre mesures différentes de la distance : somme des carrés des écarts entre les quotients de la table réelle et de la table type la plus proche, somme des valeurs absolues des écarts, et les mêmes sur les logarithmes des quotients. Seules 193 tables observées étant nommément assignées à l'une des quatre régions, on les a ventilées selon leur région et la région de la table-type régionale dont elles étaient la plus proche. En principe, toutes les tables assignées à la région Sud auraient dû être les plus proches de tables-types Sud et la même chose pour les autres régions. Ce n'est pas le cas, comme on le voit sur les tableaux suivants où l'on a croisé la région d'origine et la région de la table type la plus proche. Pour 147 tables masculines, les quatre mesures de la distance donnaient le même résultat, mais pas la concordance attendue (tous les chiffres auraient dû se trouver dans les cases diagonales). Seules 114 restaient dans leur région. Trente-trois étaient mieux ajustées par la table-type d'une autre région et pour 46, la région de la plus proche table-type variait selon la distance retenue parmi les quatre indiquées plus haut. Les résultats sont un peu moins discordants pour les tables féminines.

Région de la table-type la plus proche (hommes)

Région des tables	Ouest	Nord	Est	Sud	Plusieurs
Ouest	63	8	14	5	41
Nord	1	7	0	0	1
Est	3	0	24	1	3
Sud	1	0	0	20	1

Région de la table-type la plus proche (femmes)

Région des tables	Ouest	Nord	Est	Sud	Plusieurs
Ouest	79	8	10	0	34
Nord	0	7	0	0	2
Est	0	0	28	1	2
Sud	0	0	0	19	3

La région Ouest étant la plus mal définie, plus de la moitié de ses tables masculines sont plus proches d'une table d'une autre région pour au moins une des quatre distances et 27 d'entre elles sur les 141, sont toujours plus proches d'une autre région. Les résultats sont meilleurs pour les autres régions, ce qui tient au petit nombre de leurs représentants, mais il existe quand même des désaccords.

Il faut aussi tenir compte du fait que les tables qui ont servi au calcul des tables-types des régions Nord, Sud et Est sont peu nombreuses (62 sur 326), d'où un effet de sélection de petits noyaux éloignés au sein d'un *continuum* bien plus large.

Malgré ces défauts, l'approche par région et une entrée (l'espérance de vie à 10 ans) pourrait être plus précise que l'approche de Ledermann par deux facteurs. Le tester est possible. Pour cela, on a repris l'ensemble des 326 tables et on les a ajustées, d'une part, à la table régionale la plus proche (même si ce n'était pas la région à laquelle elles appartenaient) et, d'autre part, à la table la plus proche du réseau 201 de Ledermann (régression multiple sur le quotient de 0 à 5 ans et celui de 20 à 45 ans). On a calculé la distance globale des 326 tables à ces deux systèmes de tables-types selon les quatre distances définies plus haut. On a complété en calculant la distance moyenne au plan des deux premières composantes principales des 326 tables observées, ce qui offrirait la meilleure estimation en prenant des entrées plus proches de ces deux composantes, s'il y avait une seule région. Les résultats pour les tables masculines et féminines sont indiqués dans les deux tableaux suivants.

**Distance moyenne entre l'ensemble
des tables observées et estimées par table-type (hommes)**

Distance	Coale-Demeny	Ledermann	Composantes principales
Euclidienne	0,407	0,420	0,355
Valeur absolue	21,300	21,400	16,900
Euclidienne des Logs	385,300	83,400	52,200
Valeur absolue des Logs	485,100	367,700	269,600

**Distance moyenne entre l'ensemble
des tables observées et estimées par table-type (femmes)**

Distance	Coale-Demeny	Ledermann	Composantes principales
Euclidienne	0,258	0,402	0,216
Valeur absolue	16,000	20,600	14,100
Euclidienne des Logs	292,200	119,700	68,100
Valeur absolue des Logs	411,000	448,400	298,600

