

Conditions de vie durant l'enfance et longévité : évaluation d'une base de données créée à partir du recensement canadien de 1901 et de l'état civil québécois

Early life Conditions and Longevity : linking data from the 1901 Canadian Censuses to Quebec's civil registration

Laurence Pilon-Marien, Alain Gagnon, Bertrand Desjardins and Robert Bourbeau

Volume 38, Number 1, Spring 2009

Enjeux de l'analyse démographique et nouvelles pistes méthodologiques

URI: <https://id.erudit.org/iderudit/039992ar>

DOI: <https://doi.org/10.7202/039992ar>

[See table of contents](#)

Publisher(s)

Association des démographes du Québec

ISSN

0380-1721 (print)

1705-1495 (digital)

[Explore this journal](#)

Cite this note

Pilon-Marien, L., Gagnon, A., Desjardins, B. & Bourbeau, R. (2009). Conditions de vie durant l'enfance et longévité : évaluation d'une base de données créée à partir du recensement canadien de 1901 et de l'état civil québécois. *Cahiers québécois de démographie*, 38(1), 171–191. <https://doi.org/10.7202/039992ar>

Article abstract

It has been suspected for a long time that early life conditions could affect adult mortality, and scholars endorsing a life course perspective have recently reported compelling evidence for the “long arm of childhood.” As this niche has not yet been developed in the Canadian context, we have established a record-linkage project tracing the early life conditions of individuals who were listed as children in the 1901 Canadian Census and Quebec's death records. This paper focuses on the difficulties involved in the linkage process and on how selection biases may affect findings. First, we compare mortality patterns from our samples to the mortality tables derived from the Canadian Human Mortality Database at the Université de Montréal. Second, using logistic regression, we build a sample selection model for which the outcome is whether a death record is found. Many of the predictors included in this model (sex, urban-rural status, father's occupation, literacy, etc.) will also be used in our future models predicting mortality at older ages, helping us to assess potential biases and their impacts on findings.

Conditions de vie durant l'enfance et longévité : évaluation d'une base de données créée à partir du recensement canadien de 1901 et de l'état civil québécois¹

LAURENCE PILON-MARIEN^{*}, ALAIN GAGNON^{},
BERTRAND DESJARDINS^{*} ET ROBERT BOURBEAU^{*}**

Plusieurs études montrent que les conditions de vie vécues dans l'enfance peuvent avoir un effet à long terme sur la santé et la mortalité. Cette niche n'ayant pas été explorée pour la population canadienne, nous avons mis sur pied un projet visant le jumelage du recensement canadien de 1901 et des actes de décès de l'état civil québécois dans le but d'établir l'âge au décès d'enfants pour lesquels il existait des informations socio-économiques. Le but de la présente contribution est de tester la validité des données collectées en les comparant avec des tables de mortalité calculées à l'aide des données issues de la Base de données sur la longévité canadienne (Université de Montréal), et d'examiner des sources de biais potentiels. Parmi les biais potentiels, nous avons évalué la possibilité d'un effet de sélection dans l'échantillon à l'aide d'une régression logistique dans laquelle nous avons introduit plusieurs variables explicatives de la mortalité aux grands âges. Le but recherché étant de savoir si ces variables ont à leur tour un impact sur le fait de trouver ou non un acte de décès.

English abstract, p. 191

On sait depuis longtemps que les conditions de vie dans l'enfance ont un effet sur la santé et la mortalité aux âges subséquents (Kermack *et al.*, 1934). Des études récentes ont montré dans ce sens que les enfants exposés à des conditions défavorables ont des niveaux de morbidité et de

1. Nous tenons à remercier le Conseil de recherches en sciences humaines du Canada (CRSHC) qui a financé ce projet et sans l'appui duquel il n'aurait pu voir le jour. Nous souhaitons également souligner la précieuse collaboration du projet *Balsac* et du *Canadian Families Project*, ainsi que Madame Lisa Dillon, professeure au département de démographie de l'Université de Montréal.

^{*} Département de démographie de l'Université de Montréal, Montréal, Québec, Canada.

^{**} Département de sociologie et département d'épidémiologie et de biostatistiques, University of Western Ontario, London, Ontario, Canada.

mortalité plus élevés que la moyenne à l'âge adulte (Elo et Preston, 1992 ; Galobardes *et al.*, 2004). Cette niche n'ayant pas été explorée pour la population canadienne et québécoise récente (voir Gagnon et Mazan, 2009 pour l'époque préindustrielle), nous avons mis sur pied un projet de recherche visant à observer l'effet des conditions de vie durant l'enfance sur la mortalité aux âges avancés au Québec.

Dans le cadre de ce projet, une base de données jumelant le recensement canadien de 1901 et les actes de décès de l'état civil québécois a été créée dans le but d'établir l'âge au décès d'enfants canadiens-français nés à la fin du 19^e siècle pour lesquels il existait des informations socio-économiques dans le recensement. Grâce aux variables liées à la composition du ménage, à la profession du chef de ménage et aux moyens de subsistance de la famille, à l'alphabétisation et à l'éducation de ses membres âgés de cinq ans ou plus, au lieu de résidence et à la résidence elle-même, ainsi qu'aux divers biens fonciers que possèdent les membres du ménage, nous avons été en mesure de conceptualiser le milieu de vie initial de notre cohorte d'étude, ce qui permettra à terme d'étudier l'impact de ce milieu sur la mortalité. Le projet étant encore jeune et la base de données en construction, peu d'études ont jusqu'à présent utilisé ces données et seuls quelques résultats préliminaires sont disponibles. Le but de cette contribution est donc méthodologique. L'objet est de tester la précision, la représentativité et la validité des données collectées en les comparant avec celles de la Base de données sur la longévité canadienne (BDLC)² et en examinant des sources de biais potentiels qui pourraient émaner de problèmes d'appariement et de sélection dans l'échantillon. Ce regard approfondi sur la qualité des données est une étape cruciale en raison de son impact sur la pertinence et la justesse des recherches qui en découleront.

COLLECTE DE DONNÉES

La conception de la collecte de données s'inspire des méthodes utilisées par Preston *et al.* (1998) dans leur étude sur les liens entre les conditions de vie dans l'enfance et la longévité. Cette étude de « cas-témoins » est basée sur des Africain-Américains décédés à plus de 85 ans entre le 1^{er} et le 14 janvier 1985, et qui ont pu être retracés par la suite dans le recensement américain de 1900 ou 1910. Toutefois, contrairement à cette dernière, nous avons

2. Département de démographie de l'Université de Montréal : www.bdlc.umontreal.ca.

préféré effectuer une étude de cohorte retraçant en aval la date de décès d'individus énumérés dans la province de Québec lors du recensement de 1901.

