

Démographie et différences

Colloque international de Montréal (7-10 juin 1988)



ASSOCIATION INTERNATIONALE DES DÉMOGRAPHES DE LANGUE FRANÇAISE

AIDELF

Une méthodologie pour l'analyse comparative de la mortalité différentielle

• Graziella CASELLI*, Josiane DUCHENE** et Guillaume WUNSCH**

*Université «La Sapienza», Rome, Italie

**Université Catholique de Louvain, Belgique

La plupart des études sur la mortalité différentielle dans les pays développés utilisent, d'une part, les décès recueillis par l'état civil, et d'autre part, les renseignements individuels ou agrégés provenant d'un recensement. Certaines approches recourent au couplage des deux sources; d'autres comparent simplement les décès aux effectifs recensés selon des caractéristiques diverses. Etant donné la multiplicité des approches et des caractéristiques utilisées, l'analyse comparative de la mortalité différentielle dans le temps ou dans l'espace s'avère souvent difficile⁽¹⁾. Par ailleurs, malgré le grand nombre d'études réalisées dans ce domaine, nous sommes encore loin de pouvoir expliquer les différences de mortalité observées. En effet, le plus souvent, nous ne disposons que de simples descriptions de différences de mortalité entre groupes (par exemple, les différences entre catégories sociales⁽²⁾), où les seules variables « explicatives » prises en compte sont celles qui définissent les groupes, sans référence à une théorie sous-jacente. D'autres études n'avancent que des explications élaborées a posteriori, qui relèvent du sens commun et qui n'ont guère été confrontées aux faits.

La présente communication constitue une première étape dans la voie de l'élaboration d'un cadre conceptuel et d'une méthodologie pour l'analyse comparative de la mortalité différentielle, à partir des diverses sources statistiques disponibles et selon une optique cohérente et homogène.

I.- Choix de l'approche

Diverses approches peuvent être envisagées : individuelle ou agrégée, transversale ou longitudinale. La mortalité différentielle peut être étudiée pour l'ensemble des causes ou en distinguant les causes de décès. Nous avons décidé de privilégier une approche longitudinale qui part de données individuelles suivies et qui incorpore des variables contextuelles. Nous tenons compte, de ce fait, de l'histoire de vie de chaque individu

(1) Voir à ce sujet : Fox A.J. (1984) « Design problems and data collection strategies in studies of mortality differentials : developed countries », in J. Vallin et al (ed.) *Methodologies for the Collection and Analysis of Mortality Data*, Ordina, Liège, 89-108, ainsi que : Vallin J. « Facteurs socio-économiques de la mortalité dans les pays développés », *Délibérations de la réunion sur les facteurs et les conséquences socio-économiques de la mortalité*, Mexico, 1979, 266-302.

(2) Pour une bonne vue d'ensemble de ce sujet, voir Valkonen T. (1987) « Social inequality in the face of death », *Congrès Européen de Démographie 1987*, Jyväskylä, 1987, 201-261.

ainsi que des caractéristiques spatiales et temporelles de son cadre de vie. Par ailleurs, nous prendrons en compte les décès selon leur cause.

Comme la quasi-totalité des pays ne disposent pas des données nécessaires à une approche longitudinale, prospective et individuelle, plusieurs approches sont définies à titre palliatif :

a) Couplage des informations sur les décès survenus au cours d'une période qui suit *chaque* recensement avec les données individuelles enregistrées lors de ce recensement et de *tous* les recensements qui le précèdent. Cette approche est longitudinale tant pour les décès que pour les caractéristiques individuelles.

b) Couplage des informations sur les décès survenus au cours d'une période qui suit *chaque* recensement avec les données individuelles enregistrées lors de ce *seul* recensement. Cette approche est longitudinale pour les décès mais transversale pour les caractéristiques individuelles.

c) Couplage des informations sur les décès survenus au cours de la période qui suit le *dernier* recensement avec les données individuelles de ce recensement et de *tous* ceux qui le précèdent. Cette approche est transversale pour les décès mais longitudinale pour les caractéristiques individuelles.

