

Université de Montréal

Changements dans la répartition des décès selon
l'âge : une approche non paramétrique pour
l'étude de la mortalité adulte

par

Nadine Ouellette

Département de démographie
Faculté des arts et des sciences

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de
Philosophiæ Doctor (Ph.D.)
en Démographie

mars 2011

Université de Montréal

Faculté des études supérieures

Cette thèse intitulée

Changements dans la répartition des décès selon
l'âge : une approche non paramétrique pour
l'étude de la mortalité adulte

présentée par

Nadine Ouellette

a été évaluée par un jury composé des personnes suivantes :

Bertrand Desjardins

(président-rapporteur)

Robert Bourbeau

(directeur de recherche)

Jutta Gampe

(co-directrice)

Marc Termote

(membre du jury)

Vladimir Canudas-Romo

(examineur externe)

Jean-François Angers

(représentant du doyen de la FES)

RÉSUMÉ

Au cours du siècle dernier, nous avons pu observer une diminution remarquable de la mortalité dans toutes les régions du monde, en particulier dans les pays développés. Cette chute a été caractérisée par des modifications importantes quant à la répartition des décès selon l'âge, ces derniers ne se produisant plus principalement durant les premiers âges de la vie mais plutôt au delà de l'âge de 65 ans. Notre étude s'intéresse spécifiquement au suivi fin et détaillé des changements survenus dans la distribution des âges au décès chez les personnes âgées. Pour ce faire, nous proposons une nouvelle méthode de lissage non paramétrique souple qui repose sur l'utilisation des *P-splines* et qui mène à une expression précise de la mortalité, telle que décrite par les données observées. Les résultats de nos analyses sont présentés sous forme d'articles scientifiques, qui s'appuient sur les données de la *Human Mortality Database*, la *Base de données sur la longévité canadienne* et le *Registre de la population du Québec ancien* reconnues pour leur fiabilité.

Les conclusions du premier article suggèrent que certains pays à faible mortalité auraient récemment franchi l'ère de la compression de la mortalité aux grands âges, ère durant laquelle les décès au sein des personnes âgées tendent à se concentrer dans un intervalle d'âge progressivement plus court. En effet, depuis le début des années 1990 au Japon, l'âge modal au décès continue d'augmenter alors que le niveau d'hétérogénéité des durées de vie au-delà de cet âge demeure inchangé. Nous assistons ainsi à un déplacement de l'ensemble des durées de vie adultes vers des âges plus élevés, sans réduction parallèle de la dispersion de la mortalité aux grands âges. En France et au Canada, les femmes affichent aussi

de tels développements depuis le début des années 2000, mais le scénario de compression de la mortalité aux grands âges est toujours en cours chez les hommes. Aux États-Unis, les résultats de la dernière décennie s'avèrent inquiétants car pour plusieurs années consécutives, l'âge modal au décès, soit la durée de vie la plus commune des adultes, a diminué de manière importante chez les deux sexes.

Le second article s'inscrit dans une perspective géographique plus fine et révèle que les disparités provinciales en matière de mortalité adulte au Canada entre 1930 et 2007, bien décrites à l'aide de surfaces de mortalité lissées, sont importantes et méritent d'être suivies de près. Plus spécifiquement, sur la base des trajectoires temporelles de l'âge modal au décès et de l'écart type des âges au décès situés au-delà du mode, les différentiels de mortalité aux grands âges entre provinces ont à peine diminué durant cette période, et cela, malgré la baisse notable de la mortalité dans toutes les provinces depuis le début du XX^e siècle. Également, nous constatons que ce sont précisément les femmes issues de provinces de l'Ouest et du centre du pays qui semblent avoir franchi l'ère de la compression de la mortalité aux grands âges au Canada.

Dans le cadre du troisième et dernier article de cette thèse, nous étudions la longévité des adultes au XVIII^e siècle et apportons un nouvel éclairage sur la durée de vie la plus commune des adultes à cette époque. À la lumière de nos résultats, l'âge le plus commun au décès parmi les adultes canadiens-français a augmenté entre 1740-1754 et 1785-1799 au Québec ancien. En effet, l'âge modal au décès est passé d'environ 73 ans à près de 76 ans chez les femmes et d'environ 70 ans à 74 ans chez les hommes. Les conditions de vie particulières de la population canadienne-française à cette époque pourraient expliquer cet accroissement.

Mots-clés : âge modal au décès, compression de la mortalité aux grands âges, report de la mortalité, longévité humaine, lissage par P-splines, surfaces de mortalité, Canada, États-Unis, France, Japon, provinces canadiennes, Québec ancien

ABSTRACT

Over the course of the last century, we have witnessed major improvements in the level of mortality in regions all across the globe, in particular in developed countries. This remarkable mortality decrease has also been characterized by fundamental changes in the mortality profile by age. Indeed, deaths are no longer occurring mainly at very young ages but rather at advanced ages such as above age 65. Our research focuses on monitoring and understanding historical changes in the age-at-death distribution among the elderly population. We propose a new flexible nonparametric smoothing approach based on *P-splines* leading to detailed mortality representations, as described by actual data. The results are presented in three scientific papers, which rest upon reliable data taken from the *Human Mortality Database*, the *Canadian Human Mortality Database*, and the *Registre de la population du Québec ancien*.

Findings from the first paper suggest that some low mortality countries may have recently reached the end of the old-age compression of mortality era, where deaths among the elderly population tend to concentrate into a progressively shorter age interval over time. Indeed, since the early 1990s in Japan, the modal age at death continues to increase while reductions in the variability of age at death above the mode have stopped. Thus, the distribution of age at death at older ages has been sliding towards higher ages without changing its shape. In France and Canada, women show such developments since the early 2000s, whereas men are still boldly engaged in an old-age mortality compression regime. In the USA, the picture for the latest decade is worrying because for several consecutive years in that timeframe, women and men have both recorded important declines in

their modal age at death, which corresponds to the most common age at death among adults.

The second paper takes a look within national boundaries and examines regional adult mortality differentials in Canada between 1930 and 2007. Smoothed mortality surfaces reveal that provincial disparities among adults in general and among the elderly population in particular are substantial in this country and deserve to be monitored closely. More specifically, based on modal age at death and standard deviation above the mode time trends, provincial disparities at older ages have barely reduced during the period studied, despite the great mortality improvements recorded in all provinces since the early XXth century. Also, we find that women who have reached the end of the old-age compression of mortality era in Canada are respectively those of Western and Central provinces.

The last paper focuses on adult longevity during the XVIIIth century in historical Quebec and provides new insight on the most common adult age at death. Indeed, our analysis reveals that the modal age at death increased among French-Canadian adults between 1740-1754 and 1785-1799. In 1740-1754, it was estimated at 73 years among females and at about 70 years among males. By 1785-1799, modal age at death estimates were almost 3 years higher for females and 4 years higher for males. Specific living conditions of the French-Canadian population at the time could explain these results.

Keywords : Modal age at death, Old-age mortality compression, Shifting mortality, Human longevity, P-spline smoothing, Mortality surfaces, Canada, United States of America, France, Japan, Canadian provinces, Historical Quebec

TABLE DES MATIÈRES

Résumé	iii
Abstract	v
Liste des figures	xi
Liste des tableaux	xiv
Dédicace	xv
Remerciements	xvi
Introduction générale	1
Chapitre 1. Recension des études	8
1.1. Répartition des décès selon l'âge	8
1.1.1. Fries et ses successeurs.....	8
1.1.2. Contributions canadiennes	10
1.1.3. Les indicateurs.....	11
1.1.3.1. Mesures de tendance centrale	11
1.1.3.2. Mesures de dispersion.....	13
1.1.4. Résultats des études articulées autour de l'âge modal au décès..	15
1.2. Lois et modèles de mortalité	18
1.2.1. Le profil de la mortalité selon l'âge.....	19
1.2.2. Le profil de la mortalité selon l'âge et les années de calendrier ..	22
1.3. Des modèles paramétriques aux méthodes de lissage plus souples ..	24
1.3.1. La percée des P-splines	24
1.3.2. Objectifs et questions de recherche	25
Chapitre 2. Sources de données et méthodes d'analyse	27
2.1. Sources et qualité des données	27
2.1.1. La <i>Human Mortality Database</i>	27

2.1.2.	<i>La Base de données sur la longévité canadienne</i>	33
2.1.3.	<i>Le Registre de population du Québec ancien</i>	36
2.2.	Méthodes d'analyse	39
2.2.1.	Lissage par P-splines	39
2.2.1.1.	Les B-splines et les P-splines	39
2.2.1.2.	La fonction de vraisemblance pénalisée	43
2.2.2.	Les P-splines pour l'étude de la mortalité	44
2.2.2.1.	Lissage par P-splines des données selon l'âge	44
2.2.2.2.	Généralisation pour des données selon l'âge et l'année de calendrier	48
2.2.3.	Mesure des transformations de la fonction de densité lissée au fil du temps	54
2.2.3.1.	L'âge modal au décès et la dispersion des durées de vie au-delà de cet âge	54
Chapitre 3. Changes in the age-at-death distribution in four low mortality countries : A nonparametric approach		56
3.1.	Introduction	57
3.2.	Background	58
3.2.1.	Average lifespan	58
3.2.2.	Lifespan inequality	61
3.2.3.	Choice of a nonparametric method	62
3.3.	Data and methods	64
3.4.	Results	66
3.4.1.	Illustration of the nonparametric P-spline approach using Japanese data	66
3.4.2.	Adult mortality trends in four low mortality countries	69
3.4.2.1.	Modal age at death estimates (\hat{M})	69
3.4.2.2.	Standard deviation of ages at death above the mode estimates ($SD(\widehat{M+})$)	72
3.5.	Discussion	75
3.6.	Acknowledgements	78
3.7.	Appendix	79

3.7.1.	From B-splines to P-splines	79
3.7.2.	The penalized likelihood function for P-splines	80
3.7.3.	Smoothing mortality data with P-splines	82
3.7.4.	Comparison between HMD life table age-at-death distributions and P-spline smoothed density functions for Japan	82
Chapitre 4. Regional disparities in Canadian adult mortality : A comparative study based on smoothed mortality surfaces and age-at-death distributions		84
4.1.	Introduction	85
4.2.	Study objectives	87
4.3.	Data	88
4.4.	Methods	90
4.4.1.	Smoothing mortality data with P-splines	90
4.4.2.	Computation and use of smoothed age-at-death distributions...	91
4.5.	Results	93
4.5.1.	Uncovering provincial variations via smoothed mortality surfaces	93
4.5.2.	Summarizing changes in smoothed age-at-death distributions over time across provinces	98
4.6.	Discussion	103
4.7.	Notes	107
4.8.	Acknowledgements	108
4.9.	Appendix A : The P-spline method in the context of mortality	108
4.9.1.	Smoothing in one dimension	108
4.9.2.	Smoothing in two dimensions	109
4.10.	Appendix B : 99% bootstrap confidence intervals	111
Chapitre 5. La durée de vie la plus commune des adultes au XVIII^e siècle : l'expérience des Canadiens-français ..		116
5.1.	Introduction	117
5.2.	Répartition des décès selon l'âge d'après diverses tables de mortalité des populations du passé	118

5.2.1.	Table de mortalité de Halley : absence d'une durée de vie modale	118
5.2.2.	Tables de mortalité de Wargentin et Deparcieux : durée de vie modale autour de 70 ans	119
5.2.3.	Premières tables de mortalité complètes en Suisse : la durée de vie modale n'aurait pas changé	121
5.3.	Répartition des décès selon l'âge des Canadiens-français au XVIII ^e siècle	122
5.3.1.	Registre de la population du Québec ancien	122
5.3.2.	Tables de mortalité complètes pour les mariés	125
5.3.3.	Lissage des données de mortalité à l'aide de la méthode des P-splines	127
5.4.	Résultats	129
5.5.	Discussion	133
5.6.	Remerciements	135
	Conclusion	136
	Bibliographie	143

LISTE DES FIGURES

1.1	Concept des durées de vie normales introduit par Lexis	12
1.2	Illustration de l'âge modal au décès (M) et de l'écart type, relatif à M , des durées de vie situées au-delà de M ($SD(M+)$), femmes, Japon, 2000-2004.....	16
2.1	Schéma des données impliquées dans le calcul de la population exposée au risque de décéder	30
2.2	Méthodes de calcul des estimations de population	31
2.3	B-splines (en noir) avec noeuds (en rouge) et parties de polynômes (en bleu) : (a) linéaire, (b) quadratique, (c) cubique.....	40
2.4	Base B-spline incluant huit B-splines cubiques avec noeuds équidistants sur le domaine de x	41
2.5	Modélisation de données (simulées) à l'aide d'une base B-spline incluant huit B-splines cubiques avec noeuds équidistants sur le domaine de x .	43
2.6	Régression sur données simulées : (a) non pénalisée, (b) pénalisée	45
2.7	Taux de mortalité observés selon l'âge et force de mortalité lissée à l'aide des P-splines, France, femmes, 10 à 109 ans, 2005	48
2.8	Fonction de densité lissée à l'aide des P-splines décrivant la répartition des décès selon l'âge, France, femmes, 2005.....	49
2.9	Produit de Kronecker de deux B-splines cubiques sur le domaine des âges et années de calendrier	50
2.10	Produit de Kronecker de deux bases B-splines cubiques sur le domaine des âges et années de calendrier	51
2.11	Taux de mortalité observés (haut) et force de mortalité lissée à l'aide des P-splines en deux dimensions (bas), Colombie-Britannique, hommes, 10 à 105 ans, 1930 à 2007.....	52
2.12	Fonction de densité lissée à l'aide des P-splines en deux dimensions, Colombie-Britannique, hommes, 10 à 105 ans, 1930 à 2007	53

3.1	Life table death distribution, USA, Women, 2003	60
3.2	Comparison between actual death counts (points) and fitted death counts (lines) resulting from nonparametric P-spline estimations of Poisson regression models, Japan, 1947 to 2008	67
3.3	Smoothed density functions describing the age-at-death distribution and resulting from nonparametric P-spline estimations of Poisson regression models, Japan, 1947 to 2008.....	68
3.4	Estimated modal age at death trends in four low mortality countries based on smoothed density functions.....	70
3.5	Estimated standard deviation of ages at death above the mode in four low mortality countries based on smoothed density functions.....	73
3.6	Comparison between Fries's predicted age-at-death distribution and empirical smoothed age-at-death distributions among women in four low mortality countries	78
3.7	B-spline basis containing 8 cubic B-splines with equally-spaced knots.	80
3.8	Unpenalized (left panel) and penalized (right panel) regression using simulated data and 23 equally-spaced cubic B-splines	81
3.9	Comparison between life table deaths distributions (overplotted points and lines) downloaded from the HMD and smoothed density functions (plain solid lines) resulting from nonparametric P-spline estimations of Poisson regression models, Japan, 1947 to 2008	83
4.1	Province-specific smoothed mortality surfaces based on fitted death rates with two-dimensional Poisson P-splines, Females, Canada, 1930 to 2007.....	95
4.2	Province-specific smoothed mortality surfaces based on fitted death rates with two-dimensional Poisson P-splines, Males, Canada, 1930 to 2007.....	96
4.3	Modal age at death by province and sex estimated from two-dimensional smoothed age-at-death distribution, Canada, 1930 to 2007	100
4.4	Standard deviation above the modal age at death by province and sex estimated from two-dimensional smoothed age-at-death distribution, Canada, 1930 to 2007.....	102
4.5	Life expectancy at birth by province and sex, Canada, 1930 to 2007..	105

4.6	Two-dimensional Kronecker product of two cubic B-splines	110
4.7	Modal age at death estimates and corresponding 99% bootstrap confidence intervals, British Columbia and Quebec, 1930 to 2007	113
5.1	Répartition des décès selon l'âge chez les adultes d'après la table de mortalité complète de Halley, sexes réunis, Breslau, 1687-1691	120
5.2	Répartition des décès selon l'âge et le sexe chez les adultes en Suède, 1754-1756, et en Suisse, 1876-1880	121
5.3	Schéma descriptif des données sur les mariés tirées du Registre de la population du Québec ancien	124
5.4	Effectifs de la population mariée retenue selon le sexe, Québec ancien, 1 ^{er} juillet, 1640 à 1799	125
5.5	Nombre annuel de décès de mariés selon le sexe, Québec ancien, 1640 à 1799	126
5.6	Répartition des décès des adultes mariés selon le sexe et l'âge d'après les tables de mortalité pour les périodes 1740-1754 à 1784-1799, Québec ancien	130
5.7	Comparaison de la répartition des décès de la table de mortalité et de celle résultant d'un lissage par P-splines, Hommes mariés, Québec ancien, 1740-1754	131
5.8	Durée de vie la plus commune chez les adultes mariés selon le sexe et la période, Québec ancien, 1740-1754 à 1785-1799	132

LISTE DES TABLEAUX

- 4.1 Modal age at death estimates and corresponding 99% bootstrap confidence intervals by province and sex, Canada, 1930, 1950, 1970, 1990, and 2007 (Females); 1970, 1980, 1990, 2000, and 2007 (Males) . 114
- 4.2 Standard deviation above the modal age at death estimates and corresponding 99% bootstrap confidence intervals by province and sex, Canada, 1930, 1950, 1970, 1990, and 2007 (Females); 1970, 1980, 1990, 2000, and 2007 (Males) 115
- 5.1 Estimations de la durée de vie la plus commune chez les adultes mariés, Québec ancien, 1740-1754 à 1785-1799 133

DÉDICACE

À mes parents, Alain et Lynette,
à ma soeur Marie-Chantal,
et à mon copain Fayssal.

REMERCIEMENTS

Je tiens d’abord à remercier tout particulièrement mon directeur de recherche, Robert Bourbeau, qui m’a accordé sa précieuse confiance au cours des quatre dernières années. J’ai eu la chance d’être supervisée par un mentor très exigeant, mais qui s’avère aussi un motivateur remarquable. Ses conseils et encouragements, toujours parfaitement opportuns, ont manifestement contribué à accroître la qualité de mes travaux de recherche. De plus, sa passion contagieuse pour l’analyse démographique et son ardente détermination à défendre la place de la démographie au coeur des recherches en sciences humaines et sociales ont été des sources d’inspiration continuelles. Je suis grandement reconnaissante pour toutes les compétences qu’il m’aura permis d’acquérir durant mon parcours au doctorat.

Je suis également très reconnaissante envers ma codirectrice, Jutta Gampe, qui s’est avérée une source additionnelle d’inspiration exceptionnelle. Je me considère privilégiée d’avoir pu séjourner à plusieurs occasions au laboratoire de démographie statistique qu’elle dirige au *Max Planck Institute for Demographic Research* à Rostock en Allemagne. Mes séjours y ont tous été fort inspirants pour mes travaux de thèse et ont certainement contribué à les bonifier. J’en profite pour souligner tout particulièrement le soutien très apprécié de Carlo G. Camarda lors de mon apprentissage des P-splines et de leur adaptation aux données sur la mortalité.

Je remercie aussi mes deux autres coauteurs, Bertrand Desjardins et Jean-Marie Robine, pour avoir suscité en moi le désir de m’aventurer dans les eaux de la démographie historique. Grâce à leur appui, j’ai eu beaucoup de plaisir à

découvrir cette dimension de la discipline et à réaliser le troisième article de cette thèse.

J'aimerais aussi exprimer ma profonde gratitude envers le *Conseil de recherches en sciences humaines du Canada* et le *Fonds de recherche sur la société et la culture du Québec*, pour avoir reconnu la valeur de ce projet de recherche doctorale et cru en ma capacité de le mener à bon terme. Le soutien de ces organismes subventionnaires a été largement apprécié. De même, je remercie le Département de démographie de l'Université de Montréal, la Faculté des études supérieures et postdoctorales de l'Université de Montréal ainsi que le *Groupe financier Banque TD*, qui ont aussi contribué au financement de ce projet. Enfin, grâce au support du Réseau stratégique de connaissances *Changements de population et parcours de vie (University of Western Ontario)*, j'ai pu participer à plusieurs colloques scientifiques internationaux dans le domaine de la démographie et entre autres y présenter les résultats de mes travaux de thèse.