Les différences renvoient au mode de calcul des tables-types. Comme Coale et Demeny ont utilisé, en partie, les quotients directement, leurs tables sont un peu plus précises au sens de la distance mesurée sur les quotients. Au contraire, les tables de Ledermann construites sur les logarithmes des quotients sont nettement supérieures en termes de distance comptée selon les logarithmes. Dans un premier cas, les écarts en valeur absolue sont mieux pris en compte, dans le second, les écarts en valeur relative. Deux raisons font cependant pencher légèrement la balance en faveur des tables de Ledermann, l'une circonstancielle, l'autre de parcimonie. Circonstancielle car la comparaison a été menée sur l'échantillon qui a servi à établir les tables de Coale et Demeny en non sur celui de Ledermann,

ce qui le met en désavantage. Il était impossible de faire autrement car l'échantillon de Ledermann n'est pas régionalisé. Ce qui mène à la raison de parcimonie. La méthode de Ledermann est simple et fidèle aux canons de la statistique mathématique, celle de Coale et Demeny s'apparente en partie à du bricolage. Les tables ont été assignées aux quatre régions à vue d'œil et la méthode de régression est pour le moins non conventionnelle, avec ce mélange de quotients et de logarithmes de quotients.

Il semblait que la messe était dite ou plutôt les deux messes, l'une fondée sur l'existence de types séparés de mortalité, celle de Coale et Demeny, l'autre sur un seul type, celle de Ledermann. Cependant, un élément extérieur en a décidé autrement. L'une des deux raisons qui avaient mené à la construction des tables-types, à savoir, la possibilité d'estimer la mortalité dans des pays en développement où l'on ne disposait que d'informations partielles sur la mortalité par âge, se heurtait à la composition des échantillons. La grande majorité des 326 tables de Coale et Demeny et des 154 tables de Ledermann provenaient de pays développés. Elles pouvaient donc ne pas correspondre à des types de mortalité rencontrés en Afrique, en Amérique latine et en Asie. Les Nations unies ont alors décidé, en 1981, de construire de nouvelles tables-types de mortalité, spécifiquement pour les pays en développement (Nations unies, 1982). Cela a permis une synthèse entre la méthode de Ledermann et celle de Coale et Demeny, en partie pour des raisons de personnes. En effet, Larry Heligman, qui avait été étudiant de Coale, et l'auteur de ces lignes qui avait supervisé et préfacé l'édition des *Nouvelles tables-types* de Ledermann après sa disparition, ont été chargés de l'opération.

Trente-six tables masculines et 36 féminines couvrant l'Amérique latine, l'Asie et l'Afrique ont été retenues et leur validité soigneusement testée, puis à l'aide de méthodes de taxinomie numérique⁽¹⁾, quatre groupes de tables ont été nettement dégagés, laissant à part seulement quatre tables (Corée, Hong Kong, Koweït, Israël), ce qui revenait à adopter le principe régional, mais en le mettant en œuvre rigoureusement. Les logits des moyennes régionales ont été calculés et, au lieu des quotients de chaque table, on a retenu la différence entre le logit de chaque quotient par âge diminué du logit moyen du groupe pour cet âge. Une analyse en composantes principales a alors été effectuée globalement sur ces différences de logits. Le premier axe a concentré 90 % de la variance totale et les trois premiers 97 %, valeurs très proches de celles obtenues par Ledermann. Les coefficients de chacun des trois premiers axes et les termes constants ont été calculés pour établir directement le réseau de tables de chaque région, mais les tables publiées se sont limitées à la première composante principale. Le logit du quotient d'un âge donné dans un groupe donné y est obtenu par la somme du logit moyen à cet âge dans le groupe considéré et de la première composante principale de l'écart des logits.

(1) Une ultramétrique supérieure, avec vérification graphique des écarts entre les tables et la table moyenne «Ouest» du réseau de Coale et Demeny.