

Nouvelles perspectives de l'analyse biographique
NEW PERSPECTIVES IN BIOGRAPHICAL ANALYSIS
NUEVAS PERSPECTIVAS DEL ANALISIS BIOGRAFICO

Daniel Courgeau and Éva Lelièvre

Volume 22, Number 1, Spring 1993

URI: <https://id.erudit.org/iderudit/010133ar>

DOI: <https://doi.org/10.7202/010133ar>

[See table of contents](#)

Publisher(s)

Association des démographes du Québec

ISSN

0380-1721 (print)

1705-1495 (digital)

[Explore this journal](#)

Cite this article

Courgeau, D. & Lelièvre, É. (1993). Nouvelles perspectives de l'analyse biographique. *Cahiers québécois de démographie*, 22(1), 23–43.
<https://doi.org/10.7202/010133ar>

Article abstract

In the last decade, life-event history analysis, also called failure time data analysis or survival analysis, has been widely adopted by demographers. This methodology in demography allows to overcome major hurdles especially when analyzing longitudinal survey data. This paper describes the new perspectives opened to research in that field, and is illustrated by new results and examples of research projects. The authors concentrate on four issues: the analysis of incorrect and imperfect data, the analysis based on more complex data and lastly the study of interaction between phenomena.

Nouvelles perspectives de l'analyse biographique

Daniel COURGEAU et Éva LELIÈVRE *

Au cours de la dernière décennie, les méthodes de l'analyse biographique se sont rapidement diffusées dans tous les domaines de la démographie. Bien qu'elles constituent une généralisation de l'analyse longitudinale classique, qui s'est développée depuis la fin de la Deuxième Guerre mondiale, il est intéressant de noter que ces méthodes sont d'abord apparues dans des études cliniques ou de fiabilité industrielle portant sur quelques dizaines d'observations. Puis, au cours des années 1970, elles se sont diffusées, principalement dans des revues non démographiques : *Applied Statistics*, *Biometrics*, *Biometrika*, *Journal of The American Statistical Association*, *Journal of the Royal Statistical Society*, *Technometrics*, etc.

Un terrain favorable est apparu en démographie avec le développement des enquêtes rétrospectives, telles que les enquêtes mondiales sur la fécondité de la fin des années 1970, qui ont libéré le démographe de la carapace rigide des données administratives (état civil, registres de population, etc.). À notre avis, l'utilisation de ces enquêtes a permis d'aborder de façon précise deux problèmes qui se posaient depuis déjà longtemps en démographie.

Le premier problème est celui de l'interaction, que la démographie classique avait évacué en traitant chaque phénomène séparément, se contentant d'éliminer l'effet des autres phénomènes, considérés comme perturbateurs. Nous avons abordé ce problème dès 1976, en utilisant des enquêtes rétrospectives de l'INED (Courgeau, 1976, 1977). Depuis, l'analyse de l'interac-

* Institut national d'études démographiques (INED), Paris. Une précédente version de cet article a été présentée oralement au Congrès européen de l'EAPS-UIESP, à Paris (citée par Keilman, 1993).

tion entre phénomènes démographiques s'est développée, permettant d'étudier des situations toujours plus complexes.

Le second problème est celui de l'hétérogénéité des populations, que l'analyse longitudinale classique avait du mal à intégrer, car certaines caractéristiques d'un individu (profession, lieu de résidence, etc.) peuvent changer en permanence tout au long de sa vie. Les méthodes de l'analyse biographique ont permis de prendre cette hétérogénéité en compte. Menken et al. (1981) ont ainsi réussi à analyser le divorce en fonction de diverses caractéristiques des individus. Depuis, ces méthodes ont été appliquées à de nombreuses analyses et nous avons généralisé leur utilisation en traitant conjointement l'hétérogénéité et l'interaction (Courgeau et Lelièvre, 1986).

En dépit de ces avancées, de nombreux problèmes restent irrésolus ou partiellement résolus. Nous allons aborder ici quatre de ces problèmes, auxquels nous nous sommes attaqués au cours des dernières années. Trois sont liés à l'utilisation de données incorrectes, de données incomplètes ou de données plus complexes que celles dont on se sert habituellement; dans le quatrième cas, nous rapporterons un essai visant à dépasser l'interaction entre phénomènes.

Pour traiter ces problèmes, nous utiliserons les données de deux enquêtes. La première enquête, sur la «biographie familiale, professionnelle et migratoire», dite plus simplement «triple biographie» (3B), a été réalisée en France en 1981 et porte sur les individus nés entre 1911 et 1935 (4602 personnes enquêtées). On en trouvera une présentation détaillée dans Courgeau et Lelièvre, 1989. La seconde enquête a été réalisée en Belgique en 1988 (445 couples enquêtés) en vue de tester la qualité des données obtenues rétrospectivement comparativement à celles que fournit le registre de population belge (Poulain et al., 1991).

DONNÉES INCORRECTES

On recueille très souvent les données biographiques de façon rétrospective, en faisant appel à la mémoire des personnes enquêtées. Il en résulte un risque élevé d'erreurs dans les données recueillies, risque d'autant plus grand que les événements ont eu lieu à une date plus reculée. La date des divers événements peut être erronée et, ce qui est plus grave, certains événements pourront être oubliés.

Il importe donc de tester la qualité de ces données obtenues rétrospectivement à l'aide de données sur les mêmes personnes recueillies tout au long de leur existence. Dans un pays disposant de registres de population, un tel travail est réalisable.

Cette enquête a pu être réalisée en Belgique avec le concours du Département de démographie de l'Université catholique de Louvain. Les registres de population sont tenus de façon satisfaisante depuis 1856 dans ce pays (voir Poulain, 1978, pour une critique précise de cette source en ce qui concerne les migrations internes), et la comparaison avec des données recueillies rétrospectivement auprès des mêmes individus y est possible.

On a mené l'enquête auprès de 445 couples ¹, en mettant les individus dans les conditions les plus défavorables : les deux conjoints étaient d'abord interrogés simultanément mais séparément, dans deux pièces différentes. À l'issue de l'interview, ils confrontaient leurs récits des événements de leur vie et tranchaient, à l'aide, en particulier, des documents dont ils pouvaient alors disposer (livret de famille, quittance de loyer, etc.), produisant un troisième récit. La consultation des registres, indépendamment de l'interview, a constitué une quatrième source décrivant les mêmes événements. L'échantillon a été limité aux couples dont l'un des conjoints est né entre 1933 et 1942; l'autre devait être né dans la période plus large allant de 1933 à 1947. Leur âge à l'enquête s'étale donc de 41 ans à 55 ans. Il s'agit de personnes mises en situation de se remémorer des événements parfois très anciens.