D'autre part, mon périple au doctorat n'aurait pu être si agréable sans la présence de mes superbes collègues de l'équipe de recherche *Mortalité et longévité* au Département de démographie. Plus particulièrement, je remercie chaleureusement mes amies Camille Bouchard-Coulombe, Valérie Jarry et Chantale Lecours pour leur écoute attentive, leurs multiples encouragements, leurs nombreux conseils, bref leur support inconditionnel. Elles méritent également une mention spéciale pour leur très grande détermination à me faire sortir du C-5044 de temps en temps!

Quant à vous, mes très chers amies et amis non démographes, Sania, Nazha, Khouloud, Rita, René, Nabil, Ali, Mathieu, Malek, Rony, Rudy, Michel et tous les autres que j'ai oublié de nommer mais qui sauront se reconnaître, merci d'être là au quotidien. Vous êtes tous exceptionnels et je vous adore!

Enfin, mes derniers remerciements s'adressent de toute évidence à mes parents, ma soeur et mon copain. La dédicace en dit long ; si ce n'était de vous, ce merveilleux projet n'aurait jamais pu être réalisé.

INTRODUCTION GÉNÉRALE

Le XX^e siècle a été marqué par une diminution spectaculaire du niveau général de la mortalité, qui s'est accompagnée d'un changement important du profil selon l'âge : plutôt que de survenir aux premiers âges de la vie, la mortalité s'est concentrée à des âges de plus en plus avancés. L'expression *compression de la mortalité* est couramment utilisée pour désigner ce phénomène de concentration des décès dans un intervalle d'âge progressivement plus court. Le processus de *rectangularisation de la courbe de survie* enregistré au cours du siècle dernier dans les pays développés, c'est-à-dire le fait que l'allure de la courbe de survie soit devenue plus rectangulaire au fil du temps, est d'ailleurs étroitement lié à ce dernier phénomène de réduction de la dispersion des âges au décès (Wilmoth et Horiuchi, 1999). Le suivi des transformations de la courbe de survie ou dans la répartition des décès selon l'âge est rapidement devenu un sujet de grand intérêt auprès des chercheurs, notamment en raison de leurs profondes implications sur la gestion et l'organisation de nos sociétés. En effet, l'arrivée en proportion grandissante de la population aux âges avancés n'est pas sans répercussions majeures sur la planification des besoins en soins de santé, services sociaux et sécurité du revenu.

La recension des écrits montre l'existence de plus d'une vingtaine d'indicateurs servant à quantifier les changements survenus en matière de répartition des décès selon l'âge (Cheung *et al.*, 2005). Ces indicateurs visent essentiellement à résumer la tendance centrale et la dispersion des âges au décès. Les récents développements à ce sujet soulignent notamment l'importance de porter une attention

non seulement à l'espérance de vie à la naissance, c'est-à-dire l'âge moyen au décès, mais également à l'âge médian au décès et surtout à l'âge modal au décès chez les adultes (Canudas-Romo, 2008, 2010; Cheung *et al.*, 2005, 2008, 2009; Cheung et Robine, 2007; Kannisto, 2000, 2001, 2007; Paccaud *et al.*, 1998; Robine, 2001; Thatcher *et al.*, 2010). En effet, cet âge modal au décès est strictement influencé par les conditions de mortalité à l'âge adulte et il est ainsi nettement plus sensible aux changements qui surviennent au sein de cette sous-population (Canudas-Romo, 2010; Horiuchi, 2003; Kannisto, 2001). Considérant que depuis le milieu du XX^e siècle, l'essentiel des modifications quant à la répartition des décès selon l'âge dans les pays développés est alimenté par la baisse soutenue de la mortalité aux âges avancés, l'âge modal au décès constitue un outil particulièrement adapté au suivi et à l'explication de ces mutations de la mortalité chez les adultes.

Jusqu'à présent, l'estimation des indicateurs de tendance centrale et de dispersion des âges au décès, en particulier l'âge modal au décès, s'est réalisée dans un contexte de modélisation plutôt rigide. En effet, le calcul de ces indicateurs requiert habituellement l'utilisation de méthodes statistiques paramétriques qui imposent une structure prédéterminée aux données. Il s'agit là d'une limite qui a été soulevée à quelques reprises (Cheung *et al.*, 2005, 2008; Cheung et Robine, 2007) puisque cette pratique est susceptible d'exercer une influence sur l'estimation des indicateurs. Les modèles non paramétriques, qui permettent un relâchement des hypothèses relatives à la structure des données, représentent donc une excellente avenue de recherche qui mérite grandement d'être explorée.

Ainsi, l'objectif premier de la présente thèse consiste à proposer une nouvelle approche de lissage non paramétrique souple, permettant un meilleur suivi des changements survenus dans la répartition des décès selon l'âge, en particulier aux âges avancés. Plus spécifiquement, la méthode suggérée repose sur l'utilisation des P-splines (Eilers et Marx, 1996), qui jouissent d'une popularité grandissante dans plusieurs domaines appliqués dont celui de l'analyse de la mortalité (Camarda, 2008; Currie *et al.*, 2004, 2006; Eilers *et al.*, 2006). Soulignons que par l'entremise

de cette proposition d'une méthode statistique novatrice pour l'étude de la mortalité, notre travail de recherche s'inscrit dans un effort de rapprochement entre la démographie et la statistique.

Notre second objectif est d'analyser de plus près les changements survenus récemment dans la répartition des décès aux grands âges au sein d'un sous-ensemble de pays à faible mortalité. En effet, des travaux parus dernièrement suggèrent que certains pays à faible mortalité pourraient maintenant avoir franchi une nouvelle ère en cette matière (Canudas-Romo, 2008; Cheung *et al.*, 2008, 2009; Cheung et Robine, 2007; Thatcher *et al.*, 2010). Dans ces pays, les auteurs notent un déplacement de l'ensemble des durées de vie adultes vers des âges plus élevés, sans réduction parallèle de la dispersion de la mortalité aux grands âges. Cette nouvelle ère serait donc en conformité avec le *shifting mortality regime*, décrit par Kannisto (1996) ainsi que Bongaarts et Feeney (2002, 2003) et développé davantage par Bongaarts (2005).

En guise de troisième objectif, nous souhaitons faire le point sur les disparités régionales à propos des transformations quant à la répartition des décès selon l'âge au Canada. Jusqu'à présent, seules deux études (Martel et Bourbeau, 2003; Nagnur, 1986) ont porté sur le sujet au Canada et aucune d'entre elles n'a exploré les différences entre les provinces canadiennes.

Enfin, notre quatrième objectif consiste à apporter un nouvel éclairage sur la durée de vie la plus commune des adultes au XVIII^e siècle. À ce jour, nous détenons très peu d'information sur la durée de vie *typique* des adultes avant le XIX^e siècle, notamment en raison de la rareté de données historiques détaillées et fiables sur la mortalité de l'époque. Ainsi, nous ignorons si l'âge le plus commun au décès chez les adultes a augmenté, diminué, fluctué ou s'il a plutôt eu tendance à demeurer stable durant une très longue période de temps dans le passé. Ces informations s'avèrent toutefois essentielles à notre compréhension de la longévité humaine.

Plus généralement, cette recherche doctorale nous mènera inévitablement à nous intéresser au débat que suscite l'amélioration de la survie aux grands âges, à savoir jusqu'à quel âge pourront être repoussés les décès les plus tardifs chez les humains. Depuis les travaux précurseurs de Vincent (1951), qui laissaient entendre que la durée de vie humaine était limitée à 110 ans, l'existence d'une durée maximale de la vie humaine ou d'un âge limite à la vie humaine a alimenté plusieurs discussions. Après plus d'un demi-siècle de travaux orientés vers l'étude des limites potentielles à la vie humaine, les conclusions des chercheurs demeurent partagées. D'ailleurs, au début des années 1990, Manton *et al.* (1991) soutenaient que les différentes perspectives à ce sujet pouvaient être séparées en trois groupes : les *traditionalistes* (qui suggèrent que les limites sont immanentes), les *visionnaires* (qui envisagent des percées biomédicales révolutionnaires en matière de vieillissement), puis les *empiristes* (qui admettent que les limites, si elles existent, sont hors de portée de vue puisque des progrès importants en matière de mortalité aux âges avancés ont toujours lieu). Plus récemment, Carnes et Olshansky (2007) parlaient plutôt de *futuristes* (qui considèrent l'accessibilité à l'immortalité), d'*optimistes* (qui croient que la durée de vie moyenne pourrait atteindre et même dépasser 100 ans d'ici la fin du XXI^e siècle) et de *réalistes* (qui suggèrent une durée de vie moyenne d'environ 85 ans au maximum). Quelles que soient les étiquettes employées, celles-ci illustrent bien la persistance du litige.

En résumé, la thèse est composée de cinq chapitres, dont trois constituent des articles de recherche scientifique. Les deux premiers chapitres servent d'introduction à ces articles et sont organisés de la façon suivante. Le chapitre 1 consiste en une recension des études et est divisé en trois parties. Nous examinons d'abord les divers indicateurs issus des travaux visant à effectuer un suivi des transformations de la répartition des décès selon l'âge. Nous nous attardons ensuite aux lois de mortalité, ces modèles mathématiques qui décrivent l'évolution de la mortalité en fonction de l'âge, en raison de leur proximité avec la façon dont les décès se répartissent selon l'âge. Enfin, la percée des P-splines dans le domaine de la mortalité est abordée et nous faisons état de la pertinence de cet outil de modélisation souple quant au suivi des changements dans la répartition des décès selon l'âge.

Le second chapitre traite des sources de données utilisées dans chacun des trois articles, ainsi que des méthodes retenues pour les analyser. Dans l'ordre, la *Human Mortality Database*, la *Base de données sur la longévité canadienne* et le *Registre de la population du Québec ancien* sont présentés de manière détaillée et la qualité de ces sources d'information y est examinée. Ensuite, compte tenu de la place importante occupée par la méthode de lissage par P-splines au sein de nos travaux de recherche, celle-ci est examinée en profondeur à l'aide d'exemples illustratifs.

Les chapitres 3, 4 et 5 correspondent aux trois articles et constituent le coeur de cette thèse. Le premier article exposé au chapitre 3 sert de tribune pour la présentation de cette nouvelle approche de lissage non paramétrique par P-splines, qui offre un suivi précis des changements survenus dans la répartition des décès selon l'âge à travers le temps. À partir de données sur la mortalité au Canada (1921 à 2007), en France (1920 à 2007), au Japon (1947 à 2008) et aux États-Unis (1945 à 2006), la méthode de lissage par P-splines selon l'âge est utilisée pour étudier les modifications récentes de la mortalité aux âges avancés dans ces pays à faible mortalité. Chacun des pays considéré constitue un cas d'espèce et mérite une analyse détaillée selon le sexe.

Le chapitre 4 porte sur les disparités provinciales en matière de mortalité adulte au Canada entre 1930 et 2007. L'analyse repose sur la méthode de lissage par P-splines selon l'âge et les années de calendrier, soit une généralisation de la méthode proposée dans l'article précédent. En effet, l'approche de lissage en deux dimensions s'avère particulièrement utile pour l'étude de populations de petite taille, telles que certaines provinces canadiennes. Les écarts de mortalité à l'âge adulte entre les dix provinces canadiennes sont explorés sous l'angle des surfaces de mortalité lissées, ainsi que celui des trajectoires temporelles de l'âge modal au décès et de l'écart type des âges au décès situés au-delà du mode.

Le dernier article présenté au cinquième chapitre traite de la longévité des adultes sous l'angle de la démographie historique. Nous retraçons d'abord l'évolution de la durée de vie la plus commune des adultes à travers les diverses tables

de mortalité déjà existantes et reposant sur des données des XVII^e, XVIII^e et XIX^e siècles. Ensuite, nous procédons au suivi de l'évolution de la durée de vie modale des adultes au sein de la population des Canadiens-français au Québec ancien durant la période 1740-1754 à 1785-1799, à l'aide de la méthode de lissage par P-splines unidimensionnelle.

Il est à noter que je suis la principale et première auteure de ces trois articles. Pour chacun d'entre eux, mes contributions et celles de mes coauteurs sont brièvement décrites ci-dessous.

Article 1 : « Changes in the age-at-death distribution in four low mortality countries : A nonparametric approach »

Nadine Ouellette

- Direction du projet
- Développement de la nouvelle méthode par P-splines
- Implémentation sous R
- Analyse des résultats
- Rédaction de l'article

Robert Bourbeau

- Participation à l'analyse de certains résultats
- Relecture de l'article

Article 2 : « Regional disparities in Canadian adult mortality : A comparative study based on smoothed mortality surfaces and age-at-death distributions »

Nadine Ouellette

- Direction du projet
- Généralisation de la nouvelle méthode par P-splines afin d'introduire l'année de calendrier en tant que dimension supplémentaire
- Adaptation des intervalles de confiance bootstrap à l'approche par P-splines
- Implémentation sous R
- Mise à jour de la Base de données sur la longévité canadienne
- Programmation sous R
- Analyse des résultats

- Rédaction de l'article

Robert Bourbeau

- Participation à l'analyse de certains résultats
- Relecture de l'article

Carlo Giovanni Camarda

- Participation à l'adaptation des intervalles de confiance bootstrap à l'approche par P-splines
- Relecture de l'article

Article 3 : « La durée de vie la plus commune des adultes au XVIIIe siècle : l'expérience des Canadiens-français »

Nadine Ouellette

- Direction du projet
- Traitement des données du Registre de la Population du Québec Ancien
- Application de la nouvelle méthode par P-splines sous R
- Construction des tables de mortalité du moment
- Analyse des résultats
- Rédaction de l'article

Jean-Marie Robine

- Relecture de l'article

Robert Bourbeau

- Relecture de l'article

Bertrand Desjardins

- Transmission des données du Registre de la Population du Québec Ancien
- Relecture de l'article

La conclusion générale de cette thèse est l'occasion de synthétiser les principaux résultats issus de chacun des trois articles et de souligner leur contribution à l'avancement des connaissances en matière de longévité des adultes. Quelques pistes visant à orienter les recherches futures sont également proposées.

Chapitre 1

Recension des études

Il existe un large éventail d'études portant sur la problématique qui nous intéresse. Plusieurs de ces études sont citées dans l'un ou l'autre des trois articles (chapitres 3, 4 et 5) qui forment le coeur de cette thèse. Ce premier chapitre se veut donc une recension des principales études sur lesquelles s'appuient les analyses présentées plus loin et qui donnent le contexte général dans lequel s'inscrivent nos travaux.

Le chapitre est composé de trois sections. La première consiste en une revue générale des études portant sur le thème des transformations dans la répartition des décès selon l'âge. La seconde présente les lois et les modèles de mortalité les plus couramment rencontrés en démographie, ceux-ci étant étroitement liés à la façon dont les décès se répartissent en fonction de l'âge, de même qu'en fonction de l'âge et des années de calendrier. Enfin, la troisième section traite d'une approche de modélisation non paramétrique émergente et fort prometteuse pour l'étude de la mortalité : la méthode de lissage par P-splines. Nous expliquons en quoi l'utilisation de cette approche est pertinente pour effectuer un suivi précis des changements dans la répartition des décès selon l'âge.

1.1. RÉPARTITION DES DÉCÈS SELON L'ÂGE

1.1.1. Fries et ses successeurs

Le biologiste Comfort (1956, 1964) fut le premier à introduire le concept de *rectangularisation de la courbe de survie*. Ce concept fut ensuite abordé par d'autres

chercheurs (Gordon, 1980; Strehler, 1975; Vaupel *et al.*, 1979), avant qu'il ne soit finalement popularisé par le médecin James Fries (1980). D'après Fries, la courbe de survie chez les humains tend à devenir rectangulaire au fur et à mesure que le niveau de mortalité diminue puisque la durée de vie humaine est limitée. Dans des conditions *idéales*, soit en l'absence de mortalité prématurée, Fries fixe la durée de vie moyenne à 85 ans, principalement en raison de facteurs génétiques. Il ajoute que les décès seront alors fortement concentrés autour de celle-ci et que l'écart type pour l'ensemble des durées de vie vaudra précisément 4 ans. D'autre part, il soutient qu'un report des maladies chroniques résultera non seulement en une forte rectangularisation de la courbe de survie, mais également de celle de la courbe de morbidité, menant ainsi à une *compression de la morbidité*. La théorie de Fries ne fit pas l'unanimité auprès des scientifiques et donna rapidement lieu à des échanges enflammés (Fries, 1984; Myers et Manton, 1984a,b; Schneider et Brody, 1983). En effet, la plupart de ses opposants considéraient qu'une expansion plutôt qu'une compression de la morbidité pourrait survenir si la baisse de la mortalité aux grands âges se poursuivait.

Depuis les travaux de Fries, plusieurs chercheurs ont étudié les transformations de la courbe de survie et les changements dans la répartition de décès selon l'âge au sein de divers pays et régions. Spécifiquement, leurs recherches ont porté sur le Canada (Nagnur, 1986), les États-Unis (Eakin et Witten, 1995; Fries, 1980; Lynch et Brown, 2001; Manton et Tolley, 1991; Myers et Manton, 1984a,b; Rothenberg *et al.*, 1991), la France (Robine, 2001), la ville de Hong Kong (Cheung *et al.*, 2005), le Japon (Cheung et Robine, 2007), les Pays-Bas (Nusselder et Mackenbach, 1996, 1997), la province de Québec (Martel et Bourbeau, 2003) et enfin, la Suisse (Cheung *et al.*, 2009; Paccaud *et al.*, 1998). Des études comparatives impliquant plusieurs pays à faible mortalité ont aussi été présentées (Canudas-Romo, 2008; Cheung *et al.*, 2008; Hill, 1993; Thatcher *et al.*, 2010; Wilmoth et Horiuchi, 1999). La plupart de ces travaux confirment qu'au cours de la transition épidémiologique (Omran, 1971), les décès se sont effectivement concentrés dans un intervalle d'âge progressivement plus court et la forme de la courbe de survie est devenue de plus en plus rectangulaire. Autrement dit, les phénomènes

de *compression de la mortalité* et de rectangularisation de la courbe de survie ont été observés dans la grande majorité de ces pays et régions.

Au fil de ces recherches, plus d'une vingtaine d'indicateurs ont été proposés pour mesurer les transformations affectant la courbe de survie ou la répartition des décès selon l'âge (Cheung *et al.*, 2005). Ces indicateurs, qui visent essentiellement à résumer la tendance centrale et la dispersion des âges au décès au sein de la population, font l'objet de la sous-section 1.1.3.

1.1.2. Contributions canadiennes

Étant donné la place occupée par la situation canadienne dans les articles de cette thèse, il est utile d'apporter quelques précisions au sujet des deux études canadiennes citées ci-haut (Martel et Bourbeau, 2003; Nagnur, 1986). Celle de Nagnur (1986) porte sur le Canada, ses provinces et territoires réunis, et se concentre sur la période allant de 1921 à 1981. Elle repose sur un seul indicateur de dispersion des décès selon l'âge, soit l'entropie H_a définie par Keyfitz (1977), qui constitue une mesure d'élasticité de l'espérance de vie. Plus précisément, il s'agit du pourcentage d'amélioration de l'espérance de vie à l'âge a , pour une diminution équivalente à 1% de la mortalité se produisant à l'âge a et au-dessus. La contribution de Martel et Bourbeau (2003), tirée du mémoire de maîtrise de Martel (2002), porte quant à elle sur la province de Québec et couvre la période 1921-1999. Elle fait intervenir plusieurs indicateurs résumant la tendance centrale et la dispersion des décès selon âge, dont l'âge modal au décès et le pourcentage de décès survenant à cet âge modal (Kannisto, 2001). Martel (2002) a également eu recours à quelques indicateurs additionnels, dont l'écart type des durées de vie situées au-delà de l'âge modal au décès (Kannisto, 2001). Tout comme la majorité des auteurs ayant analysé la situation dans d'autres pays et régions, Nagnur ainsi que Martel et Bourbeau ont constaté la présence du phénomène de rectangularisation de la courbe de survie durant les périodes étudiées. Martel et Bourbeau ont cependant souligné que la compression de la mortalité semblait ralentir depuis les années 1960 au Québec. Nos analyses nous permettront notamment d'effectuer une mise à jour de la situation canadienne en matière de changements dans la

façon dont les décès se répartissent selon l'âge, particulièrement aux grands âges, et d'explorer les différentiels entre les provinces.