d) Couplage des informations sur les décès survenus au cours de la période qui suit le *dernier* recensement avec les données individuelles de ce *seul* recensement. Cette approche est transversale tant pour les décès que pour les caractéristiques individuelles.

e) Faute de possibilité de couplage, il faut se rabattre sur une approche «écologique», basée sur des données agrégées au niveau d'unités géographiques plus ou moins homogènes du point de vue de la mortalité ou des variables explicatives. Trois cas sont possibles ici :

e-1) Approche longitudinale de la mortalité et utilisation des variables explicatives à plusieurs moments dans le temps⁽³⁾.

e-2) Approche transversale de la mortalité et utilisation des variables explicatives à plusieurs moments dans le temps.

e-3) Approche transversale de la mortalité et utilisation des variables explicatives à un seul moment dans le temps.

II.- Hypothèse de base

Dans la situation idéale, où il est possible de procéder à une approche longitudinale, prospective et individuelle de la mortalité par cause, le modèle estimera l'âge au décès en tant que minimum des âges au décès fournis par les divers modèles descriptifs spécifiques à chaque cause.

La variable à expliquer (dépendante) du modèle sera l'âge au décès de l'individu; les variables explicatives (indépendantes) seront réparties entre *marqueurs* et *facteurs* de risque. Selon la terminologie épidémiologique⁽⁴⁾, un marqueur de risque est une

⁽³⁾ On pourrait aussi envisager le cas peu intéressant d'une approche longitudinale de la mortalité et de l'utilisation des variables explicatives à un seul moment dans le temps.

⁽⁴⁾ Jenicek M., Cleroux R. (1982) *Epidémiologie*, Maloine, Paris.

caractéristique sur laquelle on ne peut guère agir mais qui détermine des personnes particulièrement vulnérables (âge, sexe, ethnie...) tandis qu'un facteur de risque est toute caractéristique (endogène et exogène) sur laquelle on peut agir ou dont l'effet peut être modifié (tabagisme, alcoolisme, qualité des soins de santé...).

Le modèle se base sur les trois hypothèses suivantes :

- 1) Les marqueurs de risque sont communs à toutes les causes de décès⁽⁵⁾.
- 2) Les facteurs de risque sont spécifiques à chaque cause de décès.
- 3) Les causes de décès sont indépendantes. On pourrait toutefois relâcher cette hypothèse en précisant le type de dépendance entre processus morbides.

Notons que ces hypothèses ne sont pas nécessairement compatibles : la présence de marqueurs communs à plusieurs causes peut introduire une dépendance entre âges au décès selon les causes. Dans la mesure où le modèle serait utilisé jusqu'à l'âge limite de la vie humaine, il serait nécessaire de tenir compte explicitement de la mortalité liée au vieillissement biologique (sénescence)⁽⁶⁾.

III.- Partition des causes de décès et sélection des marqueurs et facteurs de risque

Pour établir la partition des causes, nous nous basons sur la classification des causes de décès selon l'axe étiologique élaborée par Jacques Vallin et Alfred Nizard⁽⁷⁾. En effet, l'approche explicative de la mortalité différentielle requiert un nombre limité de groupes de causes homogènes du point de vue étiologique. Tenant compte de l'importance numérique des neuf classes de la typologie de Vallin et Nizard, il a été décidé de retenir parmi celles-ci les catégories suivantes :

- 1) *Dégénérescence* (il s'agit essentiellement des maladies cardio-vasculaires);
- 2) *Processus tumoraux*;
- 3) *Accidents et suicides*;
- 4) *Processus infectieux* (y compris ceux de l'appareil respiratoire);
- 5) *Malnutrition* (y compris cirrhose et psychose alcoolique);
- 6) *Autres*.

Quant à la sélection des marqueurs de risque, une distinction peut être opérée entre marqueurs fixes qui sont permanents et marqueurs stables qui peuvent être changés par une variation importante des conditions. Dans le premier groupe, on aurait par exemple le sexe ou le rang de naissance; le second groupe comprendrait entre autres le poids, la taille, le niveau d'instruction... En ce qui concerne les facteurs de risque, ils seront spécifiques aux six groupes de causes de décès définis ci-dessus⁽⁸⁾.