La comparaison des dates données par chaque conjoint et par les conjoints réunis à celles que fournit le registre de population est portée au tableau 1 ². Si la comparaison est satisfaisante pour les dates de mariage, de naissance ou de décès d'enfants, car il s'agit d'actes d'état civil, l'inscription d'un ménage au registre l'est moins. En effet, malgré l'obligation légale de déclarer le changement de logement dans un délai de huit jours, cette déclaration peut souffrir un certain retard, qui à quelques exceptions près n'excède toutefois pas le mois. En outre, le registre de population n'enregistre pas les change-

¹ Un test limité à 50 couples interrogés avait déjà été réalisé antérieurement (Duchene, 1985 ; Courgeau, 1985) : il conduit à des résultats voisins de ceux qui sont donnés ici, mais bien sûr moins précis.

² Nous laissons de côté ici les omissions qui portent essentiellement sur des séjours de très courte durée, en l'occurrence de moins de six mois (voir Poulain, Riandey et Firdion, 1991, pour plus de détails).

TABLEAU 1 — Omissions et erreurs dans les dates de mariage, de naissance des enfants, d'émancipation des enfants et de migrations du ménage (entre parenthèses, proportion [%] par rapport aux événements datés)

Datation	Mariage			Naissance des enfants			Émancipation des enfants			Migration du ménage		
	H	F	Couple	H	F	Couple	H	F	Couple	H	F	Couple
Nombre total d'événements	445	445	445	1 078	1 078	1 078	310	310	310	1 388	1 388	1 388
Nombre d'évén. datés par ménage et registre, dont :	440	445	445	993	1 054	1 057	172	207	209	1 019	1 075	1 213
Date exacte (à ± 1 mois près)	414 (94,1)	440 (98,9)	443 (99,6)	931 (93,8)	1 032 (97,9)	1 039 (98,3)	76 (44,2)	108 (52,2)	112 (53,6)	630 (61,8)	701 (65,2)	816 (67,3)
Antidatées d'un an ou moins	8 (1,8)	1 (0,2)	0 (0,0)	20 (2,0)	6 (0,6)	8 (0,7)	36 (20,9)	37 (17,8)	37 (17,7)	193 (19,0)	202 (18,8)	238 (19,6)
Postdatées d'un an ou moins	13 (2,9)	3 (0,6)	1 (0,2)	29 (2,9)	9 (0,8)	6 (0,6)	19 (11,0)	19 (9,2)	17 (8,1)	77 (7,6)	68 (6,3)	59 (4,9)
Antidatées de plus d'un an	2 (0,5)	0 (0,0)	0 (0,0)	6 (0,6)	3 (0,3)	3 (0,3)	24 (14,0)	30 (14,5)	33 (15,8)	82 (8,0)	73 (6,8)	66 (5,4)
Postdatées de plus d'un an	3 (0,7)	1 (0,2)	1 (0,2)	7 (0,7)	4 (0,4)	1 (0,1)	17 (9,9)	13 (6,3)	10 (4,8)	37 (3,6)	31 (2,9)	34 (2,8)

Source : Poulain, Riandey et Firdion, 1991.

ments de domicile hors du pays, et enregistre mal les émigrations (ou les immigrations) vers (de) l'étranger, en particulier celles des militaires ou des coopérants.

On observe sur le tableau 1 une très nette différence entre le mariage et la naissance d'enfants d'une part (dates exactes à plus au moins un mois près dans plus de 93 % des cas) et la migration du ménage ou l'émancipation des enfants d'autre part (dates exactes à plus ou moins un mois près dans seulement 44 % à 67 % des cas). Cependant ces derniers pourcentages augmentent lorsqu'on prend un intervalle de plus ou moins un an : pour les migrations, ils passent de 88,3 % chez les hommes à 90,3 % chez les femmes et à 91,8 % chez les conjoints réunis; pour l'émancipation des enfants ils passent de 76,2 % chez les hommes, 79,2 % chez les femmes et 79,4 % chez les conjoints réunis. Notons ici combien l'émancipation des enfants est un phénomène difficile à saisir, car il se produit parfois avec un ou plusieurs allers-retours : le pourcentage d'événements ant-datés de plus d'un an croît lorsque l'on passe des hommes aux femmes et aux conjoints réunis. On obtient d'ailleurs toujours une meilleure information des femmes que des hommes, et des conjoints réunis disposant de tous leurs documents que des femmes ou des hommes. Il est donc préférable, dans la mesure du possible, d'interroger les deux conjoints réunis ou, éventuellement, les femmes, qu'il s'agisse d'événements familiaux ou de migrations.

Il est également intéressant de chercher à analyser chez les enquêtés les caractéristiques qui influent sur leur mémorisation de divers événements et par conséquent sur leur témoignage. Cela a été réalisé dans un très intéressant travail (Auriat, 1991) qui montre que certaines catégories d'individus (travailleurs manuels, enquêtés ayant une forte mobilité, etc.) ont une plus forte probabilité d'omettre une migration et que certains types de mobilité (intracommunale, mobilité précédée d'une durée de séjour de moins de 10 mois, etc.) sont plus souvent oubliés.

Cependant, ces erreurs de datation n'affectent pas ou guère les résultats d'analyses, tant non paramétriques que paramétriques ou semi-paramétriques, que nous avons réalisées avec ces différentes données (Courgeau, 1991). Nous présentons ici les résultats d'une analyse paramétrique des durées de séjour de plus de six mois dans les différents logements.