1.1.3. Les indicateurs

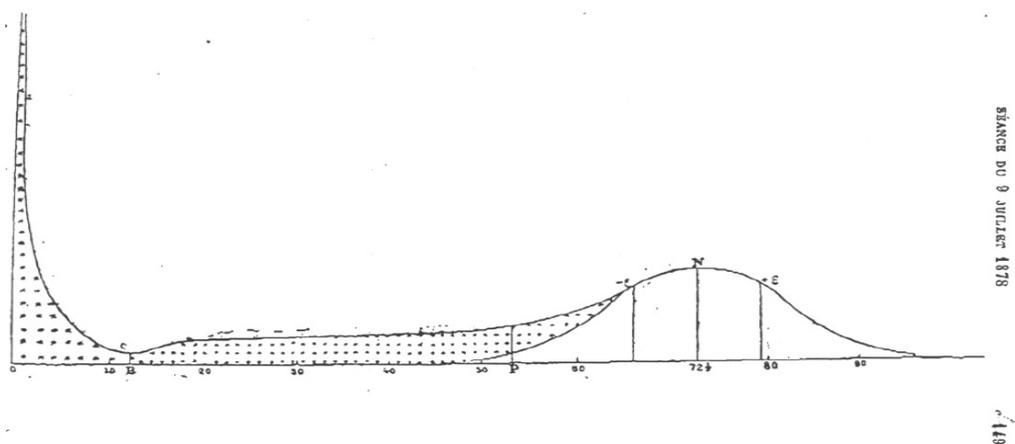
Devant l'abondance d'indicateurs proposés par les chercheurs et l'absence de consensus quant à ceux qui devraient être privilégiés pour observer correctement les changements dans la répartition des décès selon l'âge, les efforts de synthèse critique de ces différentes mesures effectuées par Wilmoth et Horiuchi (1999), Kannisto (2000), Robine (2001) et Cheung *et al.* (2005) se sont avérés grandement éclairants. On cherchera ici à présenter les principales conclusions issues de ces revues des indicateurs, tout en mettant l'accent sur les développements récents en matière de mesures résumant la tendance centrale et la dispersion des âges au décès.

1.1.3.1. *Mesures de tendance centrale*

Jusqu'à quel âge vivons-nous en moyenne ? correspond probablement à la question la plus courante en démographie. La réponse traditionnelle apportée à cette question est bien entendu l'espérance de vie à la naissance, soit la durée de vie moyenne des individus au sein de la population. Les durées de vie médiane et modale, des mesures pourtant complémentaires de tendance centrale de la répartition des décès selon l'âge, sont pour leur part beaucoup moins utilisées (Canudas-Romo, 2010). Cependant, au cours des dix dernières années, l'âge modal au décès chez les adultes a pris une place prépondérante au sein des études sur la longévité humaine (Canudas-Romo, 2008, 2010; Cheung et Robine, 2007; Cheung *et al.*, 2005, 2008, 2009; Kannisto, 2000, 2001, 2007; Paccaud *et al.*, 1998; Robine, 2001; Thatcher *et al.*, 2010). En effet, l'âge modal au décès s'avère particulièrement adapté à l'étude de la longévité des adultes puisqu'il nous informe sur la durée de vie la plus commune de ces derniers. Plus encore, contrairement à l'espérance de vie à la naissance, l'âge modal au décès est seulement influencé par les conditions de mortalité à l'âge adulte ; il est ainsi nettement plus sensible aux changements qui surviennent au sein de cette sous-population (Kannisto, 2001; Horiuchi, 2003).

Dès la fin du XIX^e siècle, le *concept des durées de vie normales* émergeant des travaux précurseurs du statisticien allemand Wilhelm Lexis (1877, 1878) identifiait l'âge modal au décès comme étant la caractéristique la plus *centrale* et *naturelle* de la longévité humaine (voir figure 1.1). D'après Lexis, les décès se produisant à cet âge et au-delà pouvaient être vus comme des décès *normaux*. De plus, par symétrie de la portion de la distribution située à la droite de l'âge modal au décès, la partie de gauche de la distribution des durées de vie *normales* pouvait être obtenue, séparant ainsi les décès prématurés (zone pointillée sur la figure 1.1) de ceux considérés normaux par Lexis (Cheung *et al.*, 2005).

FIGURE 1.1. Concept des durées de vie normales introduit par Lexis



Source : Lexis (1878)

Notes : L'axe des abscisses représente l'âge et celui des ordonnées correspond à la fréquence des décès. La zone pointillée identifie les décès prématurés.

Soulignons qu'à cette époque, les travaux de Lexis constituaient en quelque sorte l'aboutissement de plus de deux siècles de recherches sur la description de la répartition des décès selon l'âge. En effet, John Graunt (1662) avait amorcé cette exploration environ 215 années plus tôt en utilisant les bulletins de mortalité de la ville de Londres. Il fut suivi par Leibniz (1680), Halley (1693), Kersseboom (1742), Struyck (1740, 1753), Süssmilch (1765) et Deparcieux (1746, 1760), pour ne nommer que quelques-uns des contributeurs influents en la matière. Les travaux de ces pères fondateurs de la science démographique seront d'ailleurs abordés dans

le cadre du troisième article de cette thèse (voir précisément la section 5.2). Lexis s'appuya pour sa part principalement sur les travaux de Quetelet (1835, 1871) concernant *l'homme moyen* et sur la loi de la distribution normale des erreurs développée par Laplace (1812) pour formuler son concept d'une durée normale de la vie humaine (Véron et Rohrbasser, 2003).

La contribution de Lexis fut d'emblée reçue très favorablement par Bertillon (1878), de même que par d'autres statisticiens et économistes vers la fin du XIX^e siècle (Véron et Rohrbasser, 2003). Cependant, elle passa par la suite pratiquement inaperçue pendant à peu près un siècle, soit jusqu'à ce que quelques chercheurs, à commencer par Kannisto en 2000, démontrent son utilité pour mesurer les changements dans la répartition des décès selon l'âge, en particulier aux grands âges. La présentation des résultats de ces travaux articulés autour de l'âge modal au décès (Canudas-Romo, 2008; Cheung et Robine, 2007; Cheung *et al.*, 2008, 2009; Robine *et al.*, 2006; Thatcher *et al.*, 2010) est effectuée à la sous-section 1.1.4; celle-ci requiert d'abord que nous nous attardions aux indicateurs visant à résumer la dispersion des décès selon l'âge.

1.1.3.2. *Mesures de dispersion*

Quelle est l'ampleur de l'hétérogénéité des durées de vie? Voilà une autre question très pertinente en démographie qui a heureusement gagné en popularité ces dernières années, notamment grâce aux travaux de Wilmoth et Horiuchi (1999). Ces derniers ont d'abord analysé dix indicateurs résumant la variabilité des âges au décès, incluant ceux qui avaient été utilisés précédemment par Keyfitz (1977), Myers et Manton (1984a), Eakin et Witten (1995) et Nusselder et Mackenbach (1996), dans le but de réduire la confusion qui régnait à ce sujet. À l'aide de données sur la Suède, le Japon et les États-Unis, Wilmoth et Horiuchi ont démontré que les résultats obtenus pour les dix indicateurs retenus étaient fortement corrélés entre eux. À des fins pratiques, ils ont donc recommandé l'utilisation de l'écart interquartile des âges au décès, correspondant à la différence entre le premier quartile et le troisième quartile de la distribution des âges au décès, pour mesurer l'hétérogénéité des durées de vie. Sur la base de cet indicateur,

les auteurs ont pu constater que, depuis le milieu du XX^e siècle, cette hétérogénéité avait cessé de diminuer dans les trois pays à l'étude, malgré la poursuite de la baisse de la mortalité. Le phénomène de rectangularisation de la courbe de survie n'était donc plus pertinent.

En effet, la baisse importante de la mortalité infanto-juvénile, ainsi que celle chez les jeunes adultes, observée dans la plupart des pays à faible mortalité jusqu'au milieu du XX^e siècle, a provoqué une réduction marquée de l'hétérogénéité de l'*ensemble* des durées de vie. En d'autres termes, le phénomène de *compression globale de la mortalité* s'est avéré considérable jusqu'aux années 1950 dans ces pays et la forme de leurs courbes de survie respectives est devenue plus rectangulaire. Par la suite, même si des gains systématiques en matière de mortalité chez les personnes âgées ont alors commencé à être enregistrés (Kannisto *et al.*, 1994; Jeune et Vaupel, 1995), la compression globale de la mortalité s'est essouffée, de même que la rectangularisation de la courbe de survie. Ces constats ont donc conduit les chercheurs à faire la distinction entre le phénomène de compression globale de la mortalité, qui concerne l'ensemble des durées de vie, et celui de *compression de la mortalité aux grands âges*, basé sur les décès survenant aux âges avancés. Similairement, les changements dans la répartition des décès selon l'âge, tous âges considérés, sont analysés séparément de ceux portant spécifiquement sur les âges avancés (Thatcher *et al.*, 2010).

Depuis les travaux de Wilmoth et Horiuchi (1999), quelques nouveaux indicateurs de dispersion des âges au décès concernant l'ensemble des durées de vie, ou encore un éventail d'âges considérable, ont été proposés. Mentionnons d'abord l'indicateur S_{10} (Edwards et Tuljapurkar, 2005), qui correspond à l'écart type des âges au décès au-delà de l'âge de 10 ans. Celui-ci a été utilisé par Edwards et Tuljapurkar afin d'étudier la convergence des niveaux d'hétérogénéité des durées de vie au sein des pays industrialisés. Pour apprécier l'évolution de la concentration des décès autour de l'âge modal au décès (M) à travers le temps, Canudas-Romo (2008) recommanda pour sa part l'indicateur SDM , soit l'écart type, relatif à M , des âges au décès. Enfin, Zhang et Vaupel (2009) ont récemment suggéré

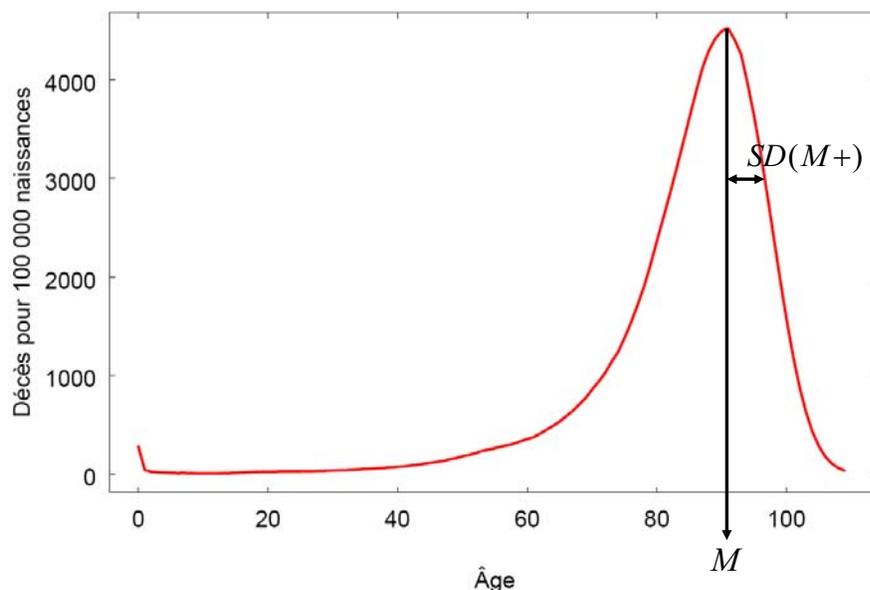
e^\dagger , qui représente le nombre moyen d'années de vie compromises en raison des décès survenus précocement, afin d'analyser la dynamique du phénomène de dispersion des âges au décès dans le modèle de la table de mortalité. Soulignons que ces trois indicateurs, S_{10} , SDM et e^\dagger , s'avèrent fortement corrélés entre eux, de même qu'avec ceux étudiés précédemment par Wilmoth et Horiuchi (1999) (Vaupel *et al.*, 2009). Tout dépendant du contexte de l'étude, pour des questions d'interprétation, d'explication et/ou de propriétés mathématiques, l'un ou l'autre de ces indicateurs peut cependant s'avérer davantage pertinent.

Dans le but de mesurer la dispersion des durées de vie parmi les personnes âgées, Kannisto (2000) proposa notamment d'utiliser l'écart type, relatif à l'âge modal au décès (M), des durées de vie situées au-delà du mode ($SD(M+)$) (voir figure 1.2). Cet indicateur est inspiré des travaux de Lexis (1877, 1878) et il repose exclusivement sur les décès survenus au-delà de l'âge modal. Il est donc peu susceptible d'être influencé par les changements relatifs à l'ampleur de la mortalité infanto-juvénile et prématurée au fil du temps. Cet indicateur, conjugué avec l'âge modal au décès, a été utilisé dans plusieurs études ultérieures afin de surveiller les transformations dans la répartition des décès aux grands âges dans une variété de pays à faible mortalité (Cheung et Robine, 2007; Cheung *et al.*, 2008, 2009; Kannisto, 2001, 2007; Thatcher *et al.*, 2010). Axés sur la longévité des adultes, ces travaux visent plus spécifiquement à déterminer si la compression de la mortalité aux grands âges poursuivra son cours, aussi longtemps que l'âge le plus commun au décès augmentera. En effet, un tel scénario pourrait être interprété comme une indication d'une résistance accrue face à la prolongation de la vie humaine (Kannisto, 2001). Les principaux résultats de ces recherches sont présentés à la sous-section suivante.

1.1.4. Résultats des études articulées autour de l'âge modal au décès

Dans leur étude portant sur le Japon pour la période 1950-2004, Cheung et Robine (2007) constatent que l'âge modal au décès (M) a augmenté sans relâche durant ces années. Cette hausse marquée de la durée de vie la plus commune s'est accompagnée d'une diminution de l'écart type des durées de vie situées au-delà de

FIGURE 1.2. Illustration de l'âge modal au décès (M) et de l'écart type, relatif à M , des durées de vie situées au-delà de M ($SD(M+)$), femmes, Japon, 2000-2004



Source : Human Mortality Database (2010)

l'âge modal au décès ($SD(M+)$) pendant plusieurs décennies. Cependant, depuis le milieu des années 1980 chez les femmes et le début des années 1990 chez les hommes, $SD(M+)$ est presque demeuré inchangé. Ainsi, à la lumière des résultats provenant des quinze dernières années au Japon, l'augmentation de l'âge le plus commun au décès se poursuit, mais la compression de la mortalité aux grands âges semble s'être essoufflée considérablement. Un déplacement de l'ensemble des durées de vie adultes vers des âges plus élevés, sans réduction de leur dispersion, aurait donc lieu depuis plus d'une décennie dans ce pays. Cette nouvelle ère serait d'ailleurs en conformité avec le *shifting mortality scenario*, décrit par Kannisto (1996) ainsi que Bongaarts et Feeney (2002, 2003) et développé davantage par Bongaarts (2005). En effet, d'après Bongaarts (2005), le *shifting mortality scenario* correspond à un régime où la mortalité adulte est décalée vers des âges progressivement plus avancés au fil du temps, sans toutefois que le profil selon l'âge soit modifié. Soulignons que la paire d'indicateurs $d(M)$ et $d(M+)$ (Kannisto, 2001), où $d(M)$ correspond au nombre de décès survenant à l'âge modal et $d(M+)$ au nombre cumulé de décès depuis l'âge modal, a également été utilisée

par Cheung et Robine (2007) et mena au même constat. Les auteurs ont donc signalé que ces résultats obtenus pour le Japon ne suggéraient aucunement que nous approchions d'une limite en matière de longévité humaine.

Les travaux subséquents réalisés par Cheung et ses collaborateurs (2008, 2009) nous informent que d'autres pays à faible mortalité affichent des résultats similaires à ceux du Japon. En effet, depuis le début des années 1990 en France et en Italie, M poursuit son ascension, mais le rythme de diminution de $SD(M+)$ a nettement ralenti (Cheung *et al.*, 2008). Il en va de même pour la Suisse depuis le milieu des années 1990 (Cheung *et al.*, 2009). Ces pays seraient donc également en cours de transition entre le régime de compression de la mortalité aux grands âges et le *shifting mortality scenario*. Le Japon semble toutefois être le plus avancé de ces pays dans cette transition. De toute évidence, ces développements récents méritent d'être surveillés de près puisqu'une concrétisation du *shifting mortality scenario* dans le futur « démultiplierait l'augmentation du nombre des personnes très âgées comme celui des centenaires mais aussi des personnes âgées de 105 ans ou de 110 ans » (Robine et Cheung, 2009, p.941). Dans le cadre des articles de cette thèse, nous proposerons une nouvelle approche, introduite à la section 1.3 ultérieure, qui permet d'effectuer un suivi très précis de l'évolution de M et $SD(M+)$.

En plus d'avoir été analysés sous une approche empirique, l'âge modal au décès et les indicateurs lui étant associés ont aussi été étudiés d'un point de vue davantage théorique (Canudas-Romo, 2008; Robine *et al.*, 2006; Thatcher *et al.*, 2010). Ainsi, la contribution de Canudas-Romo (2008) vise spécifiquement à étudier le *shifting mortality scenario* sous l'angle de l'âge modal au décès en utilisant diverses *lois* (modèles) de mortalité. Mentionnons que la sous-section 1.2.1 ci-dessous aborde les principaux modèles utilisés en démographie notamment, pour décrire l'évolution de la mortalité en fonction de l'âge. Les résultats obtenus à l'aide des différents modèles de mortalité retenus par Canudas-Romo suggèrent que le *shifting mortality scenario* constitue une avenue bien réaliste pour les pays

à faible mortalité, si les changements récents en matière de mortalité se maintiennent dans les années à venir. En effet, au sein de ces modèles, le *shifting mortality scenario* fait en sorte que l'âge modal au décès ne cesse d'augmenter au fil du temps et les mesures de dispersion y étant associées tendent vers une valeur constante, rejoignant ainsi les résultats empiriques décrits précédemment (Cheung et Robine, 2007; Cheung *et al.*, 2008, 2009).

Les travaux de Thatcher *et al.* (2010), inspirés entre autres de ceux de Robine *et al.* (2006), portent quant à eux sur la dynamique sous-jacente au phénomène de compression de la mortalité aux grands âges. Plus spécifiquement, les auteurs abordent ce phénomène à l'aide d'un modèle logistique simplifié (voir sous-section 1.2.1, équation (1.2)) reconnu pour sa représentation satisfaisante de la mortalité aux grands âges (Himes *et al.*, 1994; Kannisto, 1992; Thatcher *et al.*, 1998; Thatcher, 1999). Les résultats de cette étude révèlent que les durées de vie situées au-delà de l'âge modal au décès ont eu tendance à se concentrer dans un intervalle d'âge progressivement plus court au fil du temps puisque les taux de mortalité aux très grands âges ont diminué moins rapidement que ceux aux grands âges. Précisons que pour leur démonstration, Thatcher et ses collaborateurs (2010) ont utilisé les taux de mortalité à 90 ans et à 70 ans respectivement. Étant donné que la compression de la mortalité aux grands âges semble s'être essoufflée ces dernières années dans quelques pays à faible mortalité, les auteurs soulignent que nous ignorons encore jusqu'où diminueront les indicateurs de dispersion de la mortalité aux grands âges, en particulier $SD(M+)$.

1.2. LOIS ET MODÈLES DE MORTALITÉ

Les *lois de mortalité* sont largement utilisées en démographie et dans plusieurs autres domaines. À titre d'exemple d'application, elles interviennent souvent dans la construction des tables de mortalité. Ces lois visent à représenter la trajectoire de la mortalité selon l'âge à l'aide de modèles mathématiques. Ainsi, elles décrivent également la manière dont les décès se répartissent selon l'âge. Dans la présente section, nous abordons en premier lieu les principaux modèles proposés pour représenter le profil de la mortalité selon l'âge. Précisons que la contribution

de Gavrilov et Gavrilova (1991) procède à un examen complet de ces modèles, alors que Forfar *et al.* (1988) offrent une synthèse exhaustive des méthodes utilisées dans le domaine de l'actuariat. En second lieu, les modèles traduisant le profil de la mortalité selon l'âge et les années de calendrier sont brièvement examinés.