⁽⁵⁾ Ces marqueurs peuvent toutefois influencer différemment les diverses causes de décès; par exemple, le fait d'être du sexe féminin diminuerait le risque de mortalité ischémique du cœur mais augmenterait le risque de cancer du sein.

⁽⁶⁾ Voir à ce sujet : Duchene J., Wunsch G. (1986) « From the demographer's cauldron : single-decrement life tables and the span of life », *Working paper* no. 132, Institut de Démographie, Louvain-la-Neuve, juillet.

⁽⁷⁾ Vallin J., Nizard A. (1978) « Les causes de décès en France », *Population*, 3, 547-608.

⁽⁸⁾ Voir à ce sujet l'article des mêmes auteurs : *A methodology for the comparative analysis of differential mortality*, *Materiali di studi e di ricerca*, Dipartimento di Scienze Demografiche, Università di Roma « La Sapienza », n° 9, 1987.

IV.- Choix du modèle statistique pour une approche individuelle

L'hypothèse de base est que l'âge de décès est fonction de l'histoire de vie de l'individu depuis sa naissance, des conditions de sa gestation et de son patrimoine biologique. L'histoire de vie de l'individu est constituée par un ensemble de *états* successifs, chaque état étant défini par les modalités prises par les marqueurs et facteurs de risque présents. Par exemple, l'état de l'individu à la naissance peut être caractérisé par son sexe, son groupe sanguin, son rang de naissance, son statut social, etc. Au cours de sa vie, l'individu change d'état chaque fois qu'au moins un des marqueurs ou facteurs de risque change de modalité. L'âge au décès de l'individu est la somme des années vécues dans chaque état.

Cette hypothèse implique que le risque de décéder dans un état est fonction de cet état, des états antérieurs, de l'ordre de succession de ces états et du temps passé dans chacun d'eux. Ce modèle général peut être simplifié notamment en ne tenant pas compte de l'ordre de succession des états. Pour réduire encore le nombre d'états possibles, on peut caractériser l'individu par les marqueurs fixes ou stables devenus fixes à partir d'un certain âge, tel le niveau d'instruction, et ne définir les états qu'à partir des marqueurs non fixés et des facteurs de risque. On peut de plus regrouper les états à partir d'une analyse de classification utilisant un critère de densité.

Pour chaque état qui a été caractérisé par les modalités des marqueurs et facteurs, on relève, pour l'individu, la durée de vie passée dans cet état. Les marqueurs et facteurs qui définissent les états peuvent être soit individuels, soit contextuels.

Puisqu'à un âge donné, la survie ou le décès de l'individu dépend de la durée passée dans chacun des états, le modèle de risques (Hazard Model modifié de Cox) nous semble convenir pour traiter ce problème.

Dans de nombreux cas, on ne connaît que la modalité de la variable à un seul moment; dans ce cas, il est évidemment impossible de connaître le temps vécu avec cette modalité. Nous proposons néanmoins de tenir compte de cette information même partielle.

A titre d'exemple, supposons que nous disposions d'une observation continue et que l'individu considéré se caractérise comme suit sur les trois variables profession, résidence et groupe sanguin :

| | | | |
|------------------------|-------------------|-------------------|-------------------|
| Profession (j) | j_0 | j_1 | j_2 |
| Résidence (k) | k_1 | | k_2 |
| Groupe sanguin (i) | i_1 | | |
| Etats | (i_1, j_0, k_1) | (i_1, j_1, k_1) | (i_1, j_2, k_2) |
| Durées | D_1 | D_2 | D_4 |

On distingue quatre états caractérisés par les modalités prises par les variables considérées et à chaque état est associée une durée d'exposition au risque dans cet état.