Le canevas à l'origine de la base de données est l'échantillon de 5 % du recensement canadien de 1901 qui a été dépouillé dans le cadre du *Canadian Families Project* (Sager et Baskerville, 2007). Il s'agit d'un échantillon aléatoire et nominatif des ménages recensés au Canada le 31 mars 1901. Plus précisément, on y retrouve 5 % des demeures (*dwellings*), sélectionnées au hasard sur chaque bobine de microfilm du recensement. Pour chacune des demeures retenues, les tableaux 1 et 2 du recensement ont été dépouillés, c'est-à-dire les informations relatives aux individus et aux ménages, ainsi que celles sur la propriété (Ornstein, 2000 ; Sager, 2001). De façon générale, l'utilisation de l'échantillon du *Canadian Families Project* a permis d'épargner sur les coûts d'un nouveau dépouillement et de fonder notre recherche sur une cohorte d'individus éteinte, ce qui est fondamental pour l'étude de la mortalité aux grands âges. Toutefois, la base de données n'est pas une copie exacte de l'échantillon de 5 %, puisqu'elle prend uniquement en compte des individus d'origine canadienne-française et catholique, présents au Québec lors du recensement de 1901.

C'est en raison du manque d'informations complémentaires essentielles à la recherche des actes de décès que seules les familles d'origine canadienne-française du Québec ont été sélectionnées. Fait à noter, l'origine des familles a été établie en fonction de la langue maternelle et de la religion déclarées par le chef de ménage lors du recensement. Parmi ces familles francophones de confession catholique ont été retenues celles incluant au moins un enfant âgé de 0 à 15 ans (né entre 1885 et 1901) et ayant une relation de filiation avec le chef de ménage. Ces deux derniers critères répondent au principal objectif de la collecte qui est de fournir des données pour étudier l'effet des conditions de vie dans l'enfance sur la longévité. La sélection des individus étant basée sur le ménage, l'échantillon exclut les enfants présents dans une institution au moment du recensement (orphelinat, etc.).

Une recherche systématique a été effectuée pour trouver les dates de décès des individus de l'échantillon afin d'obtenir leur âge au décès. Celles-ci ont été retracées à l'aide de l'index des décès de 1926 à 1996 de l'Institut de la statistique du Québec (ISQ) et de la Société de généalogie du Québec, ainsi que d'un fichier spécialement conçu par l'ISQ rapportant les décès des personnes âgées de 85 ans et plus entre 1997 à 2004. Il est important de mentionner ici que les actes de décès nominatifs concernent uniquement les décès ayant eu lieu sur le territoire québécois. Cela implique que la

mortalité considérée dans l'échantillon concerne des individus d'origine canadienne-française qui étaient non seulement domiciliés au Québec en 1901, mais qui sont également décédés dans cette province après 1926.

Les décès ont été retracés et appariés manuellement en fonction d'informations essentielles : nom(s), prénom(s) et date de naissance, tels qu'indiqués dans le recensement, ainsi que les noms et prénoms des parents et conjoint(s). Cependant, comme certaines de ces indications faisaient défaut, les actes paroissiaux des mariages catholiques au Québec entre 1800 et 1940 provenant du *Projet BALSAC*³ ont été consultés dans le but d'obtenir le nom des conjoints éventuels, de même que le nom de jeune fille de la mère des sujets. Ce sont aussi pour ces raisons que la base de données se limite aux ménages québécois dont l'un des parents est d'origine canadienne-française et catholique. En plus d'être d'une grande qualité et exhaustivité, de telles données n'étaient pas facilement disponibles pour l'ensemble du Canada et pour les autres religions avant la mise en place de l'état civil canadien en 1921.

Dans ces actes, ont également été saisis la date et le lieu du mariage des parents ainsi que ceux des individus mariés, en plus des noms et prénoms des parents des époux. Grâce à ces informations, les actes de décès de certains conjoints ont pu être trouvés et pourront servir de variables de contrôle dans de futures recherches. Par exemple, la variable de l'âge au décès du conjoint pourrait contrôler indirectement l'effet des conditions de vie après l'enfance, car ayant vécu en cohabitation et dans un environnement socio-économique similaire durant la majeure partie de leur vie adulte, on peut considérer que les conjoints bénéficient mutuellement de la longévité de l'autre (Mazan et Gagnon, 2007 ; Blackburn *et al.*, 2004). De plus, les professions des parents et la profession de l'époux au moment du mariage ont également été relevées, ce qui peut permettre de vérifier dans quelle mesure le statut socioprofessionnel à l'âge adulte et une éventuelle mobilité sociale peuvent confirmer ou renverser les tendances qui se mettent en place aux plus jeunes âges.

ÉTAT DE LA COLLECTE

Notre base de données, en construction depuis mai 2007, est aujourd'hui complétée à plus de 60 %. Afin de pouvoir utiliser les données avant que la collecte ne soit terminée, l'échantillon complet a été subdivisé en cinq

3. <http://www.uqac.ca/balsac>.

sous-échantillons aléatoires représentant chacun 20 % des ménages initiaux. Jusqu'à présent, les données de 8 206 individus nés entre 1885 et mars 1901 ont été traitées sur une possibilité de 14 063. Pour 52 % d'entre eux, un acte de décès a été trouvé et parmi ces décès, 90 % sont survenus au-delà de quarante ans. À la lueur de ces résultats, nous estimons que la base de données contiendra les informations d'au moins 6 500 individus ayant survécu à leur quarantième anniversaire une fois l'échantillon traité dans son intégralité.

TABEAU 1 État de la collecte des décès

Année de naissance	Échantillon complet	Échantillon 60%	Décès collectés	% décès collectés	% décès après 40 ans
1885-1889	1491	824	403	49%	93%
1890-1895	5162	3040	1563	52%	92%
1896-1901	7410	4342	2255	52%	88%
Total	14063	8206	4221	52%	90%

DÉCÈS PAR GROUPE D'ÂGE						
Année de naissance	40-49	50-59	60-69	70-79	80-89	90 et plus
1885-1889	30	43	65	104	94	37
1890-1895	88	159	270	393	384	143
1896-1901	111	182	387	564	498	237
Total	229	384	722	1061	976	417

Sources : Base de données sur les conditions de vie durant l'enfance et la longévité et Base de données sur la longévité canadienne (BDLC).