1.2.1. Le profil de la mortalité selon l'âge

Apportons d'emblée quelques précisions relatives aux concepts, à la notation et à la terminologie utilisés dans les pages qui suivent. Nous dénotons le *quotient de mortalité* par q_x et celui-ci correspond à la probabilité de décéder entre les âges exacts x et $x + 1$. Le *taux de mortalité* à l'âge de x ans révolus est pour sa part désigné par m_x . Enfin, la *force de mortalité*, aussi connue sous l'appellation *taux instantané de mortalité*, est notée $\mu(x)$. Ces trois quantités sont étroitement liées et des relations approximatives permettent de passer aisément de l'une aux deux autres (Thatcher *et al.*, 1998, Appendix A).

La recherche de lois de mortalité décrivant l'évolution de la mortalité selon l'âge remonte à bien longtemps, soit d'aussi loin qu'à l'époque d'Abraham De Moivre et Isaac de Graaf au XVIII^e siècle, et même jusqu'aux travaux de Graunt datant de 1662 (Dupâquier, 1996; Hald, 1990). Ce fut cependant Gompertz qui, pour la première fois en 1825, formula un modèle qui en décrivait le comportement sur un éventail d'âges considérable. La loi de Gompertz stipule qu'après l'adolescence jusqu'avant d'atteindre de très grands âges, c'est-à-dire entre l'âge de 30 et 80 ans environ, la force de mortalité évolue de manière exponentielle selon l'âge. Ainsi,

$$\mu(x) = ae^{bx}, \quad (1.1)$$

où le paramètre a représente le niveau de base de la mortalité, soit à l'âge de 30 ans environ, et b correspond au rythme d'accroissement de la mortalité avec l'âge. Pendant près de 170 années, la loi de Gompertz fut appliquée aux données de mortalité de nombreux pays et constamment reconnue pour sa bonne performance lorsque limitée aux âges prescrits (Thatcher *et al.*, 1998).

De nombreuses tentatives de généralisation du modèle de Gompertz furent proposées. Tout d'abord, Makeham (1860) suggéra l'ajout d'un terme constant c à l'équation (1.1), afin de tenir compte de l'existence de causes de décès qui ne sont aucunement liées à l'âge, d'où

$$\mu(x) = c + ae^{bx}.$$

Si le modèle de Makeham ne compromettrait pas la simplicité de celui de Gompertz puisqu'il visait également à représenter la mortalité qu'entre 30 et 80 ans environ, il en fut tout autrement pour d'autres ajustements qui aspiraient à modéliser le profil de la mortalité sur une étendue d'âges plus importante. À titre d'exemple, Le Bras (1976) a souligné que pour décrire la mortalité depuis la naissance, Steffensen avait proposé d'utiliser

$$\mu(x) = c + k_1 \frac{e^{\alpha\sqrt{x}}}{\sqrt{x}} + ae^{bx},$$

alors que Risser avait plutôt suggéré le modèle suivant

$$\mu(x) = c + ke^{a_0 + a_1x + a_2x^2 + bx + \dots + a_nx^n}.$$

Parmi les auteurs ayant développé des modèles couvrant l'ensemble des durées de vie, citons Heligman et Pollard (1980), qui recommandèrent le modèle descriptif correspondant à

$$\mu(x) = A^{(x+B)^C} + De^{-E(\ln x - \ln F)^2} + \frac{GH^x}{1 + GH^x},$$

comportant huit paramètres au total. D'une part, l'estimation des paramètres d'un tel modèle s'avère beaucoup plus ardue que celles des modèles de Gompertz et Makeham. D'autre part, il est difficile de départager la signification contextuelle de chaque paramètre afin de les interpréter.

En effet, la mortalité chez les humains aux différents âges de la vie comporte un certain nombre de particularités : d'abord élevée à la naissance (i.e. mortalité infantile), la mortalité chute ensuite drastiquement durant les premières années de la vie ; les jeunes adultes affichent habituellement une surmortalité comparativement à leurs proches aînés (i.e. mortalité accidentelle) ; on observe généralement

une décélération de la mortalité aux grands âges. La formulation d'un modèle pouvant cerner correctement la mortalité à l'âge adulte ainsi qu'à plusieurs autres âges n'est pas aisée.

En guise de représentation de la trajectoire de la mortalité aux grands âges, le modèle logistique, initialement proposé par Perks (1932), puis étudié par Beard (1959, 1971), Le Bras (1976) et Vaupel *et al.* (1979) est souvent utilisé. En effet, ce modèle admet une décélération de la force de mortalité aux très grands âges et il peut être exprimé comme suit

$$\mu(x) = c + \frac{ae^{bx}}{1 + \alpha e^{bx}}.$$

Parmi une gamme de modèles de mortalité, incluant ceux de Gompertz et de Helligman et Pollard décrits précédemment, le modèle logistique s'est imposé, d'un point de vue empirique, comme étant celui qui représentait le mieux les données observées sur la mortalité des personnes âgées de 80 ans et plus (Thatcher *et al.*, 1998). Une forme simplifiée de ce modèle logistique, étudiée par Kannisto (1992) et Himes *et al.* (1994), correspondant à

$$\mu(x) = \frac{ae^{bx}}{1 + ae^{bx}} \tag{1.2}$$

offrait également une très bonne représentation de ces données. Les théories avancées par les différents chercheurs pour expliquer la décélération de la mortalité aux grands âges reposent sur le concept d'*hétérogénéité de la population* (Beard, 1971; Le Bras, 1976; Vaupel *et al.*, 1979). Pris à âge égal, pour une multitude de facteurs explicatifs, certains individus sont plus fragiles que d'autres et décèdent donc plus précocement, d'où l'expression *hétérogénéité des durées de vie* utilisée plus tôt dans ce chapitre. Par conséquent, à mesure qu'on avance en âge, une sélection des individus les plus robustes s'effectue naturellement. Cette sélectivité différentielle peut faire en sorte que les trajectoires de mortalité selon l'âge pour la population s'avèrent étonnamment différentes des trajectoires pour les divers sous-groupes d'individus qui la composent (Vaupel et Yashin, 1985).

1.2.2. Le profil de la mortalité selon l'âge et les années de calendrier

Les modèles présentés à la sous-section précédente offrent des représentations de l'évolution de la mortalité selon l'âge exclusivement. Plusieurs auteurs se sont pour leur part attardés au développement de modèles décrivant le profil de la mortalité selon l'âge et les années de calendrier.

Parmi les principales approches disponibles, mentionnons les modèles âge-période-cohorte (APC), utilisés pour étudier les variations des taux de mortalité selon trois dimensions fondamentales en démographie : l'âge, la période et la cohorte. Plus précisément, les analyses de type APC visent à départager l'effet propre à chacune de ces dimensions. Ainsi, le modèle APC s'écrit

$$\ln(m_{ijk}) = \lambda + \alpha_i + \beta_j + \gamma_k + \epsilon_{ijk}, \quad (1.3)$$

où α_i , β_j et γ_k correspondent aux effets de l'âge, la période et la cohorte respectivement. Le paramètre λ établit quant à lui un niveau général de la mortalité et ϵ_{ijk} représente le terme d'erreur. Bien que le modèle APC apparaisse simple à l'écriture, l'estimation de ses paramètres s'avère une tâche complexe, principalement parce que les dimensions âge, période et cohorte sont linéairement dépendantes : $k + i = j$. Ce problème gêne également l'interprétation des paramètres estimés. Clayton et Schifflers (1987) et Wilmoth (2001) ont procédé à un examen plus détaillé des modèles APC et par le fait même incité les utilisateurs de ces modèles à demeurer prudents lors de l'interprétation des résultats obtenus. Les développements récents au sujet de l'approche APC sont notamment proposés par Carstensen (2007).

Une approche plus simple que celle des modèles APC et qui s'avère extrêmement populaire en démographie est celle développée par Ronald Lee et Lawrence Carter (1992). Afin de modéliser le logarithme des taux de mortalité selon l'âge i et les années de calendrier j , Lee et Carter ont proposé d'utiliser

$$\ln(m_{ij}) = \alpha_i + \beta_i \gamma_j + \epsilon_{ij}, \quad (1.4)$$

où les paramètres α_i décrivent le profil moyen de la mortalité selon l'âge, les β_i décrivent le rythme de changement de la mortalité aux divers âges et les γ_j représentent le niveau de la mortalité pour les différentes années de calendrier à l'étude. Le terme d'erreur du modèle est dénoté par ϵ_{ij} . L'avantage du modèle Lee-Carter tient du fait qu'une fois estimés, les paramètres de celui-ci peuvent être utilisés à des fins de projection de mortalité. Pour estimer les nombreux paramètres de leur modèle, Lee et Carter avaient d'abord utilisé la méthode de décomposition en valeurs singulières (Good, 1969) et fixé une contrainte sur les paramètres α_i . Plusieurs variantes ont ensuite été proposées, notamment par Wilmoth (1993), Brouhns *et al.* (2002), Booth *et al.* (2002), Renshaw et Haberman (2003a,b, 2006) et Czado *et al.* (2005). Ces variantes ont contribué, chacune à leur manière, à apporter certaines améliorations au modèle original de Lee-Carter, qui comporte diverses limites et inconvénients (Paquette, 2006, sous-section 1.2.1). Cependant, jusqu'à ce jour, l'absence de consensus quant à la variante à privilégier en pratique persiste.

De manière générale, ajoutons que les approches APC et Lee-Carter imposent une structure de modélisation rigide aux données observées et ne prennent donc pas avantage de la régularité du profil de la mortalité selon l'âge et les années de calendrier. En effet, les équations (1.3) et (1.4) décrivant les modèles APC et Lee-Carter respectivement nous indiquent que ceux-ci comportent autant de paramètres à estimer qu'il y a d'années d'âge et de calendrier (et potentiellement de cohortes) sous étude (Camarda, 2008). L'approche par *P-splines* développée par Eilers et Marx (1996), une méthode de lissage *non paramétrique* émergente et bien prometteuse en sciences appliquées, est quant à elle beaucoup plus souple et parcimonieuse. Elle constitue ainsi une alternative fort intéressante pour l'étude de la mortalité (Camarda, 2008; Currie *et al.*, 2004, 2006; Eilers *et al.*, 2006). Nous l'abordons à la prochaine sous-section et expliquons en quoi son utilisation est pertinente dans le cadre de nos travaux sur les changements dans la répartition des décès selon l'âge chez les adultes.

1.3. DES MODÈLES PARAMÉTRIQUES AUX MÉTHODES DE LISSAGE PLUS SOUPLES

1.3.1. La percée des P-splines

Dans les pays et régions peuplés qui disposent d'un système d'enregistrement des événements démographiques exhaustif et fiable, les trajectoires de mortalité selon l'âge et/ou les années de calendrier décrites par les données observées sont habituellement d'une régularité impressionnante. Les méthodes de lissage souples, qui permettent un relâchement des hypothèses relatives à la structure *attendue* de ces trajectoires, s'imposent ainsi tout naturellement dans le domaine de l'analyse de la mortalité. Elles constituent un excellent complément aux lois et modèles de mortalité décrits à la section précédente.

Bien que la littérature au sujet des méthodes de lissage souples était déjà abondante (Cleveland, 1979; de Boor, 1978; Eubank, 1988; Härdle, 1990; Silverman, 1986; Odgen, 1997), Eilers et Marx (1996) ont tout de même introduit récemment une nouvelle approche reposant sur les P-splines. Cela porta fruit puisque depuis lors, les P-splines jouissent d'une popularité grandissante dans plusieurs champs de recherche appliquée, dont celui de la mortalité (Camarda, 2008; Currie *et al.*, 2004, 2006; Eilers *et al.*, 2006). Une description détaillée de cette approche est présentée au prochain chapitre (voir section 2.2.1). Pour l'instant, précisons que l'objectif des P-splines est double : obtenir une représentation fidèle des données observées, tout en s'assurant que cette représentation soit lisse. La flexibilité de cet outil de modélisation, combinée à sa simplicité d'utilisation, d'implémentation et de compréhension, expliquent sans aucun doute la grande importance que les chercheurs lui ont accordée jusqu'à maintenant. Mentionnons que l'approche de lissage par P-splines est souvent qualifiée de *non paramétrique* puisque les paramètres des modèles P-splines n'ont pas de valeur d'interprétation.

Parmi les auteurs ayant précisément eu recours aux P-splines pour étudier la mortalité, citons d'abord Currie *et al.* (2004), qui ont démontré la pertinence de cette approche pour effectuer des projections de mortalité. Les auteurs ont

illustré leurs propos à l'aide de données provenant de l'industrie de l'assurance en Grande-Bretagne et ont également procédé à une comparaison des méthodes P-splines et Lee-Carter. Les résultats indiquent que l'approche par P-splines est largement plus parcimonieuse et qu'elle offre un rendement équivalent, sinon supérieur à la méthode Lee-Carter. Dans le cadre de sa thèse de doctorat, Camarda (2008) a pour sa part procédé à une analyse approfondie des modèles P-splines et développé un ensemble de mesures d'ajustement de ces modèles adaptées à l'étude de la mortalité au sein de populations de taille considérable. L'auteur a également illustré l'utilité des P-splines pour la correction de données sur les décès comportant des erreurs de déclarations. Bref, la pertinence des P-splines dans le domaine de la démographie, particulièrement pour l'examen de la mortalité, est déjà bien établie.

1.3.2. Objectifs et questions de recherche

Dans le cadre de notre recherche portant sur les changements dans la répartition des âges au décès, les méthodes de lissage par P-splines constituent une excellente avenue de recherche qui mérite grandement d'être explorée. En effet, jusqu'à présent, l'estimation des indicateurs résumant la tendance centrale et la dispersion des âges au décès, en particulier l'âge modal au décès, s'est déroulée dans un contexte de modélisation plutôt rigide. En effet, le calcul de ces indicateurs requiert habituellement l'utilisation de méthodes statistiques paramétriques qui, contrairement à l'approche par P-splines, imposent une structure prédéterminée aux données qui s'avère souvent discutable (voir les sous-sections 3.2.1 et 3.2.2). Il s'agit d'ailleurs d'une limite qui a été soulevée par Cheung et Robine (2007) et Cheung *et al.* (2005, 2008) puisque cette pratique est susceptible d'exercer une influence sur l'estimation des indicateurs. Offrant davantage de souplesse que les méthodes paramétriques, l'approche par P-splines nous permettra d'identifier plus précisément les trajectoires de mortalité au niveau de la population. Elle entraînera ainsi un meilleur suivi de l'évolution de l'âge modal au décès et des mesures de dispersion associées.

La présente thèse vise donc à proposer une nouvelle approche de lissage non paramétrique souple reposant sur les P-splines adaptée au suivi fin et détaillé des changements survenus dans la répartition des décès selon l'âge, en particulier aux âges avancés. Celle-ci sera utilisée dans les divers articles de cette thèse qui abordent, chacun à leur manière, le thème de la longévité des adultes et tentent de répondre aux grandes questions suivantes : Une augmentation de la durée de vie modale conduit-elle nécessairement à une réduction de l'hétérogénéité des durées de vie chez les adultes ? Est-ce qu'il y a une limite à l'augmentation de la durée de vie modale ? Existe-t-il un niveau minimal d'hétérogénéité des durées de vie pouvant être atteint au sein d'une population ? La dynamique des changements dans la répartition des décès selon l'âge chez les adultes s'avère-t-elle uniforme à travers les régions d'un même pays ? Quelle était la durée de vie la plus commune des adultes avant le XX^e et comment celle-ci a-t-elle évolué au sein des populations du passé ?

Chapitre 2

Sources de données et méthodes d'analyse

Ce second chapitre vise d'abord à présenter de manière détaillée les différentes sources de données utilisées dans les trois articles de recherche de cette thèse. Nous exposons ensuite les méthodes privilégiées pour analyser ces données.

2.1. SOURCES ET QUALITÉ DES DONNÉES

Chacun des articles de cette thèse repose sur sa propre source de données. Ainsi, en respectant l'ordre de présentation des articles aux chapitres 3, 4 et 5, nous introduisons dans un premier temps la *Human Mortality Database*, ensuite la *Base de données sur la longévité canadienne* et finalement le *Registre de la population du Québec ancien*.

2.1.1. La *Human Mortality Database*

Le principal objectif visé par la *Human Mortality Database* (HMD), une base de données internationale portant actuellement sur 37 pays et régions, est de fournir un large éventail de données aux chercheurs, étudiants, journalistes, analystes de politiques, ainsi qu'à tous ceux intéressés par l'histoire de la longévité humaine. Faisant directement suite au projet de la *Berkeley Mortality Database* (BMD) initié par John Wilmoth, professeur au Département de démographie de la *University of California, Berkeley* (UC Berkeley), celui de la HMD est le fruit d'un travail collaboratif des démographes du *Max Planck Institute for Demographic Research* (MPIDR), du Département de démographie à UC Berkeley et des différents pays qui y figurent. À ce jour, près de 300 articles scientifiques reposant

sur des données de la HMD (ou de la BMD) ont été publiés dans les grandes revues de démographie, gérontologie, actuariat, économie, médecine, épidémiologie, statistique et sociologie. Plus de 75 livres ou chapitres de livres font également usage de la HMD (ou de la BMD).

La HMD regroupe des informations détaillées sur la mortalité et la population. Ces informations sont mises à jour régulièrement, soit à chaque année pour la plupart des pays et régions, et sont accessibles gratuitement. Plus précisément, les données disponibles sont les suivantes :

- nombre annuel de naissances vivantes selon le sexe
- nombre de décès selon le sexe, l'âge, l'année de décès et la génération
- estimation de la population selon le sexe et l'âge au 1^{er} janvier de chaque année de calendrier
- estimation de la population exposée au risque de décéder selon le sexe, l'âge et l'année de calendrier (suivant l'approche transversale ou longitudinale)
- taux de mortalité selon le sexe, l'âge et l'année de calendrier (suivant l'approche transversale ou longitudinale).

Une variété de tables de mortalité (du moment ou par génération ainsi que complètes ou abrégées), établies à partir des données énumérées ci-haut, sont aussi accessibles. Ajoutons que ces tables de mortalité reposent sur un protocole méthodologique (Wilmoth *et al.*, 2007) à la fine pointe des recherches en démographie.

Dans le cadre du premier article de cette thèse présenté au chapitre 3, les données extraites de la HMD portent sur le Canada (1921 à 2007), les États-Unis (1945 à 2006), la France (1920 à 2007) et le Japon (1947 à 2008). Pour chaque pays, nous avons spécifiquement utilisé les décès et les estimations de la population exposée au risque de décéder, détaillés selon le sexe, l'âge et l'année de calendrier. Pour fins de comparaison avec les résultats de nos analyses, nous avons aussi eu recours aux tables de mortalité complètes du moment par sexe et année de calendrier.

Les données sur les décès diffusées sur la HMD pour le Canada, la France, le Japon et les États-Unis proviennent majoritairement des registres de l'état civil de ces pays. Pour les années antérieures à l'établissement d'un tel registre, les données sur les décès ont généralement été obtenues des bureaux de la statistique respectifs des pays. Plusieurs méthodes de traitement de l'information manquante ou incomplète au sujet des décès sont prévues au protocole de la HMD (Wilmoth *et al.*, 2007, p.9-15). Ces méthodes sont ainsi utilisées lorsque l'âge au décès est inconnu, la génération du décédé est non identifiée, les décès sont uniquement disponibles par groupes d'âge quinquennaux ou décennaux, ou encore lorsque les décès au-delà d'un certain âge sont réunis au sein d'un groupe d'âge ouvert. Soulignons que, suite à l'application de l'une ou de plusieurs de ces méthodes, les décès résultants ne sont généralement plus des nombres entiers ; il s'agit effectivement de nombres de décès estimés plutôt qu'observés.