Au niveau individuel, on peut appliquer le modèle de régression de Cox⁽⁹⁾. Considérons la probabilité qu'un individu décède entre les âges x et $x + n$, soit ${}_nq_x$; définissons

(9) Allison P.D. (1984) *Event History Analysis*, Sage, Beverly Hills.

ensuite le taux instantané de mortalité $\mu(x)$ par l'expression :

$$\mu(x) = \lim_{n \rightarrow 0} ({}_nq_x/n)$$

Les observations étant repérées en temps discret, nous supposons le taux instantané constant sur chaque période d'observation. Notons $\mu_r(x)$ le taux de mortalité instantané pour un individu standard fictif. Le modèle de régression de Cox se définit dès lors comme :

$$\ln \mu(x) = \ln \mu_r(x) + \sum_i b_i D_i(x) \quad [1]$$

où $D_i(x)$ est le temps passé dans l'état i et b_i sont les paramètres de régression.

On privilégie la génération⁽¹⁰⁾ et on ne considère pas la période en vue d'éviter les problèmes d'identification. Dans cette optique, le temps x sera l'âge ou le groupe d'âges de l'individu. Le modèle sera estimé par la méthode de la vraisemblance partielle de Cox⁽¹¹⁾.

En ce qui concerne les paramètres du modèle, la fonction $\mu_r(x)$ ne peut pas être interprétée. Par ailleurs, on ne tient pas compte de l'ordre de succession des états puisqu'une durée de temps $D_i(x)$ passée dans un état i a le même effet b_i sur la fonction standard quel que soit l'âge x , vu l'indépendance des paramètres b_i par rapport au temps. Enfin, dans son état actuel, le modèle est linéaire et additif puisque les variables explicatives ont des actions indépendantes sur le logarithme du taux instantané de mortalité.

Le modèle doit être appliqué séparément à chaque groupe de causes de décès, vu la nature différente des variables explicatives. Lorsqu'on étudie un groupe de causes de décès donné, on tiendra compte des sorties d'observation⁽¹²⁾ à la suite des décès par autres causes, des migrations, des troncatures d'observation, etc. On supposera que ces sorties d'observation ne sélectionnent pas les individus du point de vue des variables explicatives de la cause étudiée⁽¹³⁾.

V.- Choix du modèle statistique pour une approche «écologique»

Rappelons que l'approche écologique est un palliatif à l'approche individuelle développée ci-dessus. Afin de permettre une évaluation de la validité de l'approche, nous proposons dès lors les solutions suivantes.

En ce qui concerne la variable à expliquer (dépendante), celle-ci serait le nombre absolu de décès dans la zone considérée par âge, sexe et cause. Ce choix est motivé, d'une part, par l'objectif de voir quelles conclusions tirées sur la base de l'approche «écologique» restent valables sur le plan individuel⁽¹⁴⁾, et d'autre part, d'éviter les corrélations entre dénominateurs des variables dépendantes et indépendantes.

(10) Breslow N. (1984) « Methods for identifying mortality risk factors in longitudinal studies », in J. Vallin et al. (ed.), *Methodologies for the Collection and Analysis of Mortality Data*, Ordina, Liège, 367-391.

(11) Allison, op. cit., annexe A, 67-71.

(12) Comme on a supposé un risque instantané constant par période d'observation, on utilisera la formule d'Elveback pour éliminer les perturbations.

(13) Breslow, op. cit., p. 3.

(14) Il est évident que le nombre de décès observés dans une zone est la somme des individus décédés dans cette zone.

En ce qui concerne les variables explicatives (indépendantes), il est impossible, dans le cas présent, de connaître le temps vécu dans les divers états qui caractérisent, non plus les individus, mais les unités géographiques. L'utilisation de plusieurs recensements successifs pourrait permettre de relever les changements d'état de ces unités géographiques au cours du temps.

Le bulletin de décès reprend une information généralement peu exploitée : le lieu de naissance. L'hétérogénéité de la population selon cette caractéristique peut expliquer une part des différences observées de mortalité⁽¹⁵⁾. Pour tenir compte de ce fait, il serait utile, pour les régions à forte immigration et quand cela est possible, de distinguer les décès des natifs de ceux des immigrés et d'appliquer le modèle aux deux sous-populations en prenant en compte leurs caractéristiques propres.