Étant donné que l'index des décès ne débute qu'en 1926, année de la création de l'état civil québécois, et qu'il contient nettement moins d'éléments d'information dans ses débuts, les efforts de collecte se sont surtout concentrés sur les décès au-delà de quarante ans. Dans le cas présent, cette limite n'en est pas vraiment une puisque les données recueillies serviront à étudier la survie aux grands âges ainsi que certains facteurs de la longévité. Prendre en considération les individus décédés à plus de quarante ans permet en partie d'exclure les personnes décédées de causes non reliées à la longévité, soit la mort accidentelle ainsi que la mortalité infanto-juvénile et maternelle (Mazan et Gagnon, 2007 ; Blackburn *et al.*, 2004).

En somme, cela signifie que les décès après quarante ans de 46 % des individus de notre échantillon de départ ont été retrouvés au Québec. Pour les autres 54 %, nous savons que 10 % d'entre eux sont décédés avant cet

âge et après 1926, soit 5 % des 8 157 individus. Concernant les individus restants, ils peuvent aussi bien être morts avant cet âge et avant 1926 ou bien à l'extérieur du Québec, puisqu'il s'agit ici d'une cohorte éteinte et semi-fermée, donc soumise à l'émigration. Cependant, il ne faut pas oublier la possibilité que certains actes de décès n'aient pas été retrouvés suite à diverses difficultés dans la collecte de données. Par exemple, la date de naissance inscrite au moment du recensement ou dans l'acte de décès est parfois inexacte ou incomplète, ce qui rend le jumelage difficile. Il est également difficile de trouver ou de confirmer la date de décès pour les individus dont le prénom (Joseph, Marie, etc.), ou le nom (Tremblay, Gagné, etc.) est très répandu vu les nombreuses possibilités d'appariement. Il faut ajouter à ces embûches la possibilité de prénoms multiples pour une même personne, les actes de décès sans prénom, de même que la non-standardisation des noms et des prénoms dans l'index des décès (Aurore/Orore, Omer/Homer, Malette/Malette/Malet, Provost/Prévost, etc.). Par ailleurs, il est aussi possible que d'autres facteurs (qui ne sont pas apparus lors de la collecte des données proprement dite) aient un effet sur la probabilité du jumelage, telle la connaissance de la date de naissance exacte des enfants par le chef du ménage. Afin de limiter la portée de ces problèmes, chacun des ménages a été recherché dans le recensement de 1911 par le biais du site Internet *Automated Genealogy*⁴ (un projet d'indexation des recensements canadiens de 1901 et 1911). Pour les ménages retrouvés en 1911, les noms de famille, les prénoms et les dates de naissance ont été colligés tels qu'ils y apparaissaient pour chacun des individus qui étaient présents en 1901. Cette étape nous a permis de considérer les diverses variations des informations personnelles entre 1901 et 1911 afin d'effectuer des jumelages plus éclairés.

L'examen de la répartition des 54 % de décès non retrouvés est essentiel à l'évaluation de la qualité des données et à l'identification des biais éventuels. Par contre, avant de s'y attarder plus en détail, il est primordial de s'assurer de la représentativité des décès retrouvés, c'est-à-dire que la mortalité par âge après quarante ans de notre cohorte représente bien celle expérimentée au cours du 20^e siècle par les générations correspondantes.

4. <http://automatedgenealogy.com/>.

MORTALITÉ DES INDIVIDUS DE L'ÉCHANTILLON VERSUS MORTALITÉ DE LA GÉNÉRATION

Tout d'abord, les niveaux de mortalité après quarante ans selon les groupes d'âge quinquennaux et le sexe ont été obtenus par la construction de tables de mortalité pour la cohorte d'étude (l'échantillon) et pour la génération s'apparentant le plus possible à celle-ci. Dans le cas présent, la génération 1895 semblait le choix le plus juste puisque la moyenne des dates de naissance de l'ensemble des individus de l'échantillon correspond au 15 juin 1895.

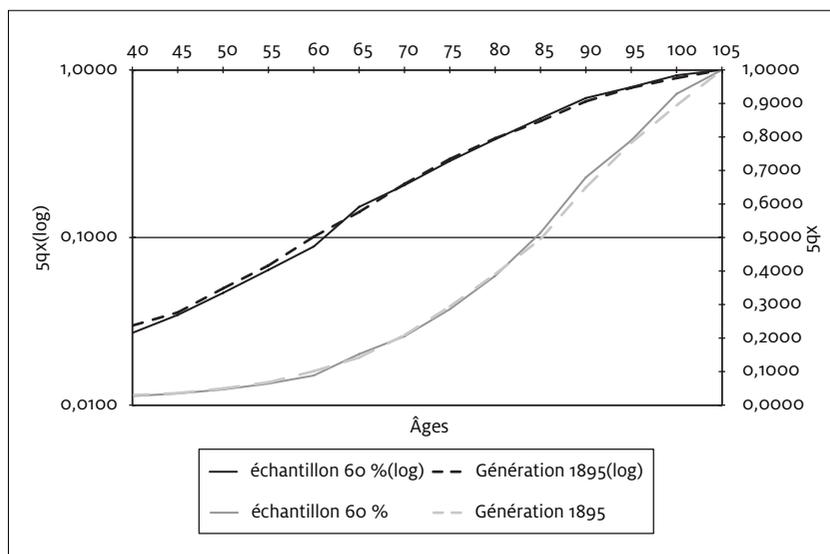
Les tables de mortalité pour l'échantillon ont été calculées en fonction des dates de décès inscrites dans l'index et des dates de naissance, telles que déclarées dans le recensement. Cependant, parmi les 3 799 individus morts après quarante ans, trente-huit âges au décès ont été estimés, car l'information concernant le jour ou le mois de naissance ou de décès était manquante. Les mois et les jours manquants ont donc été remplacés en fonction de la moyenne de mois dans une année et de jour dans un mois, soit par juin et le quinzième jour du mois. Il faut également considérer que les âges aux décès ne sont pas tous exacts au jour près étant donné que les dates de naissance, telles que déclarées par le chef du ménage lors du recensement, ne sont pas nécessairement justes. Néanmoins, si l'on se fie à Gauvreau *et al.* (2000), cette situation ne serait pas alarmante puisqu'il y aurait un manque réel de cohérence entre la date de naissance issue des registres paroissiaux et celle du recensement de 1901 pour seulement 2,2 % à 3,3 % des enfants âgés de 0 à 5 ans.

Pour ce qui est des tables de mortalité de la génération de 1895 décédée au Québec, elles ont été réalisées à l'aide des décès recueillis, pour chaque année depuis 1935, en fonction de chacun des groupes d'âge correspondant à la génération de 1895. Afin d'éviter tout biais lié à la transformation de données transversales en données longitudinales, nous avons utilisé les décès répartis par triangle de Lexis (Pressat, 1983). Le calcul de ces tables a été rendu possible grâce aux données de la BDLC.