Le tableau 2 donne les résultats de cette analyse, où nous avons fait intervenir la durée écoulée depuis le mariage et le

TABLEAU 2 — Analyse de mobilité géographique : effet de la durée écoulée depuis le mariage, de la durée de séjour en années et du statut d'occupation sur la probabilité de changer de logement, selon la source (paramètres estimés et écart type entre parenthèses)

Variables	Hommes (1260 séjours)		Femmes (1310 séjours)		Conjoints réunis (1314 séjours)		Registres (1189 séjours)	
	Modèle 1	Modèle 2	Modèle 1	Modèle 2	Modèle 1	Modèle 2	Modèle 1	Modèle 2
	Constante	-2,955 *** (0,174)	-2,322 *** (0,193)	-3,009 *** (0,179)	-2,391 *** (0,193)	-3,062 *** (0,183)	-2,264 *** (0,192)	-3,118 *** (0,187)
Début du séjour l'année du mariage	1,464 *** (0,172)	0,438 ** (0,174)	1,569 *** (0,177)	0,574 *** (0,178)	1,642 *** (0,182)	0,575 *** (0,181)	1,564 *** (0,186)	
Début du séjour entre 1 et 4 ans après le mariage	1,117 *** (0,160)	0,572 *** (0,160)	1,198 *** (0,157)	0,622 *** (0,158)	1,301 *** (0,161)	0,750 *** (0,161)	1,199 *** (0,166)	
Début du séjour entre 5 et 9 ans après le mariage	0,641 *** (0,164)	0,375 ** (0,165)	0,559 *** (0,162)	0,286 * (0,163)	0,684 *** (0,164)	0,444 *** (0,165)	0,707 *** (0,166)	
Nombre d'enfants en début de séjour	0,006 (0,040)	-0,016 (0,039)	0,052 (0,042)	0,042 (0,041)	0,051 (0,042)	0,023 (0,042)	0,033 (0,044)	
Logé par l'employeur		0,485 *** (0,091)		0,480 *** (0,090)		0,321 *** (0,078)		
Propriétaire		-2,431 *** (0,175)		-2,347 *** (0,166)		-2,538 *** (0,164)		
Durée de séjour	-0,113 *** (0,0077)	-0,056 *** (0,0076)	-0,116 *** (0,0076)	-0,058 *** (0,0075)	-0,119 *** (0,0077)	-0,064 *** (0,0075)	-0,104 ** (0,0076)	

* Résultat significatif au seuil de 10 %; ** : résultat significatif au seuil de 5 %; *** : résultat significatif au seuil de 1 %.

début du séjour considéré, le nombre d'enfants nés au début du séjour et, dans un second modèle, le statut d'occupation du logement donné par les hommes, les femmes et les conjoints réunis (cette information n'apparaît pas au registre).

Le modèle choisi suppose que l'effet de la durée de séjour diminue de façon exponentielle. Le quotient de migration (voir l'article cité pour le test de cette hypothèse, qui est vérifiée) s'écrit, à l'instant t :

$$h(t; \mathbf{z}) = \exp(\mathbf{z} \alpha + \beta t)$$

où \mathbf{z} est le vecteur des variables décrites plus haut, et α et β des paramètres à estimer qui donneront l'effet des caractéristiques et de la durée de séjour.

Le tableau 2 donne d'abord les résultats du premier modèle (caractéristiques du statut d'occupation exclues), estimés séparément sur les données des hommes, des femmes, des conjoints réunis et des registres de population³. Toutes les caractéristiques ont un effet très proche, quelle que soit la source utilisée. Le nombre d'enfants en début de séjour n'a aucun effet sur le quotient de migration, alors que la durée de séjour elle-même et la durée écoulée entre le mariage et le début du séjour influent fortement sur ce quotient.

Le tableau 2 donne également les résultats du second modèle, où intervient, en plus des caractéristiques précédentes, le statut d'occupation du logement, quelle que soit la source utilisée, avec des intervalles de confiance se recoupant parfaitement. L'effet des autres caractéristiques est réduit mais reste semblable à ce qu'il était lorsqu'elles intervenaient seules.

Cette analyse montre que des erreurs importantes sur les dates de migration et sur les durées de séjour n'entraînent pas de biais aussi importants dans l'analyse des quotients de migration en fonction de diverses caractéristiques de l'enquêté en début de séjour. Dans la majorité des cas, les résultats sont cohérents entre eux, quelle que soit la source utilisée : les quelques différences observées ne modifient pas les principales conclusions de l'analyse.

D'autres analyses ont pu être réalisées, en particulier sur les liens entre la naissance du premier enfant et la première migration après mariage (Courgeau, 1991), qui donnent des

³ Les paramètres ont été estimés avec le programme RATE mis au point par N. Tuma.

résultats encore très cohérents entre eux. Cependant, la mauvaise qualité des réponses données par les hommes sur l'âge au mariage de leur femme donne un effet affaibli de cet âge lorsqu'on l'introduit comme caractéristique dans l'analyse.

Nous pouvons donc dire que si les erreurs sur la datation de divers événements peuvent être importantes, il semble qu'elles *ne modifient pas ou modifient peu l'ordre logique dans lequel ces événements se produisent*. Cet ordre est correctement remémoré, les erreurs de datation jouant le rôle d'un bruit de fond d'où l'on peut extraire une information cohérente quelle que soit la source utilisée. La mémoire semble donc fiable, là où l'analyse l'exige. Il est cependant préférable dans tous les cas de recueillir l'information directement auprès de la personne qui a vécu l'événement plutôt qu'auprès d'un tiers. Il est également préférable, lorsque l'on réalise une enquête rétrospective, de recueillir l'information auprès des femmes, ou mieux des conjoints réunis, en encourageant la confrontation des dates données avec tous les documents disponibles dans la famille.

DONNÉES INCOMPLÈTES

Les données que l'on utilise habituellement sont très rarement complètes : pour qu'elles le soient il faudrait disposer de la biographie d'une génération observée jusqu'à la mort de son dernier membre. Lorsque l'on réalise une enquête rétrospective, on n'observe la biographie des individus survivants à l'enquête que jusqu'à la date de l'interview⁴. On a mis en place des méthodes permettant d'utiliser au mieux cette information pour estimer les probabilités de connaître les divers événements (Courgeau et Lelièvre, 1989 : 45-52). Lorsqu'on ne suit la biographie des enquêtés qu'à partir d'une certaine date (enquête rétrospective), on ne dispose que des événements survenus après cette date⁵. L'utilisation de ces données introduit des biais beaucoup plus difficiles à éliminer (Courgeau et Lelièvre, 1989 : 52-56). On recommande généralement de compléter une telle enquête par une enquête rétrospective réalisée lors du premier passage, pour éviter des hypothèses trop irréalistes.