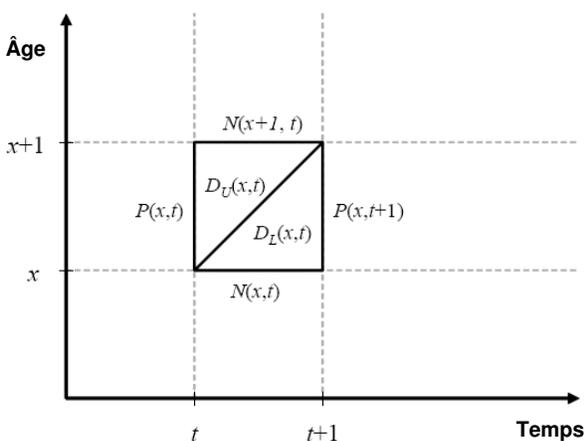
En ce qui concerne les estimations de la population exposée au risque de décéder publiées sur la HMD, celles-ci sont calculées à partir des estimations de la population au 1^{er} janvier de chaque année de calendrier et des décès. À titre d'exemple, pour un pays et un sexe donnés, la population exposée au risque de décéder âgée de x ans révolus durant l'année de calendrier t , notée E_{xt} , correspond à

$$E_{xt} = \frac{1}{2}[P(x, t) + P(x, t + 1)] + \frac{1}{6}[D_L(x, t) - D_U(x, t)], \quad (2.1)$$

où les quantités $P(x, t)$, $P(x, t + 1)$, $D_L(x, t)$ et $D_U(x, t)$ sont illustrées à la figure 2.1. La démonstration de l'équation (2.1) est effectuée à l'*Appendix E* du protocole méthodologique (Wilmoth *et al.*, 2007, p.72-77).

La figure 2.2 illustre quant à elle les différentes méthodes utilisées pour obtenir les estimations de la population au 1^{er} janvier de chaque année de calendrier, qui interviennent dans le calcul de la population exposée au risque de décéder. En deçà de l'âge de 80 ans, ce sont les estimations officielles de population, provenant généralement des instituts de la statistique des différents pays, qui sont utilisées. La méthode d'interpolation linéaire est employée lorsque ces estimations officielles sont disponibles sur une base annuelle, mais ne sont pas calculées en date du 1^{er}

FIGURE 2.1. Schéma des données impliquées dans le calcul de la population exposée au risque de décéder

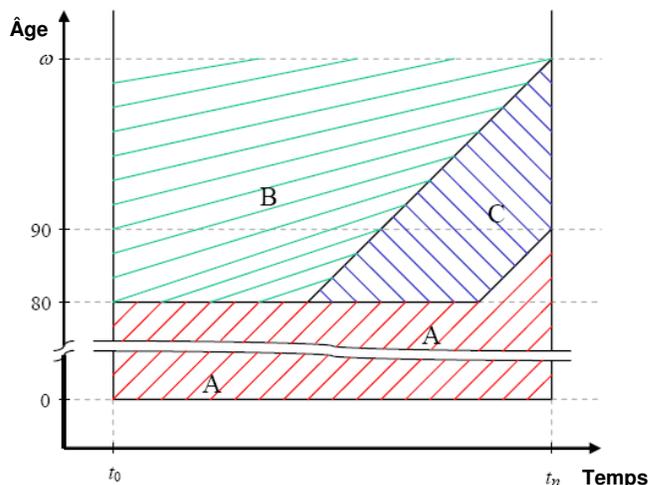


Source : Wilmoth *et al.* (2007, p.33)

janvier. Dans les cas où des estimations officielles annuelles sont non disponibles, d'autres méthodes s'appuyant notamment sur les décès observés, intitulées *inter-censal survival methods*, sont utilisées pour estimer les effectifs de la population au 1^{er} janvier de chaque année (Wilmoth *et al.*, 2007, p.16-27).

À partir de l'âge de 80 ans, les méthodes des générations éteintes et du taux de survie sont exploitées pour estimer l'effectif de la population selon le sexe et l'âge en date du 1^{er} janvier de chaque année de calendrier observée (voir figure 2.2). En effet, comme les problèmes de déclarations tendent à prendre plus d'importance au sein des populations âgées, ces méthodes exploitent les données sur les décès provenant de la statistique de l'état civil plutôt que celles tirées des recensements. Brièvement, la méthode des générations éteintes, développée par Vincent (1951), consiste à reconstituer la population à chaque âge en cumulant successivement les décès survenus au-delà de cet âge dans la génération concernée. Cette méthode n'est applicable qu'aux générations éteintes et qu'aux âges où les mouvements migratoires sont négligeables. Par définition, une génération est éteinte lorsque tous ses individus sont décédés. Ainsi, à une date t donnée, une génération est considérée éteinte si ses individus ont atteint l'âge ω , à partir duquel le nombre de décès enregistrés est virtuellement nul. Pour les générations

FIGURE 2.2. Méthodes de calcul des estimations de population



Source : Wilmoth *et al.* (2007, p.28)

Notes :

A – Estimations officielles (interpolation linéaire ; survie entre recensements)

B – Méthode des générations éteintes

C – Méthode du taux de survie

subséquentes presque éteintes, c'est-à-dire ayant au moins atteint l'âge de 90 ans à la fin de la période d'observation (voir figure 2.2), la méthode du taux de survie est appliquée. Plus spécifiquement, la méthode du taux de survie connue sous l'acronyme SR(90+) et recommandée par Thatcher *et al.* (2002) est utilisée. En résumé, pour estimer les effectifs de survivants par âge au sein des générations presque éteintes, cette méthode s'appuie sur les ratios de survivants aux divers 1^{er} janvier successifs dans les générations éteintes antérieures voisines. Pour de plus amples renseignements au sujet de la méthode des générations éteintes ou celle du taux de survie SR(90+), voir Wilmoth *et al.* (2007, p.27-32).

Dans l'ensemble, la HMD a l'avantage de réunir des données de bonne qualité puisque celles-ci ont d'abord été corrigées lorsque cela s'avérait nécessaire, puis rendues comparables, dans la mesure du possible, pour toutes les périodes observées. Les données sont effectivement soumises à plusieurs tests de qualité visant à assurer leur cohérence interne et externe. Pour ce qui est de la cohérence interne des données, des comparaisons avec des âges et années de calendrier adjacents

sont notamment effectuées. La cohérence externe des données est plutôt vérifiée par le biais de comparaisons avec des pays tels la Suède, où les données sont reconnues être de très bonne qualité.

De manière plus précise, voici ce que nous pouvons ajouter au sujet de la qualité des données dans les quatre pays analysés au chapitre 3. En premier lieu, tout porte à croire que les données pour la France disponibles sur la HMD, qui proviennent principalement de publications de l'*Institut national de la statistique et des études économiques* à la source, sont de très bonne qualité. D'après les travaux de Kannisto (1994) et Kannisto *et al.* (1994), qui procèdent à une évaluation de la qualité des données sur les décès survenus au delà de 80 ans dans 30 pays, la France s'inscrit d'ailleurs parmi ceux qui possèdent les meilleures données. Similairement, le Japon se classe aussi dans cette catégorie, quoiqu'il est précisé que la qualité des données sur les centenaires japonais avant 1960 soit moins satisfaisante que par la suite.

Au Canada, les estimations de population, basées sur le recensement et des ajustements pour le sous-dénombrement net de la population, ainsi que les données sur les décès sont publiées par *Statistique Canada* et généralement considérées comme étant exhaustives et fiables. Même chez les personnes âgées de 80 ans et plus, malgré les réserves de Kannisto et ses collègues (Kannisto, 1994; Kannisto *et al.*, 1994), les travaux de Bourbeau et Lebel (2000) attestent que la qualité des données sur les décès au Canada est bonne, du moins jusqu'à l'âge de 100 ans. L'étude de Bourbeau et Lebel (2000) souligne que les données canadiennes utilisées par Kannisto et ses collègues pour mener leurs analyses étaient insuffisantes pour qu'ils puissent se prononcer sur la qualité des données sur les décès au Canada. En effet, leurs travaux reposaient sur des données peu détaillées publiées par *Statistique Canada*, qui comportaient notamment un groupe d'âge ouvert à partir de 90 ans. Concernant la fiabilité des données sur la mortalité des centenaires, ajoutons qu'une étude récente de validation des âges au décès menée auprès de 1 900 centenaires canadiens-français de confession catholique, nés dans la province du Québec et décédés dans cette même province entre 1970 et 2007,

démontre que la qualité des données à ces âges très avancés demeure très bonne (Beaudry-Godin, 2010). Bref, parmi tous les pays figurant sur la HMD, le Canada ne fait sans doute pas partie du groupe restreint de pays où la qualité des données sur les décès est exemplaire, mais ne mérite pas non plus l'étiquette « mauvaise qualité » octroyée par Kannisto et ses collègues.

À la source, les estimations de population et les données sur les décès pour les États-Unis diffusées sur la HMD proviennent principalement de publications officielles du *US Census Bureau* et du *National Center for Health Statistics* respectivement. Parmi les 30 pays analysés par Kannisto (1994) et Kannisto *et al.* (1994), les États-Unis font également partie du sous-groupe de pays où la qualité des données sur les décès aux grands âges est faible. Les données du *Medicare Social Security Administration* sont d'ailleurs souvent utilisées en complément aux estimations de population et données de l'état civil aux grands âges ; celles-ci étant de meilleure qualité puisque les bénéficiaires de ce régime publique d'assurance santé doivent fournir une preuve d'âge au moment de leur inscription (Bell et Miller, 2005). Malgré le fait que la HMD n'exploite pas les données *Medicare* pour les États-Unis, le protocole méthodologique et les tests de cohérence interne et externe des données décrits ci-haut prévoient des mesures d'ajustement qui contribuent à améliorer la qualité des données américaines aux grands âges.

2.1.2. La Base de données sur la longévité canadienne

La *Base de données sur la longévité canadienne* (BDLC) constitue une base « satellite » de la HMD présentée ci-haut. La BDLC a effectivement été développée par l'équipe de recherche *Mortalité et longévité* du Département de démographie de l'Université de Montréal, sous la direction de Robert Bourbeau, en collaboration avec le MPIDR et le Département de démographie de UC Berkeley. Elle regroupe de l'information sur la mortalité et la population des provinces et territoires canadiens, offrant ainsi de nombreuses opportunités de recherche sur le thème des disparités régionales au Canada. Soulignons qu'à l'heure actuelle, la

BDLC est la seule HMD régionale disponible. Cependant, d'autres bases de données régionales conçues selon le modèle de la HMD pourraient bien voir le jour d'ici quelques années.

La nature des données sur la mortalité et la population diffusées sur la BDLC et la HMD est pratiquement identique, mis à part le fait que l'approche longitudinale ne soit pas offerte sur la BDLC. Ainsi, les données propres aux dix provinces et deux territoires (Territoires du Nord-Ouest et Nunavut réunis, et le Yukon) canadiens accessibles aux utilisateurs de la BDLC sont les suivantes :

- nombre annuel de naissances vivantes selon le sexe
- nombre de décès selon le sexe, l'âge, l'année de décès et la génération
- estimation de la population selon le sexe et l'âge au 1^{er} janvier de chaque année de calendrier
- estimation de la population exposée au risque de décéder selon le sexe, l'âge et l'année de calendrier (suivant l'approche transversale)
- taux de mortalité selon le sexe, l'âge et l'année de calendrier (suivant l'approche transversale),

en plus d'une variété de tables de mortalité du moment (complètes ou abrégées). Les données de la BDLC couvrent présentement la période 1921 à 2007, exception faite de Terre-Neuve-et-Labrador et des territoires pour lesquels l'information n'est accessible que depuis 1949 et 1950 respectivement. Dans les faits, Terre-Neuve-et-Labrador n'a joint officiellement la Confédération canadienne qu'au 31 mars 1949, alors que l'enregistrement des décès pour les territoires s'avère incomplet avant 1950.

Le protocole méthodologique sur lequel repose la BDLC est essentiellement le même que celui de la HMD (Wilmoth *et al.*, 2007). Le caractère régional de la BDLC requiert cependant quelques ajustements, notamment en ce qui concerne le traitement des informations manquantes sur les décès. En effet, les décès pour lesquels la province ou le territoire est inconnu ne posent pas problème lorsqu'ils sont rapportés à l'échelle nationale, comme c'est le cas sur la HMD. Pour les fins

de la BDLC, les décès dont la province ou le territoire n'est pas identifié doivent cependant être répartis, malgré qu'il y ait généralement moins d'une quinzaine de cas par année de calendrier. La procédure adoptée consiste à répartir ces décès de manière aléatoire en fonction du poids relatif des décès totaux de chaque province et territoire comparativement à l'ensemble des décès canadiens. Les détails de cette procédure sont exposés dans un document réalisé par Bourbeau *et al.* (2010). Ce document présente également l'historique complet des corrections apportées aux décès portant sur les provinces et territoires du Canada depuis 1921. Dans l'ensemble, ces corrections sont mineures et concernent moins de 1% des décès totaux à chaque année.

Le Canada compte autant de registraires de l'état civil qu'il compte de provinces et territoires. Ces bureaux de l'état civil au pays ont la responsabilité d'enregistrer les naissances, mariages et décès qui surviennent dans leur province ou territoire respectif, puis de rapporter ces informations à *Statistique Canada*, mandaté pour les publier. Les données sous-jacentes à la BDLC proviennent donc de treize administrations et il convient de s'interroger sur les divergences d'enregistrement des événements démographiques au sein de celles-ci. Précisons d'abord qu'en vertu des lois provinciales et territoriales qui commandent l'enregistrement complet des naissances, décès et mariages dans chaque province et territoire canadien, les taux de couverture sont tous très élevés. Il peut toutefois y avoir sous-dénombrement si l'enregistrement de quelques événements est transmis avec délai. Ajoutons que ces lois provinciales et territoriales sont rédigées de manière à assurer une bonne uniformité de l'enregistrement au sein des différentes administrations. Le *Conseil de la statistique de l'état civil du Canada*, un comité consultatif interprovincial/territorial, veille lui aussi à la normalisation des méthodes d'enregistrement des événements. Néanmoins, Nagnur *et al.* (1981) ont tout de même relevé certaines différences régionales quant à la qualité des statistiques de l'état civil au Canada, suggérant ainsi qu'il y a encore place à l'amélioration.

Dans le cadre du second article de cette thèse présenté au chapitre 4, les données tirées de la BDLC portent sur les dix provinces canadiennes. La période

étudiée débute en 1930, soit quelques années après les débuts de l'état civil au Canada afin d'éviter les irrégularités mineures qui auraient alors pu survenir, et prend fin en 2007. Terre-Neuve-et-Labrador fait évidemment exception puisque les données pour cette province ne débutent qu'en 1949. Les territoires canadiens ne font pas l'objet de notre étude, principalement en raison de la très petite taille de ces populations et aussi parce que leur situation sociodémographique se distingue fortement de celle des provinces canadiennes. En effet, malgré les progrès considérables réalisés en matière de mortalité au sein des territoires canadiens depuis 1950, des différences importantes persistent comparativement à l'ensemble du pays (Prud'homme, 2007; Veugelers *et al.*, 2001).

2.1.3. Le *Registre de population du Québec ancien*

Les deux sous-sections précédentes illustrent bien l'abondance de données sur la mortalité des populations récentes. Il en va cependant tout autrement pour les données qui concernent la mortalité des populations du XVIII^e siècle et antérieurement. En l'absence d'un système d'enregistrement systématique des actes de l'état civil au sein de ces populations anciennes et d'une bonne conservation des registres à travers le temps, il s'avère difficile d'étudier les phénomènes démographiques qui les touchent.

Malgré le fait que les statistiques d'état civil officielles n'ont débuté qu'en 1926 au Québec, l'enregistrement des baptêmes, mariages et sépultures par le clergé catholique remonte au XVII^e siècle. Cette pratique, instaurée par les premiers missionnaires français arrivés en Nouvelle-France durant la première moitié du XVII^e siècle, procurait de l'information sur l'état des personnes et avait donc avant tout une valeur juridique (PRDH, 2009). D'ailleurs, entre 1679 environ et ce jusqu'à la fin du XX^e siècle, l'enregistrement des naissances, mariages et décès était réalisé en deux copies; l'une d'entre elles était conservée dans les archives de la paroisse et l'autre acheminée aux autorités civiles. La population canadienne-française, pratiquement entièrement de religion catholique, a donc continuellement bénéficié d'un enregistrement de qualité de ces événements démographiques. Elle a ainsi

le potentiel d'être reconstituée exhaustivement depuis le début de la colonisation française.

Voilà qui correspond exactement au mandat du *Programme de recherche en démographie historique* (PRDH) du Département de démographie de l'Université de Montréal, un projet mis sur pied en 1966 par les professeurs Hubert Charbonneau et Jacques Légaré, aujourd'hui retraités, et maintenant sous la responsabilité de Bertrand Desjardins. Grâce au travail minutieux de dépouillement des registres paroissiaux du Québec ancien effectué par l'équipe du PRDH, ces informations sont maintenant consolidées dans une banque de données informatisée portant le nom de *Registre de la population du Québec ancien* (RPQA). À l'heure actuelle, le RPQA permet de dresser le portrait de la population du Québec depuis l'arrivée des premiers colons en 1608 jusqu'en 1799. Les sépultures de la période 1800-1850 qui concernent les personnes nées avant 1750 ont aussi été relevées par le PRDH et cette information a déjà été intégrée au RPQA. Mentionnons que le travail de reconstitution de la population canadienne-française se poursuit au sein de l'équipe du PRDH puisque l'objectif final vise une couverture du XIX^e siècle analogue à celle des XVII^e et XVIII^e siècles.

Globalement, la qualité des données du RPQA est excellente et sans équivalent au chapitre des données portant sur les populations du passé. En effet, comme le RPQA concerne tous les individus venus s'installer dans la vallée du Saint-Laurent à l'heure de la colonisation française, ainsi que leurs descendants, il a conséquemment le grand avantage de couvrir l'ensemble du territoire occupé par la population des Canadiens-français. Les problèmes de sélection et de biais qui pourraient résulter d'une couverture partielle du territoire sont ainsi éliminés (Desjardins, 1998). Les données du RPQA reposent également sur des fondations très solides; la qualité des registres a été maintes fois validée. D'après Charbonneau (1975), le clergé responsable de la tenue des registres dans les différentes paroisses de la vallée laurentienne était très instruit et cela a vraisemblablement contribué à assurer la qualité de ces derniers. Le fait que les ecclésiastiques devaient, à partir de 1679 environ, faire suivre une copie de leurs registres à chaque

année aux autorités civiles a largement restreint la disparition complète de registres, contribuant ainsi à préserver l'exhaustivité des enregistrements.

Malgré ces bonnes pratiques, un faible pourcentage de registres ont tout de même disparu au fil du temps et les actes de baptême, de mariage et de sépulture qu'ils contenaient sont donc manquants. Ces pertes, de nature aléatoire et estimées à environ 7% des années-paroisse durant la période 1608 à 1765 (Charbonneau *et al.*, 1996), toucheraient surtout « les premières années suivant la création de certaines paroisses, alors que celles-ci étaient encore peu peuplées et nouvellement établies » (Lacroix, 2009, p.26).

Le troisième article de cette thèse présenté au chapitre 5 vise notamment à apporter un nouvel éclairage sur la longévité des adultes au XVIII^e siècle et le RPQA s'avère la source idéale à consulter pour notre étude. Les données extraites du RPQA portent précisément sur tous les Canadiens-français nés et mariés sur le territoire du Québec ancien depuis les débuts de la colonie jusqu'en 1799. En raison des problèmes d'observation que posent les célibataires, nous n'avons retenu que les mariés pour notre analyse. En effet, les mariés font l'objet d'une meilleure observation que les célibataires puisqu'ils sont habituellement mieux identifiés au sein des registres paroissiaux et moins susceptibles d'effectuer des déplacements (migrations) sur de longues distances (Charbonneau, 1975; Charbonneau *et al.*, 1987). Précisons toutefois qu'au Québec ancien, puisque la très grande majorité des adultes se mariaient, seule une faible proportion d'hommes et de femmes restés célibataires de façon définitive ont été exclus.

Compte tenu de la perte aléatoire de certains registres citée ci-haut, il convient de s'interroger sur l'impact de ces pertes quant à l'enregistrement des décès de personnes mariés au sein du RPQA. Sur la base des personnes mariées avant 1800, nées avant 1750 et qui n'ont pas été identifiés comme ayant quitté le territoire du Québec ancien, nous connaissons le décès d'environ 90% d'entre eux (Lacroix, 2009). Les décès manquants, résultant principalement de la perte de certains registres, pourraient aussi correspondre à des départs non connus. Quoi qu'il en soit, cela demeure tout à fait exceptionnel dans un tel contexte historique.