Dans les deux cas, nous avons eu recours à la *méthode des générations éteintes* pour le calcul des tables de mortalité après quarante ans (Vincent, 1951), le nombre de survivants au départ, soit à quarante ans, correspondant à la totalité des décès survenus après cet âge. Par la suite, les quotients de mortalité ont été calculés avec les décès par groupe d'âge quinquennal en faisant abstraction des mouvements migratoires.

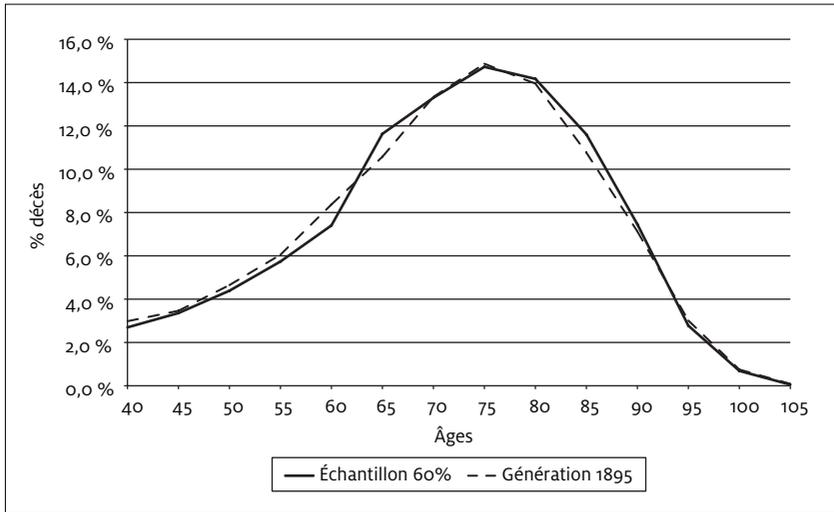
À des fins de comparaisons, nous avons retenu les quotients de mortalité (figures 1 et 3) et la répartition des décès (figure 2) pour chaque groupe d'âge quinquennal.

FIGURE 1 Quotients de mortalité de l'échantillon et de la génération 1895



Sources : Base de données sur les conditions de vie durant l'enfance et la longévité et Base de données sur la longévité canadienne (BDLC).

Dans l'ensemble, la correspondance est remarquable : les quotients de mortalité par groupes d'âge quinquennaux de l'échantillon suivent la même tendance que ceux observés pour la génération née en 1895, soit une croissance exponentielle entre 40 et 105 ans et plus. La probabilité de décéder d'environ 3 % à 9 % entre 40 ans et 60 ans s'accélère par la suite. Toutefois, il y a un léger affaissement de l'accroissement exponentiel vers 90 ans, ce qui est surtout visible à l'échelle logarithmique. Cet affaissement a été remarqué ailleurs et semble caractériser la plupart des populations (Thatcher *et al.*, 1998). C'est d'ailleurs, entre 65 et 85 ans que l'on observe la majorité des décès après quarante ans pour la cohorte issue de l'échantillon de 60 % et pour la génération de 1895, plus particulièrement aux environs du mode (à 75 ans) où surviennent environ 15 % des décès (figure 2).

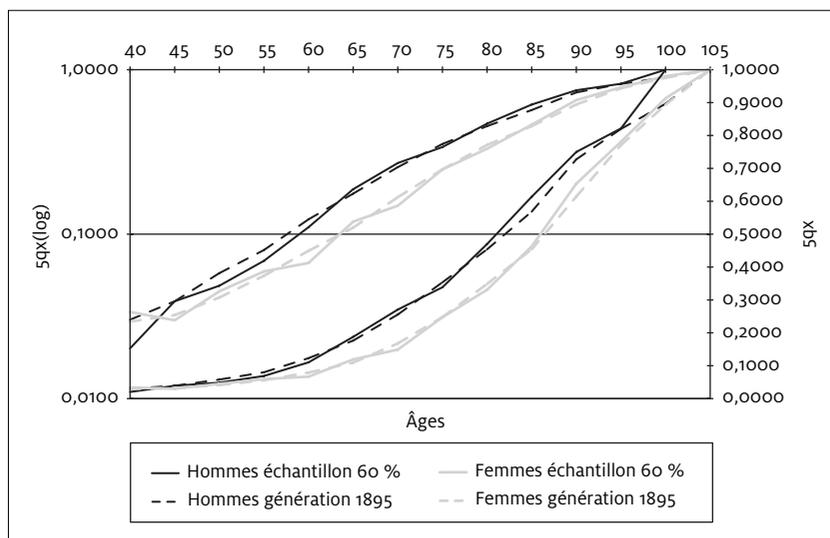
FIGURE 2 Répartition des décès après 40 ans selon le groupe d'âge

Sources : Base de données sur les conditions de vie durant l'enfance et la longévité et Base de données sur la longévité canadienne (BDLC).

Cependant, il faut être prudent dans la comparaison entre les deux tables de mortalité. La table de mortalité de référence n'est pas totalement représentative de celle de l'échantillon, car elles décrivent la mortalité concernant des populations quelque peu différentes : les niveaux de mortalité calculés pour la génération 1895 se rapportent à des individus pouvant être aussi bien nés au Québec qu'ailleurs, alors que notre échantillon se réfère exclusivement aux Canadiens français. Le léger accroissement de mortalité à 85 ans des individus de l'échantillon pourrait bien n'être qu'artificiel, étant donné la plus faible mortalité des immigrants par rapport aux natifs, même s'il y a convergence entre les deux groupes aux grands âges (Bourbeau, 2002). En somme, les légères différences entre les quotients des deux tables de mortalité ne renvoient pas nécessairement à un problème de représentativité.

Si l'on considère les sexes de manière indépendante, la mortalité des hommes et des femmes suit les mêmes tendances que celles exposées précédemment. Toutefois, on remarque clairement la mortalité différentielle selon le sexe puisque les femmes ont une mortalité beaucoup plus faible que leurs homologues masculins entre 55 et 95 ans (figure 3). Une fois de plus, les quotients de mortalité aux âges quinquennaux de l'échantillon de 60 % sont très similaires à ceux observés pour la génération de 1895 pour les deux sexes, ce qui est très encourageant.

FIGURE 3 Quotients de mortalité de l'échantillon et de la génération 1895, selon le sexe



Sources : Base de données sur les conditions de vie durant l'enfance et la longévité et Base de données sur la longévité canadienne (BDLC).