⁴ On parle alors de données tronquées à droite.

⁵ On est cette fois en présence de données tronquées à gauche.

Il existe également d'autres types de données incomplètes que nous allons examiner plus précisément ici. Ainsi, lors d'une observation par panel, on demande la situation de l'individu (état matrimonial, lieu de résidence, etc.) à chaque passage, sans s'inquiéter de la date à laquelle s'est produit l'événement amenant une nouvelle situation (mariage, migration, etc.). Il s'agit dans ce cas de données tronquées par intervalles. Cela revient à dire que la variable aléatoire T (date à laquelle survient l'événement) est comprise entre les dates de deux passages des enquêteurs t et t' . L'hypothèse que la date de passage des enquêteurs est indépendante des événements démographiques enregistrés permet d'estimer correctement la probabilité que T soit comprise entre t et t' .

Les données de l'échantillon démographique permanent (EDP) de l'INSEE, dont nous avons commencé l'analyse, entrent parfaitement dans ce cas. Il s'agit d'un fichier exceptionnel, tant par sa taille (un peu plus élevée que celle d'un échantillon au 1/100 de la population vivant en France) que par les informations qu'il comporte, tirées des bulletins individuels des recensements depuis 1968 et des bulletins de l'état civil relatifs aux événements familiaux des mêmes personnes (mariages, naissances, décès, etc.). Il permet d'éliminer les problèmes liés à la mémoire, dont nous avons indiqué l'importance dans la première partie de cet article, mais aussi d'en estimer l'effet, en comparant la résidence occupée lors d'un recensement à celle qui a été déclarée lors du recensement suivant, par la question sur le lieu de résidence au 1er janvier de l'année du recensement antérieur.

Si l'analyse longitudinale des événements familiaux est facile à entreprendre car on dispose de leurs dates précises, celle de la mobilité spatiale ou professionnelle n'est pas directement réalisable car on ne saisit la résidence et la profession des individus qu'à certaines dates : dates des recensements ou des événements familiaux. Nous sommes bien là dans le cas de troncatures par intervalles : où les dates des migrations ou des changements professionnels sont comprises entre événements familiaux et recensements.

L'hypothèse selon laquelle les dates des recensements et celles des divers événements démographiques (migration, changement professionnel, etc.) sont indépendantes est parfaitement vérifiée. Cependant, du fait de l'éloignement relatif entre les dates des recensements, on utilise conjointement l'information aux dates des événements familiaux saisis par l'état civil et

également domiciliés dans l'EDP, bien que l'hypothèse d'indépendance risque ainsi de ne pas être vérifiée.

Nous avons dans ce cas développé des méthodes non paramétriques d'analyse de ces données incomplètes, en utilisant un algorithme proposé par De Guttola et Lagakos (1989), des méthodes d'estimation semi-paramétrique de l'effet de diverses caractéristiques sur ces données (Kim, De Gruttola et Lagakos, 1993), ainsi que des méthodes d'estimation paramétrique à partir de ces observations (Courgeau, 1993). Pour tester ces méthodes, nous avons artificiellement tronqué les données de l'enquête 3B en créant un fichier «vu au travers des recensements» et un fichier «vu au travers des recensements et de l'état civil». La figure 1 donne la fonction de répartition cumulée pour les premiers changements de logement.

On voit que dans les deux cas les courbes estimées sont proches de la courbe réellement observée, avec cependant une

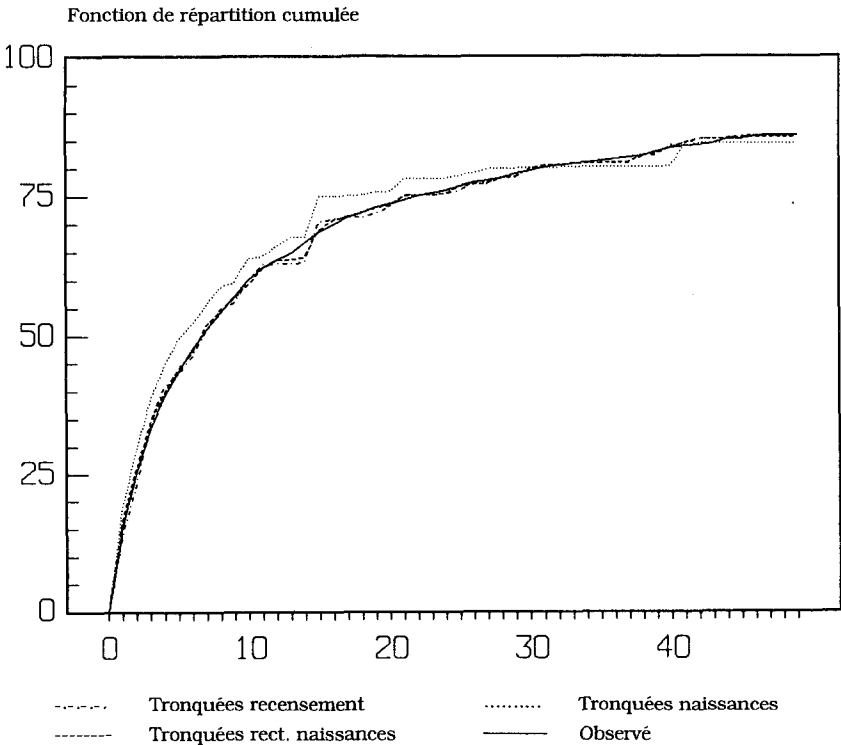


FIGURE 1 — Changements de logement
(données observées ou diversement tronquées)

amélioration pour la courbe utilisant à la fois les données de recensement et celles des événements familiaux. Cette amélioration apparaît clairement sur les variances des estimations : lorsque l'on ne fait intervenir que les données de recensement, la variance est de $4,10 \times 10^{-4}$ à une durée d'un an et de $4,79 \times 10^{-4}$ à 10 ans, contre respectivement $1,03 \times 10^{-4}$ et $2,80 \times 10^{-4}$ lorsque l'on fait intervenir également les événements familiaux. On arrive donc à une estimation convenable en utilisant ces données incomplètes, tronquées par intervalles.