Puisque la variable dépendante est le nombre total de décès par âge, sexe et cause, la structure de chaque zone selon ces trois variables doit être contrôlée, en introduisant, parmi les variables indépendantes, par exemple le nombre de personnes et le nombre de décès toutes causes du sexe et du groupe d'âges considérés. Par ailleurs, étant donné que dans les pays développés, on dispose de longues séries de recensements, on a avantage à utiliser le maximum d'informations à différents moments dans le temps. Une analyse écologique liant mortalité et variables explicatives à un seul moment du temps est basée sur l'hypothèse implicite et erronée que la mortalité d'une année dépend essentiellement des conditions de cette période⁽¹⁶⁾.

Nous proposons de maintenir, au plan écologique, le modèle statistique qui a été suggéré pour l'approche individuelle. Ce choix est opéré dans le but de comparer au mieux les résultats des diverses approches. En analyse transversale, le modèle peut être écrit comme suit, pour un sexe donné :

$$d_c^r(x) = P_r(x) M_r(x) p_c^r(x) \exp \left(\sum_i b_i D_{ir}(x) \right) \quad [2]$$

où :

d_c^r = le nombre de décès par la cause c observés à l'âge x dans une génération donnée et dans une zone géographique r ;

$P_r(x)$ = population d'âge x dans la région r au recensement;

$M_r(x)$ = taux de mortalité toutes causes à l'âge x dans la région r ;

p_c^r = proportion de décès par la cause c à l'âge x parmi l'ensemble des décès au même âge dans une unité géographique fictive;

$D_{ir}(x)$ = temps passé par l'unité géographique r dans l'état i par référence à l'âge x . Cette durée n'a de sens que si l'on dispose d'au moins trois recensements. Sinon chaque région n'occupera qu'un seul état et le temps passé dans cet état sera le même pour toutes les régions.

(15) Voir, par exemple : Young C. (1986) « Survival experience and cause of death profiles of birthplace groups in Australia », International Seminar on the Socio-Economic Aspects of Differential Mortality, Zamardi (Hongrie), 9-12 septembre 1986.

(16) Cette hypothèse est vraie pour certaines variables, telles une épidémie de grippe, mais ne l'est pas pour la plupart des facteurs (alcoolisme, tabagisme).

Remarquons par ailleurs que le modèle est équivalent à :

$$\ln(d_r^c(x)) = \ln(P_r(x)) + \ln(M_r(x)) + \ln(p_r^c(x)) + \sum_i b_i D_{ir}(x)$$

ce qui peut s'écrire :

$$\ln(M_r^c(x)) = \ln(M_r(x)) + \ln(p_r^c(x)) + \sum_i b_i D_{ir}(x)$$

ou enfin :

$$\ln(p_r^c(x)) = \ln(p_r^c(x)) + \sum_i b_i D_{ir}(x) \quad [3]$$

où :

$M_r^c(x)$ = taux de mortalité par la cause c à l'âge x dans la région r , soit $d_r^c(x)/P_r(x)$;

$p_r^c(x)$ = proportion de décès par la cause c à l'âge x parmi l'ensemble des décès au même âge dans la région r , soit $M_r^c(x)/M_r(x)$.

On a intérêt à développer l'approche écologique sur des zones les plus homogènes possibles qui ne coïncideront généralement pas avec un découpage administratif. Faut-il assurer l'homogénéité sur le plan de la mortalité ou au niveau des variables explicatives? A priori, nous suggérons d'homogénéiser sur les variables explicatives en vue de faciliter la démarche causale ultérieure.

VI.- Conclusions

Le cadre conceptuel présenté ici est essentiellement *descriptif*. Son intérêt est d'abord de dresser la liste des variables pouvant avoir une incidence sur la mortalité par cause et de servir ainsi de cadre de référence pour les études qui ne prennent en compte qu'une partie de ces variables. Il devrait permettre ensuite de mesurer l'incidence de chaque variable en contrôlant l'impact des autres, en distinguant marqueurs et facteurs de risque. Enfin, le présent schéma permet de faire le pont entre approches individuelle et « écologique ».

Dans son état actuel, le modèle ne permet pas l'établissement des cheminements de *causalité*. Vu la mauvaise connaissance des liaisons causales, il nous semble prématuré de proposer à ce stade un modèle de causalité.