2.2. MÉTHODES D'ANALYSE

La présente section aborde les diverses méthodes utilisées pour analyser les données dans les différents articles de cette thèse. Rappelons que l'approche de lissage par P-splines revêt une importance considérable au sein de nos travaux de recherche. En effet, celle-ci n'impose aucune structure prédéterminée aux données observées et mène à une expression plus précise de la mortalité. Les pages qui suivent permettront ainsi aux lecteurs de se familiariser avec cet outil, tant dans le contexte de la modélisation du profil de la mortalité selon l'âge exclusivement que selon l'âge et les années de calendrier simultanément. Les mesures de tendance centrale et de dispersion retenues pour effectuer un suivi des transformations de la mortalité aux âges avancés seront ensuite exposées.

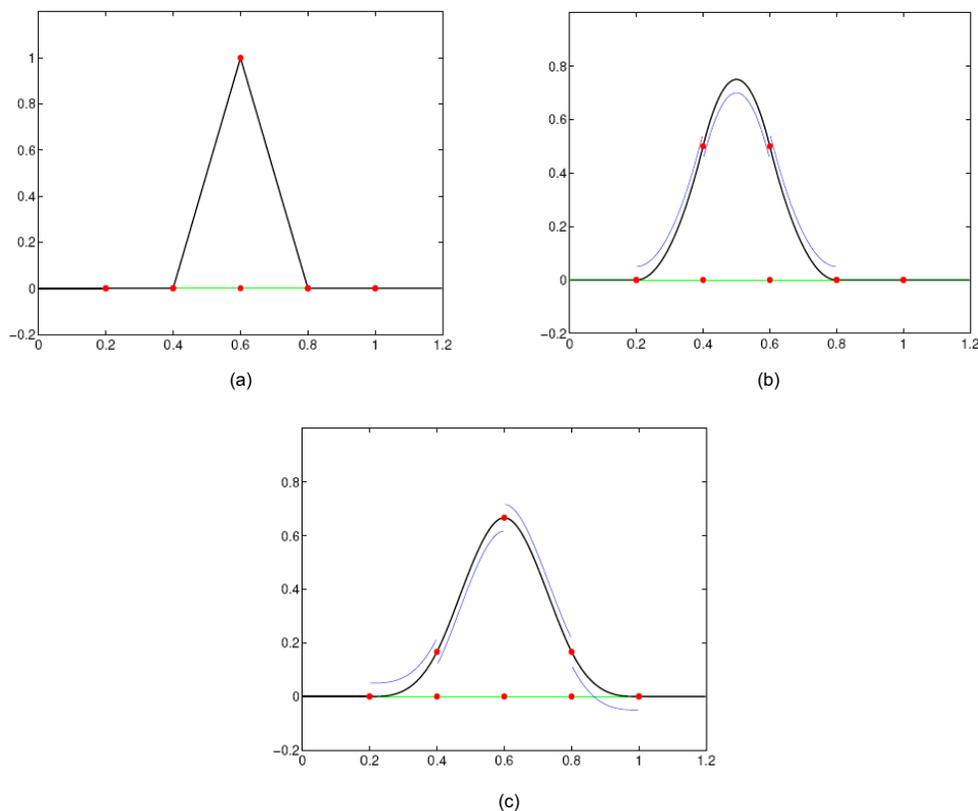
2.2.1. Lissage par P-splines

Par définition, la méthode de lissage par P-splines combine les notions de B-splines et de fonction de vraisemblance pénalisée. Brièvement, grâce aux B-splines, le processus de modélisation de données à l'aide des P-splines s'avère très flexible. En même temps, la fonction de vraisemblance pénalisée assure que l'ajustement aux données demeure lisse. Examinons ces concepts de B-splines et de fonction de vraisemblance pénalisée plus en profondeur.

2.2.1.1. *Les B-splines et les P-splines*

Les B-splines, tout comme les splines en général, sont étroitement liés aux polynômes. En réalité, les B-splines sont construits avec des parties de polynômes, joints les uns aux autres. Les divers points d'abscisse où se produisent ces jonctions sont appelées les *noeuds*. Tel qu'illustré à la figure 2.3, le degré d'un B-spline correspond au degré des parties de polynômes utilisés pour le construire. Mentionnons que le terme « B-splines » est une abréviation de l'expression « basis spline » en anglais, qui fait référence à la propriété suivante des B-splines : chaque fonction B-spline est non nulle sur une portion limitée du domaine seulement. En effet, à partir de la figure 2.3 (c), nous constatons que la fonction B-spline représentée n'est non nulle qu'entre les valeurs d'abscisse 0,2 et 1.

FIGURE 2.3. B-splines (en noir) avec noeuds (en rouge) et parties de polynômes (en bleu) : (a) linéaire, (b) quadratique, (c) cubique

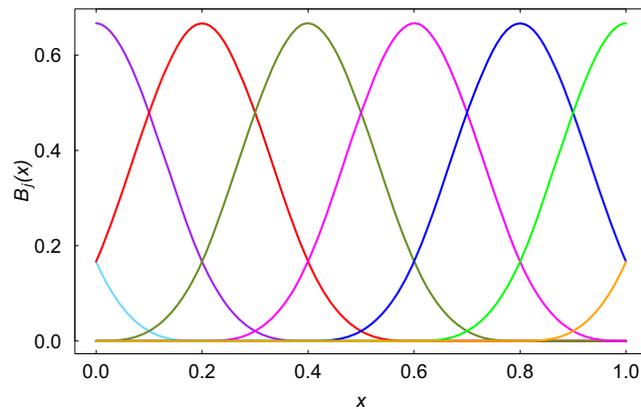


Source : Eilers et Marx (2009)

Un ensemble de plusieurs B-splines porte le nom de *base B-spline*. À titre d'exemple, la figure 2.4 montre une base B-spline composée de huit B-splines cubiques sur le domaine $x \in [0, 1]$. Comme chaque fonction B-spline est non nulle que sur une portion limitée du domaine de x , il en résulte qu'en tout point d'abscisse situé à l'intérieur de ce domaine, seules quelques fonctions B-splines $B_j(x)$, $j = 1, \dots, 8$, sont non nulles. Voilà qui explique en partie la grande popularité des bases B-splines à titre d'outils de modélisation : elles offrent un excellent contrôle *local* lors du lissage de données. Précisons que plus le nombre de noeuds sur le domaine de modélisation est élevé, plus le nombre de B-splines inclus dans la base B-spline l'est à son tour. Les huit splines cubiques formant la base B-spline

représentée à la figure 2.4 résultent de six noeuds (non représentés) répartis uniformément sur le domaine $[0,1]$. En général, l'aptitude d'une base B-spline à saisir les variations au sein des données est d'autant plus grande que le nombre de B-splines qu'elle comporte est élevé.

FIGURE 2.4. Base B-spline incluant huit B-splines cubiques avec noeuds équidistants sur le domaine de x



Source : Calculs effectués par l'auteur

À titre d'illustration du fonctionnement d'une base B-spline, considérons l'exemple suivant : nous souhaitons lisser les données, simulées aux fins de l'exemple, définies par les points (x_i, z_i) , $i = 1, \dots, 100$ sur le domaine $[0, 1]$ à l'aide de la base B-spline cubique illustrée à la figure 2.4. Plus précisément, à partir des couples (x_i, z_i) , nous désirons estimer une fonction inconnue g , présumée lisse et telle que

$$z_i = g(x_i) + \epsilon_i.$$

Dans le cas le plus simple, les erreurs ϵ_i sont indépendantes et normalement distribuées de moyenne 0 et d'écart type σ , c'est-à-dire $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$, $\forall i$. Nous verrons à la sous-section 2.2.2 que dans le contexte de données sur la mortalité, la loi de distribution des décès ne correspond pas à la loi normale mais plutôt à celle de Poisson. Quelques ajustements à la présentation ci-dessous seront ainsi apportés avant de procéder au lissage de données sur la mortalité.

La matrice \mathbf{B} associée à la base B-spline illustrée à la figure 2.4 et au jeu de données (x_i, z_i) , $i = 1, \dots, 100$ se définit comme suit

$$\mathbf{B} = \begin{pmatrix} B_1(x_1) & B_2(x_1) & \dots & B_8(x_1) \\ B_1(x_2) & B_2(x_2) & \dots & B_8(x_2) \\ \vdots & \vdots & \vdots & \vdots \\ B_1(x_{100}) & B_2(x_{100}) & \dots & B_8(x_{100}) \end{pmatrix},$$

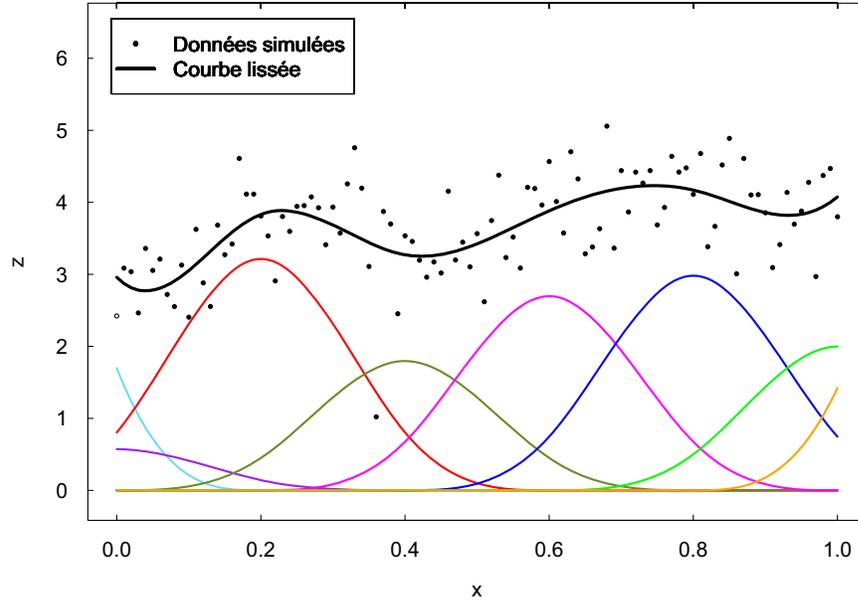
où $B_j(x_i)$, $j = 1, \dots, 8$ correspond à la valeur du j^{e} B-spline en l'abscisse x_i . La courbe lissée \hat{z} obtenue à partir des données (x_i, z_i) est illustrée à la figure 2.5. Précisément, cette courbe lissée s'écrit $\hat{z}(x_i) = \sum_{j=1}^8 \hat{\alpha}_j B_j(x_i)$, où $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_8)^\top$ est le vecteur des huit paramètres de régression estimés. Chaque paramètre de régression est donc associé à un seul B-spline dans la base B-spline considérée et il dicte la *hauteur* de ce B-spline (voir la partie inférieure de la figure 2.5). Sous forme matricielle, nous avons

$$\hat{\mathbf{z}} = \mathbf{B}\hat{\boldsymbol{\alpha}}. \quad (2.2)$$

L'équation (2.2) confirme qu'en adoptant une approche de modélisation par base B-spline, nous demeurons en terrain bien connu, soit celui de la régression classique.

Aux fins de l'exemple décrit ci-haut, nous avons d'emblée utilisé une base B-spline qui comptait huit B-splines, résultat de six noeuds équidistants initialement placés sur le domaine de modélisation $[0, 1]$. Cependant, plusieurs algorithmes ont été développés afin de déterminer le nombre optimal de noeuds à utiliser ainsi que leur positionnement sur le domaine, compte tenu des données à modéliser (Friedman et Silverman, 1989; Kooperberg et Stone, 1991, 1992). Ces tâches, essentielles à tout exercice de modélisation par base B-splines mené en bonne et due forme, restent toutefois passablement complexes. Grâce aux P-splines développés par Eilers et Marx (1996), ces tâches peuvent être évitées. En effet, l'idée derrière les P-splines consiste à utiliser un nombre relativement élevé de B-splines, conséquence de plusieurs noeuds répartis uniformément sur le domaine, tout en limitant

FIGURE 2.5. Modélisation de données (simulées) à l'aide d'une base B-spline incluant huit B-splines cubiques avec noeuds équidistants sur le domaine de x



Source : Calculs effectués par l'auteur

(i.e. pénalisant) les variations au sein des paramètres associés aux B-splines adjacents. L'approche par P-splines combine ainsi les concepts de B-splines et de fonction de vraisemblance pénalisée.

2.2.1.2. La fonction de vraisemblance pénalisée

Lorsque la méthode de lissage par P-splines est employée, le vecteur des paramètres de régression estimés $\hat{\alpha}$ dans l'équation (2.2) est obtenu en maximisant la fonction de log-vraisemblance pénalisée suivante

$$l^* = l(\alpha; \mathbf{B}; z) - \frac{1}{2} \alpha' \mathbf{P} \alpha. \quad (2.3)$$

Le premier terme de cette équation correspond à la log-vraisemblance habituelle d'un modèle de régression linéaire, alors que le second en est un de pénalité. Ce dernier assure une certaine régularité au sein du vecteur des paramètres estimés $\hat{\alpha}$; il en résulte que les valeurs des paramètres estimés associés aux B-splines adjacents ne changent pas trop brusquement. Le compromis entre la régularité

(*smoothness* en anglais) et la précision (i.e. l'adéquation aux données) du modèle est gouverné par un paramètre de lissage, inclus dans la matrice de pénalité \mathbf{P} . Plus ce paramètre de lissage est grand, plus l'importance accordée à la régularité sera élevée comparativement à la précision et vice-versa.

Le rôle du terme de pénalité dans l'équation (2.3) est illustré à la figure 2.6, à l'aide d'un exemple reposant sur de nouvelles données simulées (x_i, z_i) , $i = 1, \dots, 100$. La figure 2.6 (a) présente le résultat obtenu lorsque le terme de pénalité est omis, c'est-à-dire lorsque le paramètre de lissage est nul. Dans un tel cas, la courbe lissée est très irrégulière puisque toute l'importance est accordée à l'adéquation aux données, qui fluctuent passablement ici. En effet, en l'absence du terme de pénalité, les variations au sein des paramètres associés aux B-splines adjacents sont permises et par conséquent, la hauteur de B-splines adjacents peut changer inopinément. La figure 2.6 (b) présente une courbe lissée beaucoup plus régulière, puisque le terme de pénalité est pris en compte dans le processus d'estimation des paramètres associés aux B-splines. Tel que prévu, la pénalité fait en sorte que les paramètres estimés des B-splines adjacents évoluent de manière régulière et nous ne percevons plus de changements soudains de hauteur entre B-splines adjacents.

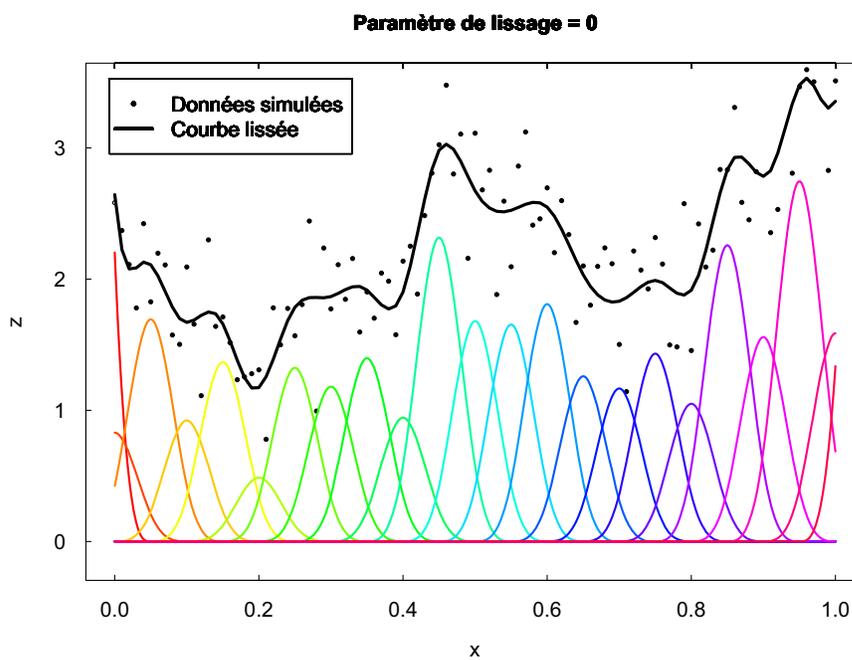
2.2.2. Les P-splines pour l'étude de la mortalité

Tel que mentionné au chapitre 1, les travaux de Camarda (2008), Currie *et al.* (2004, 2006) et Eilers *et al.* (2006) témoignent de la pertinence des P-splines pour l'étude de la mortalité. En premier lieu, examinons comment les P-splines peuvent être utilisés lorsque l'objectif visé consiste à modéliser des données sur la mortalité selon l'âge. Ces propos seront ensuite généralisés, de manière à admettre une modélisation par P-splines suivant deux dimensions prises simultanément, soient l'âge et les années de calendrier.

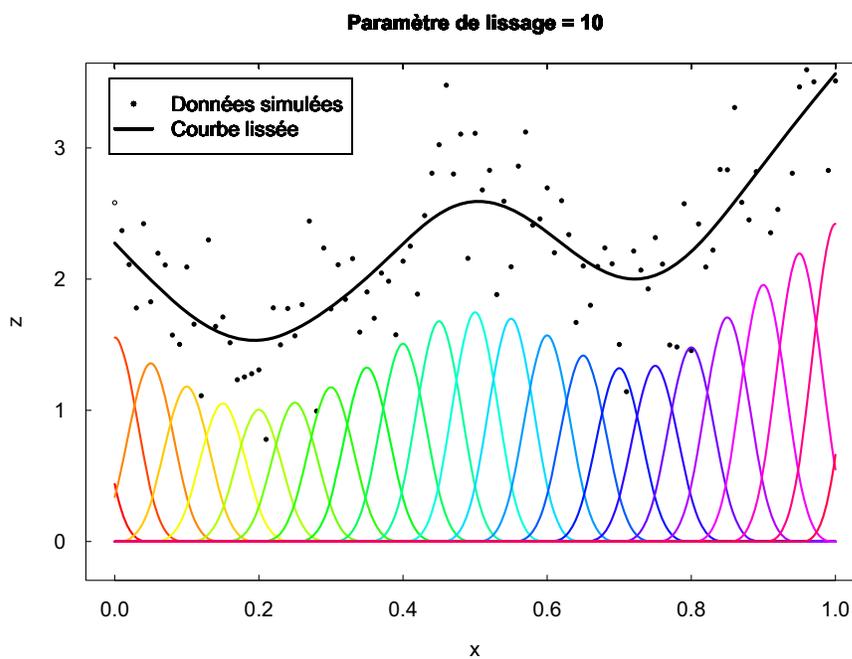
2.2.2.1. Lissage par P-splines des données selon l'âge

Pour une population prise une année de calendrier donnée, dénotons les décès observés et les effectifs de la population exposée au risque de décéder selon l'âge i

FIGURE 2.6. Régression sur données simulées : (a) non pénalisée, (b) pénalisée



(a)



(b)

Source : Calculs effectués par l'auteur

par d_i et e_i respectivement. Ainsi, les taux de mortalité selon l'âge correspondent à $m_i = d_i/e_i$ et la force de mortalité μ_i , aussi appelée le *taux instantané de mortalité*, est telle que $m_i \simeq \mu_{i+\frac{1}{2}}$ (Thatcher *et al.*, 1998, Appendix A). Sous l'hypothèse d'une force de mortalité constante à l'intérieur de chaque intervalle d'âge $[i, i+1)$, c'est-à-dire $\mu(x) = \mu_i \forall x \in [i, i+1)$, les décès observés d_i peuvent être vus comme des réalisations d'une loi de Poisson de moyenne $e_i \cdot \mu_i$, c'est-à-dire

$$d_i \sim \text{Poisson}(e_i \cdot \mu_i). \quad (2.4)$$

Afin d'estimer la force de mortalité sous-jacente aux données sur les décès observés, nous pouvons ainsi utiliser le modèle de régression de Poisson, qui s'inscrit dans la classe des modèles linéaires généralisés (McCullagh et Nelder, 1989). En effet, d'après l'équation (2.4), il en résulte

$$\begin{aligned} \ln(\mathbb{E}[\mathbf{d}]) &= \ln(\mathbf{e} \cdot \boldsymbol{\mu}) \\ &= \ln(\mathbf{e}) + \ln(\boldsymbol{\mu}), \end{aligned}$$

où les vecteurs \mathbf{d} , \mathbf{e} et $\boldsymbol{\mu}$ contiennent respectivement l'information, détaillée selon l'âge, sur les décès observés, les effectifs de population exposée au risque de décéder ainsi que les forces de mortalité. En utilisant les P-splines pour estimer les paramètres d'un tel modèle de régression de Poisson, nous obtenons

$$\begin{aligned} \ln\left(\hat{\mathbb{E}}(\mathbf{d})\right) &= \ln(\mathbf{e} \cdot \hat{\boldsymbol{\mu}}) \\ &= \ln(\mathbf{e}) + \ln(\hat{\boldsymbol{\mu}}) \end{aligned} \quad (2.5)$$

$$= \ln(\mathbf{e}) + \mathbf{B}\hat{\boldsymbol{\alpha}}, \quad (2.6)$$

où \mathbf{B} est la matrice associée à la base B-spline évaluée aux différents âges i et le vecteur $\hat{\boldsymbol{\alpha}}$ contient les paramètres estimés associés à chacun des B-splines inclus dans la base \mathbf{B} .