Une autre situation se produit en démographie historique lorsque l'on utilise seulement les registres d'état civil, qui renseignent sur la résidence des intéressés lors de l'enregistrement des divers actes. L'enquête sur la mobilité sociale, géographique et patrimoniale en France au XIXe et au XXe siècle (Dupâquier, 1981) reconstitue la généalogie descendante, en ligne masculine, de 300 couples repérés sous le Premier Empire.

On dispose dans ce cas, pour chaque individu enquêté, non pas de la date des migrations successives, mais seulement de celle des événements familiaux encadrant une ou plusieurs de ces migrations.

Dans un tel fichier, l'hypothèse d'indépendance entre les événements dont on estime l'intensité (les migrations) et ceux par lesquels on a l'information (les événements familiaux) est sans doute moins bien vérifiée que dans le cas précédent, où l'on utilisait en plus l'information aux divers recensements.

Nous avons également porté sur la figure 1 l'estimation que l'on obtient lorsque l'on ne dispose que des résidences aux divers événements familiaux. Il apparaît clairement, pour les changements de logement, que l'hypothèse d'indépendance n'est pas vérifiée et que la courbe obtenue est significativement différente de la courbe observée. La probabilité d'une nouvelle naissance est en fait beaucoup plus importante après un premier changement de logement (Courgeau, 1993). Ce résultat est cohérent avec ce que nous avons déjà observé (Courgeau, 1985) et conduit à une fonction de répartition qui tend beaucoup plus rapidement vers sa limite supérieure que dans le cas des données observées. Notons cependant que la probabilité de rester définitivement sédentaire n'est pas altérée par les troncatures (0,14).

Pour les changements à plus longue distance, comme les changements de commune ou de département, les deux fonctions de répartition se rapprochent fortement (données non présentées). Cela signifie que les changements de logement en

vue d'un accroissement de la taille de la famille se font essentiellement à l'intérieur d'une commune. On peut alors penser que cette observation sera à plus forte raison vérifiée dans le passé, où la commune marquait encore davantage les limites des interactions familiales. Les changements de commune décelés dans l'enquête française peuvent, sous cette hypothèse, être utilisés pour une estimation de la fonction de répartition des diverses migrations du passé. Bien entendu, il est nécessaire d'effectuer des tests plus précis de cette hypothèse, dans des pays où existent des registres de population. Si elle n'est pas vérifiée, d'autres méthodes devront être utilisées, qui restent encore à mettre en place.

DONNÉES PLUS COMPLEXES

Passage des données individuelles à des données de groupes plus complexes

Lors des travaux pluridisciplinaires conduits sur le thème des «Approches longitudinales», le groupe, formé de 19 chercheurs de disciplines différentes, avait tenté de structurer dans un cadre commun le champ des travaux longitudinaux (Lelièvre et Courgeau, 1991). Si l'objet d'étude «histoires de vie» est le même, chaque discipline l'aborde, le collecte et le traite différemment. La définition commune adoptée rendait compte de cette diversité : «les approches longitudinales ont pour caractéristique l'étude d'événements ou d'états, objectifs ou subjectifs, dans leur succession et leurs interactions en rapport avec un temps historiquement défini, survenus à une *même entité* (*individu, famille, organisation, etc.*) au sein d'un *groupe bien défini* (*génération, promotion, etc.*). Les approches longitudinales veulent répondre à des objectifs précis, par divers modes de recueil des données et par des méthodes d'analyse et des modèles particuliers». Le passage de l'individu à une entité collective, à un groupe plus complexe était déjà prévu à ce stade conceptuel, et il est le fait de nombreuses disciplines. L'étude de la famille et des ménages dans leur dynamique est plus récente en démographie, bien que les divers événements marquant les étapes de transformation de ces unités soient l'objet principal de l'étude démographique.

Dans une perspective biographique, l'étude d'un groupe plus complexe (ménage ou famille) se propose de révéler la logique de l'influence que deux strates exercent l'une sur

l'autre. Elle veut mettre en évidence le pouvoir du groupe sur le devenir d'un individu et, réciproquement, montrer comment l'acteur individuel peut influencer une action collective. Cette démarche se place donc dans une optique qui n'est pas celle de la description des différents types de familles et de l'étude de l'évolution de leur répartition. Elle partage, par contre, ses préoccupations avec l'étude des conséquences démographiques de la structure des ménages et de la famille dans ses deux aspects : les changements de l'environnement familial des individus au cours du cycle de vie et, réciproquement, les effets des caractéristiques du ménage et de la famille sur les processus démographiques individuels (Bongaarts, 1983; Courgeau et Lelièvre, 1993).

Cette démarche nécessite néanmoins la collecte de données longitudinales sur le ménage, ce qui suppose un grand nombre de repérages préalables (Bonvalet, 1993). Un ménage rassemblant un nombre variable de membres doit être défini à l'aide d'au moins une caractéristique stable dans le temps, qui peut être de nature très diverse. On peut par exemple décider de suivre le ménage de l'individu X, ménage dont on recueille les caractéristiques au long de la trajectoire individuelle de X, ou encore on peut sélectionner un ménage défini par référence à un parent de l'individu X. Celui-ci peut alors ne faire partie du ménage que pendant une période de sa trajectoire, mais on continue d'observer les caractéristiques du ménage après le départ de X comme on les avait recueillies avant son arrivée (naissance, alliance etc.).

Ainsi, selon la première optique, l'enquête 3B, qui saisit à chaque moment la présence d'un conjoint et de divers enfants auprès de l'enquêté, permet de suivre les variations de la taille des ménages dans lesquels vivent les enquêtés. Nous en avons fait une première analyse (Courgeau, 1994) en suivant deux approches différentes pour modéliser la dépendance temporelle des divers quotients à estimer. Il est d'abord possible d'utiliser pour les divers épisodes un temps commun, qui partirait de la date de la formation du ménage avec au moins un membre. Dans ce cas, les instants successifs d'augmentation ou de diminution de la taille de ce ménage sont comptés à partir de cette origine. Une telle approche, en se concentrant sur les durées depuis l'instant initial, suppose que la durée passée dans un état influe peu sur le quotient de sortie de cet état. Cette hypothèse nous semble peu vraisemblable dans la mesure où l'entrée et le séjour dans une nouvelle étape devraient à notre

avis modifier les chances d'en sortir. Dès lors, une autre approche, utilisant les durées de séjour dans chacun des états, est envisageable : l'horloge utilisée se remet à zéro lorsque le ménage entame une nouvelle étape de son parcours.