VALEUR MANQUANTE : IMPACT DE LA MORTALITÉ AVANT QUARANTE ANS ET DE L'ÉMIGRATION SUR LA FRÉQUENCE DES APPARIEMENTS

Pour 54 % des individus de la base de données, aucun décès après quarante ans n'a été retrouvé. Trois facteurs sont en cause. Les deux premiers sont des phénomènes démographiques : la mortalité et l'émigration. Même s'ils sont impossibles à contrôler, nous pouvons cependant mesurer leurs effets respectifs. Le troisième facteur, quant à lui, est lié aux difficultés de la collecte de données ; il concerne les cas de personnes décédées au Québec, dont la trace de décès n'a pas été retrouvée pour diverses raisons. L'effet de ce dernier facteur peut simplement être évalué par une méthode résiduelle, c'est-à-dire en soustrayant le nombre approximatif d'individus « disparus » en raison d'une migration ou d'un décès avant quarante ans de l'ensemble des cas manquants.

Dans un premier temps, la mortalité avant l'âge de quarante ans de la génération de 1895 a été estimée à l'aide des tables de mortalité par génération du Québec de 1891 et 1901. Les registres paroissiaux du Québec pour le 19^e siècle n'étant pas encore dépouillés et l'état civil québécois ne commen-

çant qu'en 1926, le calcul direct de la mortalité de la génération 1895 avant quarante ans est pour l'instant impossible. Nous avons donc utilisé les tables par générations de Bourbeau-Légaré qui ont été estimées à l'aide de la méthode des écarts et de tables-types de mortalité (Bourbeau *et al.*, 1997). La table de mortalité de la génération de 1895 n'existant pas, nous avons estimé celle de 1896 par interpolation linéaire grâce aux tables de 1891 et 1901.

Selon cette dernière, la probabilité de décéder avant quarante ans de cette génération serait de 35 %, soit 35,6 % pour les hommes et 34,5 % pour les femmes. Il est possible d'évaluer grossièrement que pour environ 35 % des individus de l'échantillon, la date de décès n'a pu être retrouvée en raison d'un décès prématuré. Si l'on soustrait ce nombre du 54 % initial, il reste un écart de 20 % que nous pouvons interpréter par le biais de l'émigration et des problèmes de jumelage.

L'estimation de l'émigration est plus ardue que celle de la mortalité et s'avère donc assez approximative. Elle est en effet difficile à estimer à la base avec les données canadiennes, d'autant plus qu'elle doit l'être à la fois dans une perspective longitudinale et historique. Nous avons tenté de faire cette évaluation en recourant aux effectifs de population par âge et sexe correspondant à la cohorte âgée de 0 à 4 ans au recensement de 1901 dans les recensements canadiens postérieur à 1901. Tout en considérant le lieu de naissance et lieu de résidence mentionnés dans les recensements, ces effectifs de population ont été comparés à la population théorique attendue à chaque recensement. Par population théorique attendue, nous faisons référence à la population qu'il devrait y avoir s'il s'agissait d'une population fermée, c'est-à-dire une population uniquement soumise à la mortalité. En appliquant les principes des méthodes résiduelles, nous avons cherché à estimer l'émigration interprovinciale et internationale nette, comprenant les émigrants de retour, en prenant en compte les stocks d'émigrants à chaque recensement ainsi que les flux d'émigrants entre ceux-ci (voir résultats en annexe) dans le but d'obtenir le nombre total d'émigrants de la génération.

Cependant, nous n'avons pu mener nos démarches à terme. Au moment d'estimer l'émigration, nous avons remarqué que la population théorique aux différents âges était moins importante que la population canadienne née au Québec selon les recensements. Ce résultat est d'ailleurs très surprenant, car il impliquerait une émigration négative dans un contexte où notre population d'intérêt n'est pas soumise à l'immigration étant donné qu'elle est née au Québec. Selon nous, cette situation pourrait être le fruit d'une sous-estimation de la population théorique due à une

surestimation de la mortalité en bas âge. Du reste, les quotients de mortalité prospectifs issus des tables de mortalité Bourbeau-Légaré ne se prêtaient pas parfaitement à l'exercice, puisque les auteurs considèrent les mouvements migratoires.

En somme, nous n'avons retrouvé aucun décès après quarante ans pour 54 % des individus de la base de données ; cette proportion est la même aussi bien pour les hommes que pour les femmes. Pour la génération de 1895, la probabilité de survie jusqu'à l'âge de quarante ans est de 64,9 %. Cela implique qu'environ 35 % des individus nés à cette période n'ont pas atteint cet âge et qu'en conséquence leur acte de décès n'était pas disponible. Pour des raisons liées à la collecte de données et aux décès ayant eu lieu à l'extérieur du Québec, 20 % de l'échantillon serait incomplet. N'étant pas en mesure d'évaluer la part de l'émigration dans le non-appariement des individus avec les actes de décès, il nous a donc été impossible de mesurer l'ampleur des problèmes reliés à la collecte de données. Cependant, il y a fort à supposer que la part de ce problème est inférieure à 20 %.

EFFET DE SÉLECTION DANS L'ÉCHANTILLON ET PRÉSENTATION DES VARIABLES RELIÉES AUX CONDITIONS DE VIE DANS L'ENFANCE

Nous avons établi que les décès des individus qui n'ont pas quitté le Québec et qui ont atteint quarante ans représentent bien la mortalité de la génération, même si ce ne sont pas tous les décès qui ont été identifiés ; cette « validité » de la base de données signifie en principe que celle-ci peut être utilisée pour l'étude de la survie et de la longévité.

Toutefois, il n'est pas exclu qu'il puisse exister un problème de validité « interne », causé par la fréquence d'appariement. Autrement dit, nous n'écartons pas la possibilité d'un effet de sélection au sein de l'échantillon, car, pour 54 % des individus, la variable d'intérêt (l'âge au décès) est manquante. Généralement, les modèles utilisés pour l'analyse de survie tiennent compte de la censure à droite en supposant qu'il y a indépendance entre la sortie de l'observation et le phénomène à l'étude. Dans notre cas, il était essentiel de nous assurer que la probabilité de trouver un acte de décès et la longévité ne soit pas expliquée par les mêmes facteurs. Dans le cas contraire, il pourrait y avoir une erreur de spécification et les estimateurs du modèle explicatif de la longévité s'en trouveraient biaisés. Cette erreur de spécification aurait un effet similaire à celui d'une variable indépendante omise ayant une relation avec la variable dépendante, ainsi

qu'avec une ou plusieurs des variables indépendantes sélectionnées (Heckman, 1979 ; Berk, 1983). La sélection pourrait donc masquer l'effet réel de certains facteurs reliés à l'enfance sur la longévité.