La fonction de log-vraisemblance pénalisée à maximiser pour estimer ces paramètres est la suivante

$$l^* = l(\boldsymbol{\alpha}; \mathbf{B}; \mathbf{d}) - \frac{1}{2} \boldsymbol{\alpha}' \mathbf{P} \boldsymbol{\alpha}. \quad (2.7)$$

Cette équation est légèrement différente de celle vue précédemment à l'équation (2.3) puisque le premier terme correspond ici à la log-vraisemblance habituelle d'un modèle linéaire généralisé plutôt que celle d'un modèle linéaire. Cela vient du fait que la loi de distribution des décès ne correspond pas à la loi normale, mais plutôt à celle de Poisson. Le second terme de l'équation 2.7, tout comme c'était le cas pour l'équation (2.3), est un terme de pénalité qui vise à assurer une certaine régularité au sein du vecteur des paramètres estimés $\hat{\boldsymbol{\alpha}}$. Pour plus de détails concernant la matrice de pénalité \mathbf{P} dans le contexte de modélisation de données sur la mortalité, voir la sous-section 3.7.3, Currie *et al.* (2004) ou encore Camarda (2008, Chap. 2).

À partir des équations (2.5) et (2.6), nous arrivons à l'expression suivante pour la force de mortalité lissée selon l'âge à l'aide des P-splines

$$\hat{\mu}(x) = \exp(\mathbf{B}(x)\hat{\boldsymbol{\alpha}}). \quad (2.8)$$

À titre d'exemple, la figure 2.7 présente les taux de mortalité observés selon l'âge, ainsi que la force de mortalité lissée obtenue à l'aide de la méthode des P-splines pour les femmes françaises en 2005.

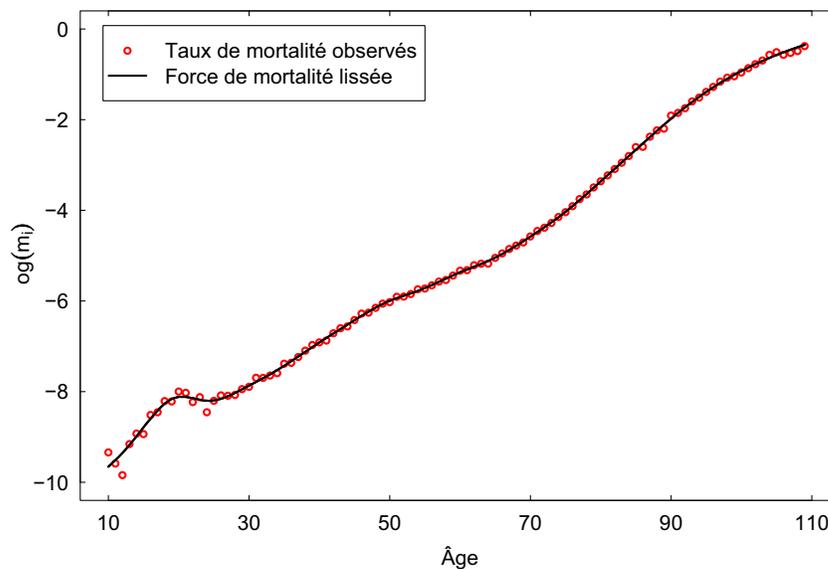
Suivant les relations d'équivalence entre la force de mortalité $\mu(x)$, la fonction de survie $S(x)$ et la fonction de densité $f(x)$ (Klein et Moeschberger, 1997, Chap. 2), nous obtenons

$$\begin{aligned} \hat{f}(x) &= \hat{\mu}(x) \hat{S}(x) \\ &= \hat{\mu}(x) \exp\left(-\int_0^x \hat{\mu}(t)dt\right). \end{aligned} \quad (2.9)$$

Ainsi, à partir de la force de mortalité lissée $\hat{\mu}(x)$ tirée de l'équation (2.8), nous pouvons évaluer la fonction de densité lissée $\hat{f}(x)$ correspondante à l'aide de méthodes d'intégration numérique usuelles. Cette fonction de densité lissée $\hat{f}(x)$ décrit la répartition des décès selon l'âge. En effet, la figure 2.8 présente le résultat obtenu pour les femmes françaises en 2005.

Dans le cadre des articles 1 et 3 présentés aux chapitres 3 et 5 de cette thèse, nous avons recours à la méthode des P-splines décrite ci-haut afin d'obtenir une

FIGURE 2.7. Taux de mortalité observés selon l'âge et force de mortalité lissée à l'aide des P-splines, France, femmes, 10 à 109 ans, 2005



Source : Human Mortality Database (2010)

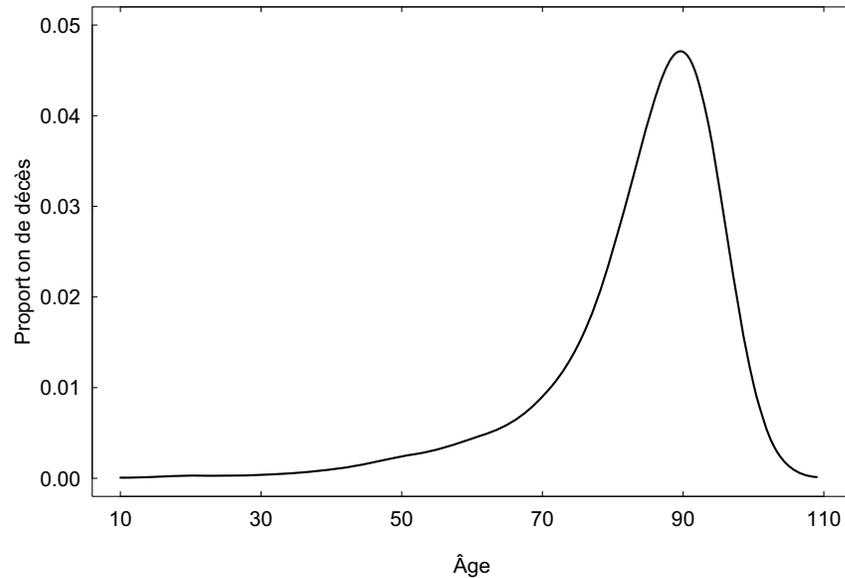
Note : Pour le lissage par P-splines, la base B-spline à noeuds équidistants comportait 23 B-splines et le paramètre de lissage valait 2.

description lisse de la répartition des décès selon l'âge, $\hat{f}(x)$, pour différentes populations prises à divers moments dans le temps. La sous-section 2.2.3 expose les mesures retenues dans ces articles pour effectuer un suivi des transformations de la mortalité au sein des populations âgées à partir de $\hat{f}(x)$. Pour l'instant, attardons-nous à une généralisation de la méthode de lissage par P-splines selon l'âge, utilisée dans l'article 2 présenté au chapitre 4, qui permet d'obtenir une description lisse de la répartition des décès selon l'âge et les années de calendrier. Soulignons d'emblée que cette approche de lissage en deux dimensions s'avère particulièrement utile pour l'étude de populations de petite taille, comme certaines provinces canadiennes.

2.2.2.2. Généralisation pour des données selon l'âge et l'année de calendrier

Une telle généralisation de la méthode de lissage par P-splines selon l'âge nécessite tout d'abord l'introduction d'une nouvelle base B-spline \mathbf{B} . Si \mathbf{B}_x dénote

FIGURE 2.8. Fonction de densité lissée à l'aide des P-splines décrivant la répartition des décès selon l'âge, France, femmes, 2005



Source : Human Mortality Database (2010)

Note : Pour le lissage par P-splines, la base B-spline à noeuds équidistants comportait 23 B-splines et le paramètre de lissage valait 2.

la matrice associée à la base B-spline sur le domaine des âges x et \mathbf{B}_y correspond à celle sur le domaine des années de calendrier y , alors la nouvelle base B-spline à utiliser est définie par

$$\mathbf{B} = \mathbf{B}_y \otimes \mathbf{B}_x, \quad (2.10)$$

où le symbole \otimes représente le produit de Kronecker.

En guise d'illustration, la figure 2.9 présente le produit de Kronecker de deux B-splines cubiques sur le domaine des âges et des années de calendrier, qui donne lieu à une forme de colline.

En suivant les mêmes principes que ceux décrits pour le lissage par P-splines unidimensionnel, nous utilisons un nombre relativement élevé de B-splines sur le domaine des âges et sur celui des années de calendrier. Ainsi, le domaine des âges et des années de calendrier devient automatiquement parsemé de multiples collines, telle celle représentée à la figure 2.9. La figure 2.10 suggère l'allure de la nouvelle base \mathbf{B} , définie par l'équation 2.10, à utiliser lors du lissage par P-splines

en deux dimensions. Il est à noter que par souci de clarté, seul un sous-ensemble de collines est présenté ici.

Pour une population donnée, dénotons la matrice des décès observés et celle des estimations de la population exposée au risque de décéder selon l'âge et les années de calendrier par \mathbf{D} et \mathbf{E} respectivement. Précisons que les lignes de ces matrices réfèrent aux âges, alors que leurs colonnes correspondent aux années de calendrier. Aux fins de la régression, définissons les deux vecteurs colonne suivants

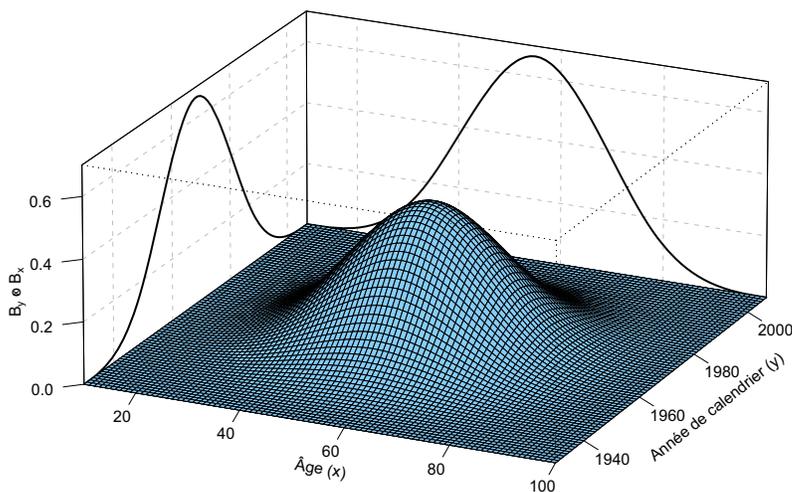
$$\mathbf{d} = \text{vec}(\mathbf{D}) \quad (2.11)$$

$$\mathbf{e} = \text{vec}(\mathbf{E}), \quad (2.12)$$

où l'opérateur vec a pour effet de placer les différentes colonnes d'une matrice les unes au-dessous des autres.

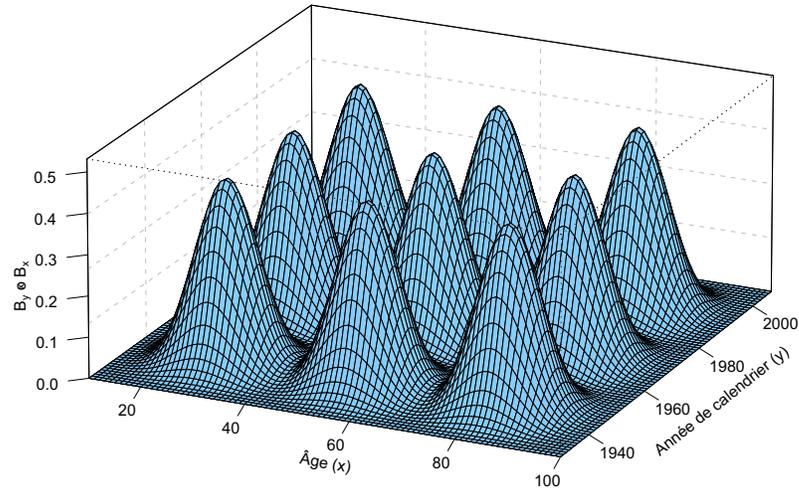
Sous l'hypothèse d'une force de mortalité $\mu(x, y)$ constante à l'intérieur de chaque intervalle d'âge et d'année de calendrier combinés, nous pouvons obtenir la force de mortalité lissée $\hat{\mu}(x, y)$ à l'aide d'un modèle de régression de Poisson et d'une approche par P-splines en deux dimensions. Cette hypothèse nous ramène effectivement aux expressions (2.5) et (2.6), où \mathbf{B} , \mathbf{d} et \mathbf{e} sont respectivement

FIGURE 2.9. Produit de Kronecker de deux B-splines cubiques sur le domaine des âges et années de calendrier



Source : Calculs effectués par l'auteur

FIGURE 2.10. Produit de Kronecker de deux bases B-splines cubiques sur le domaine des âges et années de calendrier



Source : Calculs effectués par l'auteur

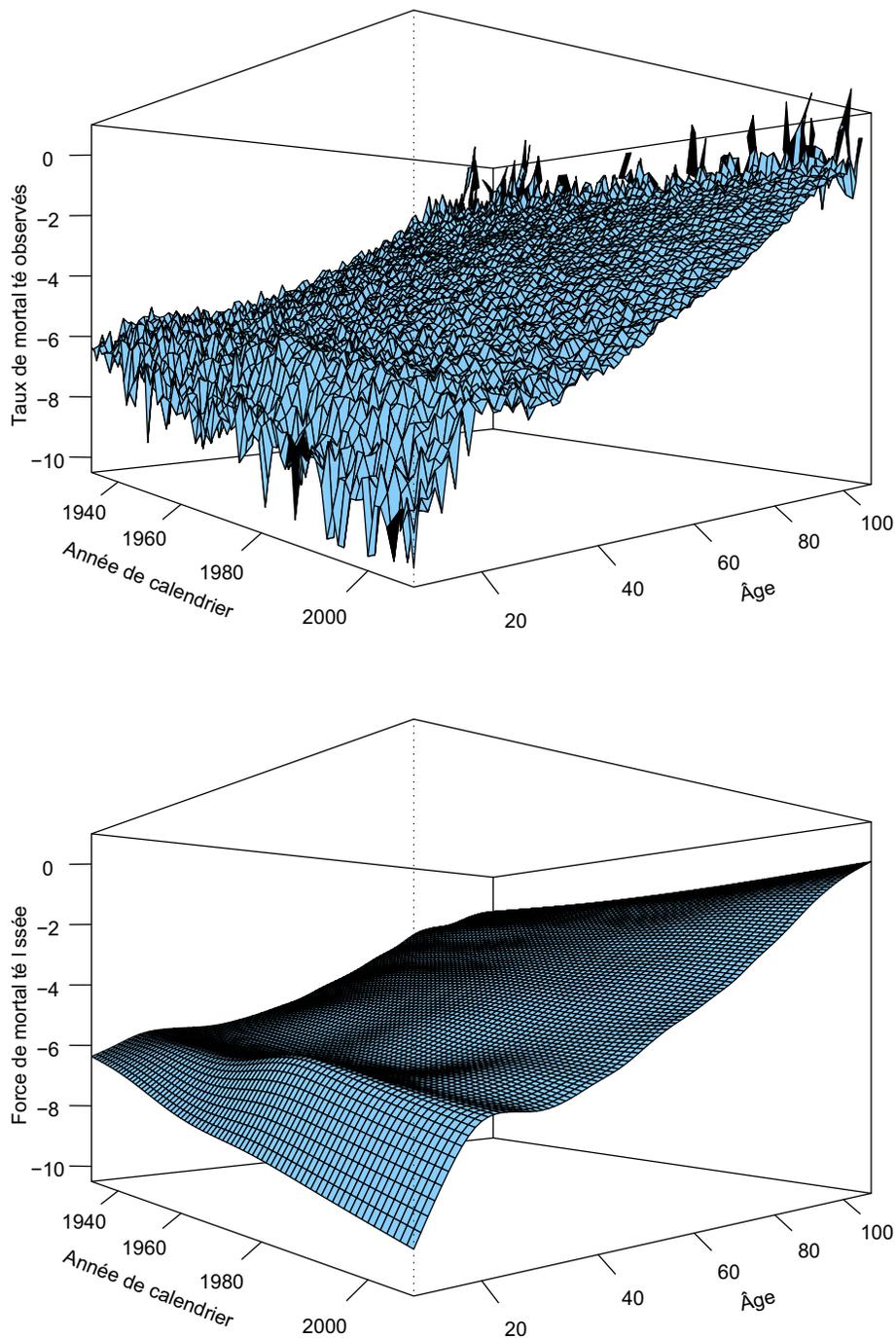
définis par les équations (2.10), (2.11) et (2.12). Comme auparavant, le vecteur des coefficients estimés $\hat{\alpha}$ est obtenu en maximisant la fonction de vraisemblance pénalisée donnée par l'équation (2.7), qui limitera les variations au sein des paramètres associés aux B-splines adjacents. Le compromis entre la régularité et la précision du modèle est toutefois gouverné par deux paramètres de lissage ici, l'un agissant sur la dimension des âges et l'autre sur la dimension des années de calendrier.

À titre d'exemple, la figure 2.11 présente les taux de mortalité observés selon l'âge et les années de calendrier, ainsi que la force de mortalité lissée obtenue à l'aide de la méthode des P-splines en deux dimensions pour les hommes de la Colombie-Britannique.

Grâce aux relations unissant la force de mortalité $\mu(x, y)$, la fonction de survie $S(x, y)$ et la fonction de densité $f(x, y)$ (Klein et Moeschberger, 1997, Chap. 2), ces trois fonctions étant continues sur les domaines de l'âge x et des années de calendrier y , nous avons

$$\hat{f}(x, y) = \hat{\mu}(x, y) \hat{S}(x, y)$$

FIGURE 2.11. Taux de mortalité observés (haut) et force de mortalité lissée à l'aide des P-splines en deux dimensions (bas), Colombie-Britannique, hommes, 10 à 105 ans, 1930 à 2007



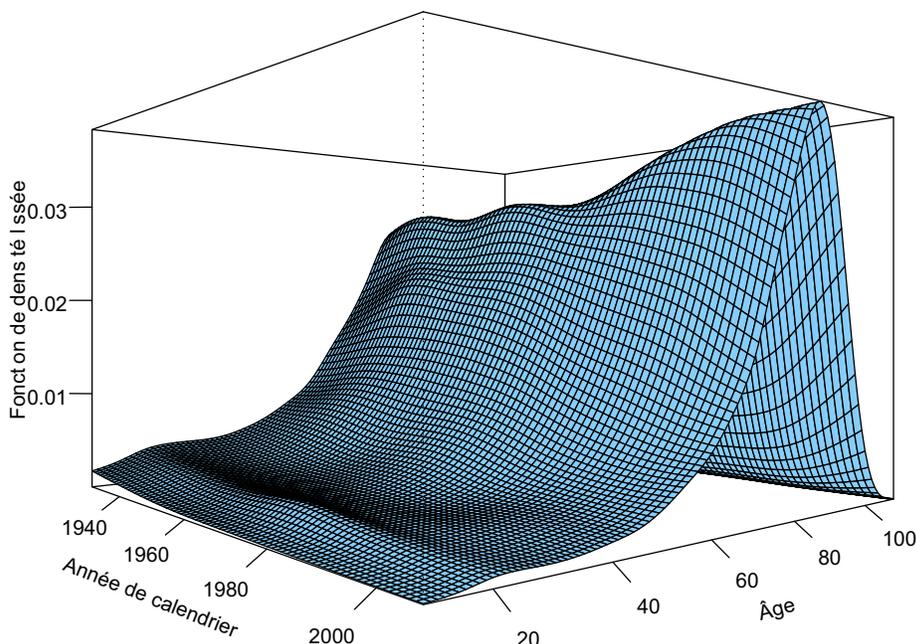
Source : Base de données sur la longévité canadienne (2010)

Note : Base B-spline cubique, noeuds équidistants (Âge : 23 B-splines, paramètre de lissage = 3; Années de calendrier : 18 B-splines, paramètre de lissage = 316).

$$= \hat{\mu}(x, y) \exp \left(- \int_0^x \hat{\mu}(u, y) du \right). \quad (2.13)$$

Ainsi, la fonction de densité lissée $\hat{f}(x, y)$, décrivant la répartition des décès selon l'âge et les années de calendrier, peut être obtenue à partir de $\hat{\mu}(x, y)$ et de méthodes d'intégration numérique usuelles. La figure 2.12 présente le résultat obtenu pour les hommes de la Colombie-Britannique.