L'examen des résultats obtenus (Courgeau, 1994) montre que la prise en compte de la durée passée dans l'état précédent conduit à des estimations plus discriminantes, rendant les résultats plus clairs et plus faciles à interpréter que lorsque l'on travaille avec la durée totale du processus. Le choix entre les deux horloges pourrait cependant ne pas être exclusif : en effet, la durée totale du processus est la somme des durées des étapes successives. Une autre solution, plus complexe mais plus satisfaisante, serait de définir le temps dans un espace à deux dimensions ou plus, chacune de ces dimensions représentant la durée écoulée entre divers événements initiaux que l'on peut choisir et l'événement étudié. Cependant, la complexité de tels modèles restreint l'estimation à des cas très simples (Andersen et al., 1992).

Bien entendu, cette approche doit être poursuivie avec des données plus pertinentes car elle semble pouvoir procurer une dimension fondamentale qui échappe au recueil des caractéristiques individuelles : celle de la structure et de son impact sur les décisions individuelles. Deux projets en cours en Inde et au Sénégal ont mis ces réflexions à profit lors de deux procédures d'enquête portant sur l'insertion des migrants en ville : projet sur la dynamique des villes secondaires et les processus migratoires en Inde (Dupont, ORSTOM) et projet sur l'insertion urbaine des migrants à Dakar (Antoine, ORSTOM-IFAN).

Le cas de l'insertion urbaine des migrants, le passage de l'individu à son espace de vie

Le maintien de forts courants migratoires dans les pays en développement a amené les chercheurs à se pencher sur les différents contextes de ces flux afin d'identifier des particularismes ou des similitudes dans ces processus. En effet, les migrations semblent «constituer des réponses individuelles ou collectives permettant au ménage ou à la famille d'élargir son espace de vie et de modifier la structure de ses activités économiques en vue d'assurer les moyens de sa reproduction» (Antoine et al., 1990). De plus, l'étude de la dynamique urbaine ne peut se concevoir qu'intégrée dans l'environnement rural et ne devient alors compréhensible «qu'en replaçant l'observation vers les villages, où se façonnent les stratégies familiales dans

lesquelles s'inscrivent les déplacements des individus vers la ville» (Dupont et Lelièvre, 1990).

L'information visée ici n'est pas seulement biographique au sens d'une collecte longitudinale portant sur des données de domaines différents, mais elle est de nature à saisir des stratégies individuelles et des réseaux sociaux, parentaux, culturels. Au delà de l'individu, le contexte de ses actions et les modifications de ce contexte sont à l'étude.

À ce stade, de nouvelles perspectives conceptuelles se présentent pour l'analyse biographique et ce sont elles qui ont motivé la mise en place, dans les deux projets, d'un système d'observation complexe. Plusieurs types d'enquêtes auprès de différents interlocuteurs et selon divers angles d'observation géographiques (de la ville aux villages d'origine) ou méthodologiques (études anthropologiques, psycho-sociologiques associées) ont été adoptés.

Sur le plan formel, l'analyse de l'interaction entre événements doit être repensée en termes d'analyse des interactions entre individus pour l'étude d'un processus complexe faisant intervenir plusieurs domaines : l'étude de l'insertion des migrants en termes de stratégie familiale par exemple. Pour cela, le recueil adopté dans ces deux projets semble bien adapté, à la restriction près que pour l'instant sa nature paraît plus qualitative que quantitative, ces données se prêtant difficilement à un traitement formel.

DÉPASSEMENT DE L'INTERACTION ENTRE PHÉNOMÈNES

Existence de situations plus complexes que celles d'interactions entre phénomènes

La mise en œuvre d'une analyse d'interaction nécessite l'identification de phénomènes distincts dont on teste la concurrence à des moments clefs de la trajectoire individuelle. Cette démarche a l'avantage d'offrir une structure à une interaction particulière qu'on ne saurait qualifier autrement. En effet, on peut constater l'amplitude du célibat masculin dans l'agriculture mais sans véritablement en identifier le mécanisme à moins d'en faire l'analyse dans une perspective longitudinale. De même, des différences dans la fécondité des couples apparaissent entre zones urbaines et zones moins urbanisées, et c'est seulement grâce à une analyse d'interaction que le rôle des migrations peut être mis en évidence.

Cependant, si le choix de phénomènes distincts s'impose facilement dans les exemples cités ci-dessus, d'autres interactions semblent plus difficiles à caractériser lorsque deux (ou plusieurs) processus distincts entrent en jeu. Dans le cas où l'apparition d'un phénomène détermine une modification radicale des conditions d'évolution d'un second processus qui lui-même affecte en retour le premier processus, les décisions qui s'ensuivent dans l'un ou l'autre des deux domaines peuvent paraître plus complexes, et leur mécanisme difficile à évaluer. En général, une analyse d'interaction ne réussit qu'à identifier une forte influence réciproque d'un événement sur l'autre sans que l'on soit à même de déterminer un axe de dépendance dominante.

On se trouve alors face à une dépendance circulaire que l'analyse n'a pas permis de dépasser. Il faut noter qu'avant la mise en œuvre du concept d'interaction, ce genre de problème ne pouvait même pas être identifié, dans la mesure où chaque phénomène constituait séparément un objet d'étude et que l'influence des autres processus était mesurée en l'absence d'évaluation de son influence propre sur ces autres processus.

L'exemple d'une telle circularité, où de nombreuses études (voir E. Bernhardt, 1988, pour un bilan) n'ont pas réussi à déterminer d'axe dominant, est l'interaction entre la fécondité et l'activité féminine. Le débat idéologique s'est alors souvent substitué à la recherche empirique ou théorique dans ce domaine, du fait de la complexité de l'interaction et en l'absence d'approche conceptuelle ou méthodologique susceptible de proposer une analyse dans un cadre formel scientifiquement satisfaisant.

Nous sommes bien ici dans un cas limite d'application de l'analyse des interactions entre phénomènes : le changement de l'effet de la maternité sur l'activité professionnelle selon l'âge des femmes. Sa mise en œuvre (voir Lelièvre, 1987) apporte des résultats éclairants. Cependant, une réponse globale n'est pas donnée par l'analyse, qui permet de constater une forte influence réciproque des phénomènes l'un sur l'autre, ceci dans un cadre formel certes moins restrictif que celui des analyses plus conventionnelles.