Il n'est pas question de corriger ici l'effet de sélection, puisque nous n'en sommes pas encore à une étude empirique sur l'effet des conditions de vie dans l'enfance sur la survie. L'intérêt de cette discussion est plutôt de soulever la question de la sélection dans l'échantillon ainsi que de vérifier et d'illustrer les sources de biais potentiels dans l'appariement des individus avec leur acte de décès. Vu les implications méthodologiques, cette étape est primordiale pour les recherches qui s'appuieront sur cette base de données, à savoir s'il est nécessaire d'ajuster les résultats en fonction d'un effet de sélection.

Jumelage des actes de décès et conditions socio-économiques dans l'enfance

En ce qui concerne la possibilité d'un effet de sélection dans l'échantillon, une régression logistique a été effectuée en prenant comme variable dépendante l'appariement ou non d'un acte de décès et une série de variables explicatives reliées aux caractéristiques socio-économiques dans l'enfance. En somme, cette étape consiste à vérifier que le jumelage des actes de décès n'est pas dû à des caractéristiques aussi reliées à la mortalité aux âges avancés.

Parmi les nombreuses possibilités offertes dans le recensement de 1901, nous avons choisi les variables qui se sont avérées jouer un rôle significatif sur la longévité dans des études antérieures. Ainsi, dans une revue de la littérature sur le sujet, Galobardes *et al.* (2004) ont comparé vingt-deux études parmi lesquelles dix-huit rendent compte d'une mortalité plus élevée pour les individus ayant vécu dans des conditions socio-économiques défavorables (logement mal aéré ou surpeuplé, absence d'eau courante, père au chômage ou détenant un emploi manuel, etc.). L'étude déjà citée de Preston *et al.* (1998) révèle que vivre son enfance en milieu rural, sur une ferme, dans un ménage biparental où le chef est propriétaire du logement et avoir un père et une mère alphabétisés augmentent les probabilités de survivre jusqu'à des âges avancés. Par exemple, les individus ayant grandi sur une ferme ont 46 % plus de chance que les autres de vivre jusqu'à 85 ans. Les auteurs constatent également que vivre son enfance dans un ménage où le chef est analphabète réduit de 27 % la probabilité de vivre jusqu'à cet âge, alors que la scolarisation de la mère augmente cette probabilité de 36.

Au total, notre modèle est composé de huit variables explicatives, dont cinq reflètent les conditions socio-économiques des ménages en 1901 (tableau 2) : emploi, résidence (urbaine ou rurale), type de ménage (monoparental ou biparental), statut (propriétaire ou locataire) et alphabétisation. Parmi ces cinq variables, deux ne sont pas issues directement du recensement. Ainsi, les variables concernant l'alphabétisation du chef de ménage et le milieu de résidence ont été créées à partir des informations disponibles. Un chef de ménage est considéré alphabétisé s'il sait au minimum lire ou écrire, tandis que dans le cas contraire, il est qualifié d'analphabète. Pour ce qui est du milieu de résidence, les lieux habités par plus de mille personnes ont été considérés comme urbains, sous ce seuil, il s'agit de milieux ruraux (Sager et Baskerville, 2007).

Trois autres variables ne touchant pas au domaine socio-économique ont aussi été incluses dans le modèle, soit le sexe, la fréquence du patronyme et la région de résidence. Après avoir constaté que les appariements étaient plus difficiles pour certains noms de famille et dans certaines régions du Québec, nous avons inclus la fréquence du patronyme et la région de résidence. Nous avons donc regroupé les soixante-cinq divisions de recensement du Québec au sein de cinq régions, soit la région de Montréal, la région de Québec, le Centre du Québec, les régions situées près des frontières ontariennes et états-uniennes, ainsi que les régions éloignées (districts ne se situant pas sur l'axe du fleuve Saint-Laurent, ni à proximité des frontières considérées). Pour ce qui est des noms de famille, la difficulté d'appariement des actes de décès touchait surtout les individus ayant un nom de famille très répandu. Nous avons alors créé une variable considérant la fréquence des patronymes dans les actes des mariages québécois qui ont eu lieu entre 1880 et 1940, et avons par la suite distingué les individus ayant l'un des dix noms les plus fréquents⁵ au Québec des autres.

Le tableau 2 montre les rapports de cote (« *odds ratios* » en anglais ou OR) pour chaque variable retenue, ainsi que les pourcentages d'individus dans chaque catégorie pour les variables catégorielles. On peut constater qu'il y a un peu plus de chances de retrouver un acte de décès après quarante ans pour un individu de sexe féminin que de sexe masculin. Le rapport de cote est de 1,0990, ce qui signifie que la « cote » pour retrouver un décès est environ 10 % plus élevée chez les femmes que chez les hommes. Ce résultat est peu étonnant, puisque nous aurions cru que la

5. Les dix noms les plus fréquents au Québec entre 1880-1940 : Bélanger, Bouchard, Côté, Gagné, Gagnon, Gauthier, Morin, Pelletier, Roy et Tremblay.

TABLEAU 2 Régression logistique : jumelage des actes de décès et conditions socio-économiques dans l'enfance

VARIABLES	RAPPORT DE COTE (Odds ratio)	% CATÉGORIES (N = 8207)		
		Décès	Sans décès	Total
Sexe	1,0990 *	52 %	48 %	4186
Référence: hommes	–	51 %	49 %	4019
Statut urbain-rural	1,0970	54 %	46 %	2504
Référence: rural	–	50 %	50 %	5701
Chef du ménage locataire	0,9348 **	52 %	48 %	2539
Référence: propriétaire	–	51 %	49%	5666
Famille monoparentale	0,5307 **	37 %	63%	277
Référence: famille biparentale	–	52 %	48%	7928
Alphabétisation du chef de ménage				
Analphabète	0,8102 *	45 %	55 %	1110
Inconnu	0,8589	44 %	56 %	136
Référence: alphabétisé		53 %	47 %	6959
Fréquence du patronyme	1,6568 ***	9 %	91 %	3984
Référence: 10 noms les plus fréquents	–	5 %	95 %	4221
Emploi du chef de ménage				
Manufacturier/marchand/professionnel/col blanc/négociant	0,9161	56 %	44 %	1212
Artisan qualifié ou semi-qualifié	0,8421	53 %	47 %	1748
Journalier	0,6023 ***	41 %	59 %	1368
Inconnu	0,9862	53 %	47 %	326
Référence: fermier	–	53 %	47 %	3551
Région de résidence				
Région de Québec	0,7427 *	51 %	49 %	1027
Centre du Québec	0,8475	55 %	45 %	1502
Régions éloignées	0,6103 ***	46 %	54 %	1144
Régions frontières Ontario et États-Unis	0,5207 ***	42 %	58 %	1836
Référence: régions Montréal	–	59 %	41 %	2696

Prob > ch₂ = 0,0000

Erreurs standards ajustées selon le ménage d'appartenance (2536 grappes)

p<0,05 *

p<0,01 **

p<0,001 ***

Source : Base de données sur les conditions de vie durant l'enfance et la longévité (BDLC).

présence d'informations complémentaires plus fréquentes dans les actes de décès des femmes, tel le nom du conjoint, permet un meilleur couplage.