Ainsi, une réflexion critique sur la nature et les implications de ces limites (Kempeneers et Lelièvre, 1990) a révélé que l'impasse était plutôt de nature conceptuelle et que le problème devait être posé en des termes dépassant la simple interaction entre phénomènes.

Le travail féminin comme concept dépassant ceux d'activité et de fécondité

L'extrême complexité des arbitrages entre la famille et l'emploi provient en grande partie de la difficulté de prendre en compte de façon formelle les temporalités mises en jeu dans ce lien. Mais elle provient également de l'identification inadéquate des phénomènes, mal décrits. Cela se traduit ensuite par l'incapacité des méthodes de donner un éclairage satisfaisant sur l'interaction complexe.

En effet, «l'activité féminine» ne relève pas d'un domaine spécifique au sein des études démographiques ou économiques. De ce fait, les caractéristiques recueillies habituellement pour son analyse sont les mêmes que pour l'analyse de l'activité des hommes, ce qui conduit à des problèmes de spécification, dans la mesure où les modalités propres au travail féminin sont mal saisies. Ce constat conduit à critiquer les données et à retenir de l'analyse des interactions le principe de la saisie simultanée d'événements relevant de deux domaines distincts. On appréhenderait ainsi, au stade même de la collecte, le travail féminin comme un phénomène unique en ne dissociant pas la composante rémunérée de la composante domestique. Le nouveau type de données serait alors analysé à l'aide de méthodes biographiques *event history analysis* (Kempeneers et Lelièvre, 1990).

Dépasser l'interaction entre activité et fécondité se révélait nécessaire. L'existence d'une situation plus complexe a amené la redéfinition même des concepts de départ et à préconiser la saisie de l'interaction à sa source : en amont de l'analyse proprement dite. Ainsi, l'avancement proposé dans ce cas n'est pas du ressort d'une modélisation plus savante des phénomènes en jeu mais plutôt d'une amélioration supplémentaire des données. L'étude du *travail* féminin nécessite la collecte simultanée des données nécessaires pour saisir l'interaction entre la sphère domestique (fécondité, présence ou absence d'un conjoint) et la sphère rémunérée (modalités de l'activité, telles que salaire, horaire et contraintes du marché de l'emploi).

Une analyse biographique appliquée ensuite aux données complexes sur le travail féminin permettrait d'appréhender les arbitrages dans leur intégralité.

CONCLUSION

De nouvelles perspectives s'offrent à l'analyse biographique confrontée à des données incorrectes, incomplètes ou complexes. Notre première préoccupation a en effet été de tester la robustesse des méthodes d'analyse biographique *event history analysis* grâce à une collecte spécifique de données.

La plupart des échantillons disponibles pour de telles analyses proviennent d'enquêtes rétrospectives où l'information sollicitée a trait à des événements souvent anciens. Il convenait de s'assurer que la mémoire, si imparfaite soit-elle, était fiable pour l'utilisation des méthodes. Nous avons donc conduit le test sur quatre sources de données disponibles touchant les mêmes événements en évaluant la variation des estimations obtenues à partir des quatre échantillons.

Le deuxième axe de recherche exploré a été celui de la capacité des méthodes de prendre en compte des données incomplètes : tronquées par intervalles. De nouveaux développements formels, ainsi que le test des hypothèses de travail, ont alors permis de s'attaquer à des données historiques où le récit de vie est reconstitué par bouts à partir des renseignements collectés à chaque événement familial notifié dans les registres d'état civil, ainsi qu'aux données de l'échantillon démographique permanent de l'INSEE.

Les deux derniers défis évoqués ici sont liés à la complexité grandissante des processus dont on cherche à identifier les déterminants. Deux voies de recherche coexistent : le passage de l'analyse de données individuelles à celle de données biographiques concernant des groupes plus complexes (famille, ménage) et la prise en compte d'interrelations entre phénomènes dépassant la simple interaction. Des exemples illustrent ces deux démarches complémentaires, toutes deux résolument biographiques et visant à identifier des réseaux. Les deux premiers concernent l'insertion des migrants dans les villes des pays en développement, le troisième le travail féminin comme concept dépassant ceux d'activité et fécondité.

Ces étapes sont les premiers jalons d'une réflexion conceptuelle nécessaire à la poursuite du développement des méthodes formelles. Tout comme l'enquête 3B a permis la mise en place de l'analyse des interactions entre phénomènes, de nouvelles collectes constituent le préalable du développement de futurs modèles d'analyse.

RÉFÉRENCES BIBLIOGRAPHIQUES

- ANDERSEN, P. K., O. BORGAN, R. GILL et N. KEIDING, 1992. *Statistical Models Based on Counting Processes*. New York, Springer Verlag, vii, 768 p.
- ANTOINE, P., P. BOCQUIER, A. S. FALL et Y. GUISE, 1990. «Étude de l'insertion urbaine à Dakar : présentation de la méthodologie d'enquête», in É. LELIÈVRE et C. LEVY-VROELANT, éd. *La Ville en mouvement : habitats et habitants*. Paris, L'Harmattan.
- AURIAT, N., 1991. «Who Forgets? An Analysis of Memory Effects in a Retrospective Survey on Migration History», *European Journal of Population*, 7, 4.
- BERNHARDT, E. M., 1989. *Fertility and Employment*. Stockholm, Université de Stockholm, Research Report in Demography, no 55.
- BONGAARTS, J., 1983. «The Formal Demography of Families and Households: An Overview», *IUSSP Newsletter*, 17.
- BONVALET, C., 1993. *Structure des ménages et cycle de vie. Histoire familiale d'une génération de Parisiens*, «Document de recherche INED».
- COURGEAU, D., 1976. «Mobilité géographique, nuptialité et fécondité», *Population*, 31, 4-5 : 901-915.
- COURGEAU, D., 1977. «Interférences entre phénomènes démographiques», *Population*, 32, no spécial : 81-93.
- COURGEAU, D., 1985a. «Interaction Between Spatial Mobility, Family and Career Life-Cycle: A French Survey», *European Sociological Review*, 1, 2 : 139-162.
- COURGEAU, D., 1985b. «Effet de déclarations erronées sur une analyse de données migratoires», in JEZIERSKI, éd., *Chaire Quetelet : migrations internes*, Louvain-la-Neuve : 151-156.
- COURGEAU, D., 1991. «Analyse de données biographiques erronées», *Population*, 46, 1 : 89-104.
- COURGEAU, D., 1992. «Impact of Response Errors on Event History Analysis», *Population, An English Selection*, 4 : 97-110.
- COURGEAU, D., 1993. «An Attempt to Analyse Individual Migration Histories from Data on Place of Usual Residence at the Time of Certain Vital Events: France during the Nineteenth Century», in D. S. REHER et R. S. SCHOFIELD, éd. *Old and New Methods in Historical Demography*. Oxford, Oxford University Press : 206-222.
- COURGEAU, D., 1994. «Event History Analysis of Household Formation and Dissolution», in E. van IMHOFF, A. KUYSTEN, P. HOOYMEYER et L. van WISSEN, éd. *Household Demography and Household Modelling*. Londres, Plenum Publishing Corporation, «Demographic Methods and Population Analysis», à paraître.
- COURGEAU, D., et É. LELIÈVRE, 1986. «Nuptialité et agriculture», *Population*, 41, 2 : 303-326.
- COURGEAU, D., et É. LELIÈVRE, 1989. *Analyse démographique de biographies*. Paris, Éditions de l'INED, 210 p.