Par ailleurs, la présence des deux parents dans le ménage est très importante pour le jumelage de l'acte de décès après quarante ans et renvoie probablement à une plus grande mortalité (avant 40 ans) chez ceux qui ont perdu un parent dans leur enfance ou à une plus forte probabilité d'émigrer ($OR = 0,5307$). Ces individus sont toutefois rares dans l'ensemble de l'échantillon (2,5 %). De la même manière, on peut penser que l'analphabétisme du chef du ménage est lié à une plus forte mortalité et donc à une plus faible probabilité de retrouver un acte de décès après quarante ans pour les individus concernés ($OR = 0,8102$). Le même constat s'applique pour les enfants de journaliers pour lesquels la mortalité « précoce » (c.-à-d. avant 40 ans) peut également être en cause pour le plus faible jumelage ($OR = 0,6023$). Les autres statuts socioprofessionnels des chefs de ménage semblent quant à eux sans conséquence pour le couplage. Bien que la mortalité différentielle puisse être mise en cause, il se peut également que les familles les plus désavantagées socio-économiquement (familles monoparentales, chef journalier, chef analphabète, etc.) soient plus promptes à émigrer à l'extérieur du Québec, notamment en Ontario, dans l'ouest canadien ou aux États-Unis, afin d'améliorer leur sort, ce qui représenterait une source additionnelle d'attrition dans notre échantillon.

Nous sommes également un peu moins susceptibles de retrouver les individus ayant vécu leur enfance (ou une partie de leur enfance) dans un ménage où le chef était locataire. Le rapport de cote ne dénote pas un effet très important (0,9348, ou une réduction d'environ 6 % de la cote) mais souligne peut-être l'importance de l'enracinement. En effet, une famille dont, à l'inverse, le chef de ménage est propriétaire a probablement moins de chance de migrer vers d'autres régions du Québec ou ailleurs, ce qui augmente la probabilité de retrouver ses membres à des âges subséquents. Dans le même ordre d'idée, la résidence près d'une frontière ou dans une région éloignée est également associée à une plus grande probabilité d'émigrer, c'est pourquoi on a moins de chance de retrouver les actes de décès des représentants de ces régions ($OR = 0,5207$). Enfin, comme nous nous y attendions, une fréquence patronymique moins élevée est associée à une plus grande probabilité de couplage ($OR = 1,6568$).

DISCUSSION

Ce compte-rendu de l'avancement de nos travaux sur la reconstitution des conditions de vie dans l'enfance et leur impact sur la mortalité subséquente révèle que les méthodes d'échantillonnage et de collecte de données mettent finalement à notre disposition un échantillon représentatif de la mortalité de notre cohorte. Comme la comparaison avec les quotients de mortalité de la génération de 1895 l'a démontré, il ne semble pas y avoir de biais dans l'estimation de la mortalité après quarante ans. Même s'il y a quelques variations avec la courbe de référence, il y a peu de chances qu'elles soient le résultat de biais systématiques. Notre échantillon réfère en outre à une population quelque peu différente de celle représentée dans les données de la BDLC.

Cependant, la régression logistique explorant les causes de non-appariement laisse présager la présence d'effets de sélection dans nos analyses futures, puisque la « réussite » du jumelage est due en partie à des variables ayant également un effet sur la mortalité ou la longévité. Cette situation n'est pas dramatique dans la mesure où il est possible de contrôler cet effet de sélection au moyen de la méthode en deux étapes de Heckman (*Heckman two-stage modelling strategy*). Brièvement, cette méthode consiste à modéliser le risque de ne pas être sélectionné dans l'analyse (ne pas avoir été jumelé à un acte de décès) au moyen d'une régression probit (Heckman, 1979). On appelle ce risque le « ratio inverse de Mills » (*Inverse Mills' Ratio*), ratio qui est ensuite utilisé comme variable indépendante dans le modèle explicatif d'intérêt (ici la mortalité après 40 ans).

Nous pensons qu'une fois entièrement complétée cette base de données sera d'une qualité appréciable. À court terme, elle permettra l'étude générale de l'impact des conditions de vie dans l'enfance sur la longévité au Québec. Ainsi, il sera possible de confirmer ou d'infirmer les résultats obtenus par des études semblables portant sur d'autres régions du monde. Parallèlement, cet échantillon servira de population de référence à une étude similaire sur la longévité des frères et sœurs de centenaires québécois. Ce projet, actuellement en cours, a pour objectif de retrouver les familles de ces centenaires dans les recensements canadiens de 1901 et 1911 afin d'identifier leur fratrie et de retracer leurs décès dans l'état civil québécois.

D'autre part, au fur et à mesure que les actes paroissiaux du 19^e et du 20^e siècles seront dépouillés, nous pourrions maximiser le jumelage des individus aux actes de décès en concentrant les efforts sur la collecte des décès avant 1926. La base de données sera donc plus complète grâce à un meilleur appariement. L'accès à ces actes de décès permettra aussi de retrouver les décès des parents et d'estimer la mortalité infanto-juvénile dans les ménages, ce qui pourra servir de variable « *proxy* » représentant l'exposition à des maladies infectieuses dans l'enfance.

Finalement, dans les prochaines années, voire même dans les prochains mois, notre base de données évoluera selon la disponibilité de nouvelles données et les intérêts des chercheurs qui l'utiliseront. D'autres variables pourraient être créées grâce à une association avec d'autres équipes de recherche sur les populations historiques du Québec et du Canada. Pour ce faire, nous n'écarterons pas la possibilité d'utiliser les bases de données rendues disponibles comme celle du PHSVQ de l'Université Laval⁶, de Balsac, déjà cité, et du projet MAP (*Montréal, l'avenir du passé : MAP*).