- COURGEAU, D., et É. LELIÈVRE, 1993. «Vers une analyse biographique des ménages». Communication présentée au Congrès général de l'UIESP, Montréal.
- DE GRUTTOLA, V., et S. W. LAGAKOS, 1989. «Analysis of Doubly-Censored Survival Data, with Application to AIDS», *Biometrics*, 45 : 1-11.
- DUCHÊNE, J., 1985. «Un test de fiabilité des enquêtes rétrospectives. Biographie familiale, professionnelle et migratoire», in JEZIERSKI, éd., *Chaire Quetelet : migrations internes*. Louvain-la-Neuve.
- DUPÂQUIER, J., 1981. «Une grande enquête sur la mobilité géographique et sociale aux XIXe et XXe siècles», *Population*, 36, 6.
- DUPONT, V., et É. LELIÈVRE, 1990. «La ville antenne villageoise, observations indiennes», Colloque international AIDELF, Rabat.
- KEILMAN, N., 1993. «Emerging Issues in Demographic Methodology», in A. BLUM et J.-L. RALLU, éd., *European Population, II. Demographic Dynamics*. John Libbey et INED : 483-508.
- KEMPENEERS, M., et É. LELIÈVRE, 1991. «Analyse biographique du travail féminin», *Revue européenne de démographie* : 377-400.
- KIM, M. Y., V. G. DE GRUTTOLA et S. W. LAGAKOS, 1993. «Analysing Doubly Censored Data with Covariates, with Application to AIDS», *Biometrics*, 49 : 13-22.
- LELIÈVRE, É., 1987. «Activité professionnelle et fécondité : les choix et les déterminants des femmes françaises entre 1930 et 1960», *Cahiers québécois de démographie*, 16, 2.
- LELIÈVRE, É., et D. COURGEAU, 1991. «Approches longitudinales», in F. STENDLER et P. WATIER, éd. *Interrogations et parcours sociologiques*. Méridiens Klincksieck.
- MENKEN, J., J. TRUSSEL, D. STEMPEL et O. BABAKOL, 1981. «Proportional Hazard Life Table Models: An Illustrative Analysis of Socio-demographic Influences on Marriage Dissolution in the United States», *Demography*, 18 : 181-200.
- POULAIN, M., 1978. «Du registre de population aux statistiques de migration interne en Belgique : critique des sources et correction des données», *Population et famille*, 45, 3 : 1-46.
- POULAIN, M., B. RIANDEY et J.-M. FIRDION, 1991. «Enquête biographique et registre belge de population : une confrontation des données», *Population*, 46, 1 : 65-88.

RÉSUMÉ — SUMMARY — RESUMEN

COURGEAU Daniel et LELIÈVRE Éva. NOUVELLES PERSPECTIVES DE L'ANALYSE BIOGRAPHIQUE

Au cours de la dernière décennie, les méthodes d'analyse biographique, dites aussi de durée de séjour ou encore de survie, se sont diffusées rapidement en démographie. Ces méthodes permettent en effet de répondre aux besoins de l'analyse démographique, en particulier en présence de données d'enquêtes longitudinales. Cet article présente les perspectives nouvelles qui s'offrent aux chercheurs dans ce domaine méthodologique en étayant l'exposé d'exemples concrets tirés d'analyses originales d'enquêtes ou de projets de recherche en cours. Les auteurs développent ainsi quatre nouveaux aspects : l'utilisation de données incomplètes ou incorrectes et de données plus complexes et enfin l'interaction entre phénomènes.

COURGEAU Daniel and LELIÈVRE Éva. NEW PERSPECTIVES IN BIOGRAPHICAL ANALYSIS

In the last decade, life-event history analysis, also called failure time data analysis or survival analysis, has been widely adopted by demographers. This methodology in demography allows to overcome major hurdles especially when analyzing longitudinal survey data. This paper describes the new perspectives opened to research in that field, and is illustrated by new results and examples of research projects. The authors concentrate on four issues: the analysis of incorrect and imperfect data, the analysis based on more complex data and lastly the study of interaction between phenomena.

COURGEAU Daniel y LELIÈVRE Éva. NUEVAS PERSPECTIVAS DEL ANÁLISIS BIOGRÁFICO

En el transcurso de la última década, muchos demógrafos han adoptado los métodos de análisis biográfico, llamados también de duración de la estadía, o de sobrevivencia. Efectivamente, estos métodos permiten responder más adecuadamente a las necesidades demográficas, especialmente cuando se trata de analizar datos de encuestas longitudinales. Este artículo presenta las nuevas perspectivas que se abren a los investigadores en este campo, ilustrándose por nuevos resultados y ejemplos concretos de análisis de encuestas o de proyectos de investigación. Los autores enfocan cuatro nuevos aspectos: el uso de datos incompletos o incorrectos, el de datos más complejos, así como el estudio de interacción entre fenómenos.