BIBLIOGRAPHIE

- BASE DE DONNÉES SUR LA LONGÉVITÉ CANADIENNE, 2008. *Décès du Québec répartis par triangles de Lexis*. Université de Montréal, février 2008. http://www.prdh.umontreal.ca/BDLC/data/que/Deaths_1XI.txt.
- BLACKBURN, M-E., R. BOURBEAU et B. DESJARDINS. 2004. « Hérité et longévité au Québec ancien », *Cahiers québécois de démographie*, 33, 1 : 9-28.
- BERK, R. 1983. « An introduction sample selection bias in sociological data », *American Sociological Review*, 48, 3 : 386-398.
- BOURBEAU, R. 2002. « L'effet de la sélection d'immigrants en bonne santé sur la mortalité canadienne aux grands âges », *Cahiers québécois de démographie*, 31, 2 : 249-274.
- BOURBEAU, R., J. LÉGARÉ et J. ÉMOND V. 1997. « Nouvelles tables de mortalité par génération au Canada et au Québec, 1801-1991 ». Ottawa, Statistique Canada, Division de la démographie, Document démographique n° 3, septembre 2007, 91 p.
- ELO, I., et S. PRESTON. 1992. « Effects of early-life conditions on adult mortality : A review », *Population Index*, 58, 2 : 186-212.

6. Population et histoire sociale de la ville de Québec (PHSVQ) : <http://www.phsvq.cieq.ulaval.ca/>.

- GAGNON, A., et R. MAZAN. 2009. « Does exposure to infectious diseases in infancy affect old age mortality ? Evidence from a pre-industrial population », *Social Science & Medecine*, 68 : 1609-1616.
- GAGNON, A., K. SMITH, R. TREMBLAY, M. VÉZINA, H. PARÉ, P-P. et B. DESJARDINS. 2009. « Is there a trade-off between fertility and longevity ? A comparative study of women from three large historical databases accounting for mortality selection », *American Journal of Human Biology* (sous-presse).
- GALOBARDES, B., J. W. LYNCH et G. D. SMITH. 2004. « Childhood socioeconomic circumstances and cause-specific mortality in Adulthood : Systematic review and interpretation », *Epidemiologic Reviews*, 26 : 7-21.
- GAUVREAU, D., P. GOSSAGE et L. GINGRAS. 2000. « Measuring Fertility with the 1901 Canadian Census : A Critical Assessment », *Historical Methods*, 33, 4 : 219-228.
- GILLILAND, J., et S. OLSON. 2003. *Montréal, l'avenir du passé. Le Plan géomatique du gouvernement du Québec*. Janvier-février 2003 : 5-7. http://geography.uwo.ca/faculty/gilliland/MAP_HGIS.pdf.
- HECKMAN, J. J. 1979. « Sample selection bias as a specification error », *Econometrica*, 47, 1 : 153-161.
- KERMACK, W. G., A. G. MCKENDRICK et P. L. MCKINLAY. 1934. « Death rates in Great Britain and Sweden : Some regularities and their Significance », *The Lancet*, 31 mars 1934 : 698-703.
- MAZAN, R., et A. GAGNON. 2007. « Influence des facteurs familiaux et environnementaux sur la longévité au Québec ancien », *Population*, 62, 2 : 315-338.
- ORNSTEIN, M. 2000. « Analysis of Household Samples : The 1901 Census of Canada ». *Historical Methods*, 33, 4 : 195-198.
- PRESSAT, R. 1983. *L'analyse démographique : Concepts-méthodes-résultats*. Paris, Presses Universitaires de France, 295 p.
- PRESTON, S. H., M. HILL et G. E. DREVENSTEDT. 1998. « Childhood conditions that predict survival to advanced ages among African-Americans », *Social Science & Medecine*, 47, 9 : 1231-1246.
- SAGER, E. W., et P. BASKERVILLE. 2007. *Household counts : Canadian households and families in 1901*. Toronto, University of Toronto Press, 486 p.
- SAGER, E. W. 2001. « Introduction : The Canadian families projects », *Journal of Family History*, 26, 2 : 157-161.
- THATCHER, A. R., V. KANNISTO et J. W. VAUPEL. 1999. *The force of mortality at ages 80 to 120*. Odense, Odense University Press, 124 p.
- VINCENT, P. 1951. « La mortalité des vieillards », *Population*, 6, 2 : 181-204.

ANNEXE

TABLEAU 3 Estimation de l'émigration pour la génération de 1895, sexes réunis

Recensement	Groupes d'âge	Px* (estimation si pop fermée)	Quotients perspectifs kx	D(x+n)*	Stock		Évaluation émigration hors- Canada	Émigration totale entre période	Survie		Évaluation nouveaux émigrants
					Pop née QC hors- Québec	Stock Pop Qc née au Qc			émigration tot 1941	Survie Stock émigration tot 1951	
1901	0-4	237603	0,1809	42985							
	5-9	194618	0,0716	13943							
1911	10-14	180675	0,0172	3112							
	15-19	177563	0,0183	3247							
1921	20-24	174316	0,0247	4309							
	25-29	170007	0,0267	4532							
1931	30-34	165475	0,0254	4211	15401	154640	170041	10835	10835		10835
	35-39	161264	0,0246	3964				10568			
1941	40-44	157300	0,0298	4690	12637	147360	159997	9940	9940		
	45-49	152610	0,0357	5448				9887	9644		
1951	50-54	147162	0,0497	7314	13658	137981	151639	9180	9180		- 167
	55-59	139847	0,0681	9518				8756	8837		
1961	60-64	130330	0,1010	13168	11621	118407	130028	11922	8236	11922	3843
	65-69	117161	0,1420	16635				7872	8130		
1971	70-74	100527	0,2087	20982							
	75-79	79545	0,2941	23392							
1981	80-84	56153	0,3909	21949							
	85-89	34204	0,4951	16936							
1991	90-94	17268	0,6493	11213							
	95-99	6056	0,7825	4739							
2001	100-104	1317	0,8942	1178							
	105-109	139	1,0000	139							
	110+	0		0							
Total	% de la génération										14197 6,0%

Sources : BDLC et recensements canadiens.

ABSTRACT

**Laurence PILON-MARIEN, Alain GAGNON, Bertrand DESJARDINS and Robert BOURBEAU
Early life Conditions and Longevity: linking data from the 1901 Canadian Censuses to
Quebec's civil registration**

It has been suspected for a long time that early life conditions could affect adult mortality, and scholars endorsing a life course perspective have recently reported compelling evidence for the "long arm of childhood." As this niche has not yet been developed in the Canadian context, we have established a record-linkage project tracing the early life conditions of individuals who were listed as children in the 1901 Canadian Census and Quebec's death records. This paper focuses on the difficulties involved in the linkage process and on how selection biases may affect findings. First, we compare mortality patterns from our samples to the mortality tables derived from the Canadian Human Mortality Database at the Université de Montréal. Second, using logistic regression, we build a sample selection model for which the outcome is whether a death record is found. Many of the predictors included in this model (sex, urban-rural status, father's occupation, literacy, etc.) will also be used in our future models predicting mortality at older ages, helping us to assess potential biases and their impacts on findings.