FIGURE 2.12. Fonction de densité lissée à l'aide des P-splines en deux dimensions, Colombie-Britannique, hommes, 10 à 105 ans, 1930 à 2007



Source : Base de données sur la longévité canadienne (2010)

Note : Base B-spline cubique, noeuds équidistants (Âge : 23 B-splines, paramètre de lissage = 3 ; Années de calendrier : 18 B-splines, paramètre de lissage = 316).

Précisons qu'un module intitulé *MortalitySmooth*, conçu par Camarda (2009) pour le logiciel R (R Development Core Team, 2010) et procédant au lissage par P-splines de données sur la mortalité selon une ou deux dimensions, est disponible sur le Web depuis le mois de novembre 2009. Ce module conduit spécifiquement aux forces de mortalité lissées selon l'âge ou selon l'âge et les années de calendrier, respectivement décrites par les équations (2.9) et (2.13). Souhaitons que ce module

puisse faire en sorte que les P-splines continuent de gagner en popularité dans le domaine de la mortalité.

2.2.3. Mesure des transformations de la fonction de densité lissée au fil du temps

Tel que souligné au chapitre 1, les mesures de tendance centrale et de dispersion privilégiées dans les articles de cette thèse pour suivre les changements en matière de répartition des décès selon l'âge chez les personnes âgées sont les suivantes : l'âge modal au décès et l'écart type des âges au décès au-dessus de cet âge modal. Attardons-nous maintenant au calcul de ces mesures, en exploitant les fonctions de densité lissées \hat{f} présentées aux équations (2.9) et (2.13).

2.2.3.1. L'âge modal au décès et la dispersion des durées de vie au-delà de cet âge

À partir de la fonction de densité lissée $\hat{f}(x)$ décrivant la répartition des décès selon l'âge, nous estimons l'âge modal au décès comme suit

$$\hat{M} = \max_x \hat{f}(x). \quad (2.14)$$

Nous pouvons ensuite estimer l'écart type, relatif à l'âge modal au décès, des âges au décès au-dessus de cet âge modal à l'aide de l'équation

$$SD(\widehat{M+}) = \sqrt{\frac{\int_{\hat{M}}^{\infty} (x - \hat{M})^2 \hat{f}(x) dx}{\int_{\hat{M}}^{\infty} \hat{f}(x) dx}}. \quad (2.15)$$

Lorsque la fonction de densité lissée décrit la répartition des décès selon l'âge et les années de calendrier, les équations (2.14) et (2.15) deviennent respectivement

$$\hat{M}(y) = \max_x \hat{f}(x, y),$$

et

$$SD(\widehat{M+})(y) = \sqrt{\frac{\int_{\hat{M}(y)}^{\infty} (x - \hat{M}(y))^2 \hat{f}(x, y) dx}{\int_{\hat{M}(y)}^{\infty} \hat{f}(x, y) dx}}.$$

Ce chapitre a permis de présenter les données et les méthodes mises à contribution dans les articles qui suivent et qui forment le coeur de cette thèse.

Chapitre 3

Changes in the age-at-death distribution in four low mortality countries : A nonparametric approach

Nadine Ouellette, Robert Bourbeau

Department of Demography, Université de Montréal

ABSTRACT

Since the beginning of the twentieth century, important transformations have occurred in the age-at-death distribution within human populations. We propose a flexible nonparametric smoothing approach based on P-splines to refine the monitoring of these changes. Using data from the Human Mortality Database for four low mortality countries, namely Canada (1921-2007), France (1920-2007), Japan (1947-2008), and the USA (1945-2006), we find that the general scenario of compression of mortality no longer describes appropriately some of the recent adult mortality trends recorded. Indeed, reductions in the variability of age at death above the mode have stopped since the early 1990s in Japan and since the early 2000s for Canadian and French women, while their respective modal age at death continued to increase. These findings provide additional support to the shifting mortality scenario, using an alternative method free from any assumption on the shape of the age-at-death distribution.

Keywords : Modal Age at Death, Smoothing, *P*-splines, Compression of Mortality, Shifting Mortality, Canada, France, Japan, USA

Article soumis à la revue *Demographic Research*

3.1. INTRODUCTION

Over the course of the last century, we have witnessed major improvements in the level of mortality in regions all across the globe. This remarkable mortality decrease has also been characterized by important changes in the age-pattern of mortality, which inevitably led to substantial modifications in the shape of the distribution of age at death and survival curve over time. Measuring transformations in the age-at-death distribution or in the survival curve quickly became a topic of great interest among researchers, as their implications on societies are profound. For example, with accurate historical trends on average lifespan and lifespan inequality in hands, governments and policymakers are in better position to ensure sustainability of social-security and health-care systems.

Efforts to document such trends have indeed been made for several countries and regions : Canada (Martel et Bourbeau, 2003; Nagnur, 1986), France (Robine, 2001), Hong Kong (Cheung *et al.*, 2005), Japan (Cheung et Robine, 2007), the Netherlands (Nusselder et Mackenbach, 1996, 1997), Switzerland (Cheung *et al.*, 2009; Paccaud *et al.*, 1998), and the USA (Eakin et Witten, 1995; Fries, 1980; Lynch et Brown, 2001; Manton et Tolley, 1991; Myers et Manton, 1984a,b; Rothenberg *et al.*, 1991). Comparative studies involving various low mortality countries have also been undertaken (Canudas-Romo, 2008; Cheung *et al.*, 2008; Hill, 1993; Thatcher *et al.*, 2010; Wilmoth et Horiuchi, 1999).

Recently, Cheung *et al.* (2005) listed and reviewed more than 20 indicators used in these studies, where each indicator aimed at quantifying either the central tendency or the dispersion (variability) of age at death across individuals. Since the computation of these indicators often involves the use of parametric statistical modelling techniques (e.g., quadratic model, normal model, logistic model) that impose a predetermined structure on data, an exploration of nonparametric statistical methods, free from assumptions related to the structure of the data, is worth considering. Indeed, concerns over the potential influence of parametric modelling on the computation of indicators have already been addressed in previous studies (Cheung et Robine, 2007; Cheung *et al.*, 2005, 2008; Paccaud *et al.*,

1998). However, the use of a comprehensive nonparametric approach has not been explored extensively yet.

Accordingly, the first objective of this paper is to introduce a flexible nonparametric smoothing approach based on P-splines (Eilers et Marx, 1996) to refine the monitoring of changes in the age-at-death distribution. Our inspiration stems from previous work by Currie *et al.* (2004) and Camarda (2008) who specifically applied the P-spline method to mortality data. The second objective is to use this nonparametric approach to study recent trends in adult mortality in a subset of four low mortality countries : Canada, France, Japan, and the USA. Japan and France have repeatedly been identified as countries where average lifespan is very high and where compression of mortality has either stopped or slowed down substantially, even at older ages (Canudas-Romo, 2008; Cheung et Robine, 2007; Cheung *et al.*, 2008; Thatcher *et al.*, 2010). In contrast, adult mortality progress in the USA has been very slow in recent decades and the level of lifespan inequality is high compared to other low mortality countries (Canudas-Romo, 2008; Wilmoth et Horiuchi, 1999). Finally, rather few studies on this topic have included Canada (Martel et Bourbeau, 2003; Nagnur, 1986), although comparisons with the USA, its neighbouring country, should be informative. This subset of countries will therefore allow us to study differences in average lifespan and lifespan inequality among low mortality populations, and to revisit conclusions put forward by recent studies.

3.2. BACKGROUND

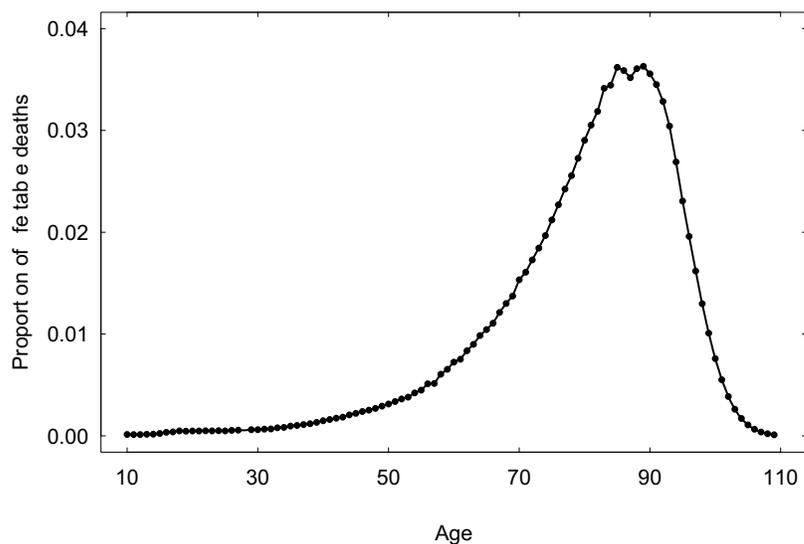
3.2.1. Average lifespan

How long do we live in average? is undoubtedly one of the most recurrent demographic question. The traditional way of answering this question is in terms of life expectancy at birth. However, the late modal age at death (referred to hereafter as the modal age at death) is another strong candidate that has received much recognition lately (Canudas-Romo, 2008, 2010; Cheung et Robine, 2007; Cheung *et al.*, 2005, 2008, 2009; Kannisto, 2000, 2001, 2007; Paccaud *et al.*, 1998;

Robine, 2001; Thatcher *et al.*, 2010). Dating back to the nineteenth century, pioneer work by Lexis (1877, 1878) on the concept of normal life durations identified the modal age at death as the most central and natural characteristic of human longevity. Indeed, unlike the life expectancy at birth, the modal age at death is solely influenced by adult mortality and consequently more sensitive to changes occurring among the elderly population (Canudas-Romo, 2010; Horiuchi, 2003; Kannisto, 2001). In low mortality countries where most deaths occur at older ages, it then becomes a strong indicator to monitor and explain recent changes in the age-at-death distribution.

Since the age distribution of life table deaths often tends to be erratic in the area surrounding the mode, various methods have been used to estimate the modal age at death. In order to find the value at which the maximum of a given frequency distribution is reached, Pearson (1902) recommended interpolating a curve through the top of ordinates and using the maximum of this interpolated curve to identify the modal value. Many followed this advice and fitted a quadratic polynomial model to life table deaths around the age M^* where the highest number of deaths occurs in the life table (Canudas-Romo, 2008, 2010; Cheung *et al.*, 2008; Kannisto, 2001, 2007; Thatcher *et al.*, 2010). All except Cheung *et al.* (2008) applied Kannisto's (2001) quadratic procedure, which consists in fitting a quadratic model using life table deaths at ages M^* , $M^* - 1$, and $M^* + 1$. The resulting modal age at death estimated thru this procedure thus always lies between M^* and $M^* + 1$ (Canudas-Romo, 2008). The main difficulty encountered with this approach is the following. Often, the roughness of the age distribution of life table deaths is such that the age with the highest number of deaths M^* does not clearly stand out (see Figure 3.1). Instead, there are several age candidates, a few years apart from each other, which lead to modal ages at death estimates that are also a few years apart from each other. This then translates into substantial artificial fluctuations in the estimated modal age at death trend over time. Note that Cheung *et al.* (2008) fitted their quadratic models based on the $[M^* - 5, M^* + 5]$ life table age range, raising other concerns, as explained below.

FIGURE 3.1. Life table death distribution, USA, Women, 2003



Source : Human Mortality Database (2010)

Cheung et Robine (2007) and Cheung *et al.* (2009) have instead been inspired by Lexis’s concept of normal life durations (Lexis, 1877, 1878) to estimate the modal age at death. According to Lexis, if premature deaths are removed from all deaths, the “normal” deaths that remain are described by a normal distribution along age. Therefore, Cheung and her colleagues fitted a scaled normal model to life table deaths, starting five years before the age where the highest number of deaths occurs in the life table. Although this procedure is appealing because it is theoretically oriented as proposed by Lexis, the scaled normal model imposes a strict “bell”-shaped structure to the selected life table deaths and any departure from this structure in the data is ignored. Furthermore, the age at which the fitting procedure starts first requires identifying a single age with the highest number of deaths in the life table. As explained above, this can be problematic when there are several age candidates to choose from. Finally, systematically starting the fitting procedure five years before the age with the highest number of deaths in the life table to avoid including premature deaths into the modelling task is questionable. Indeed, the magnitude and the age profile of premature mortality have dramatically changed over the years and vary substantially across populations.

A simplified version of the logistic model of mortality, namely the Kannisto model of old-age mortality (Thatcher *et al.*, 1998; Thatcher, 1999), has also been exploited to estimate the modal age at death (Thatcher *et al.*, 2010). However, this study by Thatcher *et al.* (2010) was oriented towards understanding the dynamics behind changes in the modal age at death over time rather than at monitoring these changes closely. Using empirical death rates at ages 70 and 90 only to fit this simple logistic model, the authors found that the resulting modal age at death estimates were reasonably close to those computed with Kannisto's (2001) quadratic method described above. Further theoretical considerations regarding the modal age at death and its related measures in several other mathematical models of mortality are provided by Canudas-Romo (2008) and Robine *et al.* (2006).

To our knowledge, Paccaud *et al.* (1998) were the only ones to rely on a nonparametric method to estimate the modal age at death. Nonparametric models are seldom used for such task, although they are generally more flexible than parametric models because they do not impose a predetermined structure on the actual data. In this paper, we demonstrate that the P-spline method provides the ability to refine the monitoring of changes in modal age at death over time.

3.2.2. Lifespan inequality

How variable is the age we die at? is another relevant demographic question worth studying and which has fortunately gained in popularity. Wilmoth et Horiuchi (1999) analysed ten measures of variability of age at death, including those used in previous studies (Eakin et Witten, 1995; Keyfitz, 1977; Myers et Manton, 1984a; Nusselder et Mackenbach, 1996). Based on period data from Sweden, Japan, and the USA, they found strong empirical correlation between these indicators. For ease of interpretation and computation, they recommended using the interquartile range to monitor changes in the variability of age at death within human populations. Shortly after, Kannisto (2000) rather argued in favour of the C-family of indicators – i.e., the shortest age interval in which a given proportion of deaths occur – in particular because it allowed for a more complete analysis

of mortality dispersion. He also suggested two other indicators to measure variability in old age mortality instead of over the entire age range. One of them, a mode-based indicator called the standard deviation of ages at death above the mode, has been used in several subsequent studies on low mortality countries (Cheung et Robine, 2007; Cheung *et al.*, 2008, 2009; Kannisto, 2001, 2007; Thatcher *et al.*, 2010). As this measure of variability solely involves deaths occurring at ages above the mode, changes in the magnitude of premature mortality over time are less likely to influence the results.

In this paper, recent changes in the variability of adult mortality will be monitored with the standard deviation of ages at death above the mode. Specifically, this indicator corresponds to the root mean square of ages at death from the mode for individuals who lived up to or beyond the mode. A decline in this measure over time thus indicates that deaths above the mode have been compressed into a shorter age interval, hence the expressions *compression of deaths above the mode* and *old-age mortality compression*. Given our objective to study recent adult mortality changes in low mortality countries, the fact that the standard deviation of ages at death above the mode is directly measuring variability in old age mortality is an important asset.

Most variability indicators discussed in the literature have been computed with deaths extracted from life tables closed with a parametric model and/or involve the modal age at death which was estimated using a parametric model. With the nonparametric smoothing approach based on P-splines, standard deviation of ages at death above the mode estimates are free from any kind of predetermined data structure susceptible to influence results.

3.2.3. Choice of a nonparametric method

When it comes to choosing a nonparametric modelling method to smooth data, there are several avenues. Under the running statistics category, there are notably kernel smoothers (Silverman, 1986; Härdle, 1990) and LOWESS – i.e., locally weighted scatterplot smoothing – (Cleveland, 1979). Then, spline methods

are plentiful : for example, there are smoothing splines, regression splines (Eubank, 1988), B-splines (de Boor, 1978), and P-splines (Eilers et Marx, 1996). Among all these techniques available, why selecting P-splines? First of all, although the good properties of P-splines can be found in at least one of the other smoothing methods listed, P-splines corresponds to the smoother that features most of these good properties simultaneously (see the Rejoinder section in Eilers et Marx (1996)). Furthermore, P-splines are easy to use, to program and to understand, while this is not necessarily the case for other nonparametric approaches. Finally, previous studies have already demonstrated that P-splines are well-suited to the task of smoothing mortality data (Currie *et al.*, 2004; Camarda, 2008).

In this paper, the P-splines smoothing procedure will be applied to mortality data from age 10 and onwards. Infant and child mortality are excluded because they present unique features that would require the use of a smoothing method suited for ill-posed data ; this goes well beyond the scope of the present study. Why not restricting the fitting procedure to the surroundings of the modal age at death ? Starting at age 10 avoids selecting another age interval, perhaps more subjectively, over which the smoothing procedure should take place. As discussed in subsection 3.2.1, the usual roughness in the mode area makes it difficult to determine a single age with the highest life table deaths. Restricting the fit to the surroundings of the mode would also involve choosing the length of the age interval over which the fitting procedure should take place. Whether this length should vary over time and across populations would also have to be debated. Another argument in favour of applying the fitting procedure from age 10 and onwards is that the analysis can then rely on a single smoothing method instead of several ones. Indeed, modal age at death estimates and standard deviation of ages at death above the mode estimates can all be computed based on the P-spline method.

3.3. DATA AND METHODS

Actual death counts d_i and exposure to risk e_i by single year of age i , single calendar year, and sex for Canada, France, Japan, and the USA were taken from the Human Mortality Database (HMD 2010). Complete period life tables were also extracted from the HMD for comparison purposes. Data selected for France (1920-2007) refer to the civilian population and begin after the influenza pandemic of 1918-19 to prevent sudden disruption in patterns over time. For the USA (1945-2006), years from 1945 and onwards were selected. All years available on the HMD were used for Canada (1921-2007) and Japan (1947-2008).

Let $m_i = d_i/e_i$ denote the central death rate at age i for a given country, calendar year, and sex. Also, let μ_i be the force of mortality (or instantaneous death rate) at age i , such that $m_i \simeq \mu_{i+\frac{1}{2}}$ (Thatcher *et al.*, 1998, Appendix A). Under the assumption that the force of mortality is a piecewise constant function within each age and time interval – i.e., $\mu(x) = \mu_i$ for all $x \in [i, i + 1)$ – then death counts d_i are a realization of a Poisson distribution with mean $e_i \cdot \mu_i$:

$$d_i \sim \text{Poisson}(e_i \cdot \mu_i). \quad (3.1)$$

The Poisson regression model used in this paper is based on this assumption. Since our response variable \mathbf{d} , the vector of observed death counts, is non-normally distributed, we introduce a linear predictor $\boldsymbol{\eta}$ such that

$$\boldsymbol{\eta} = \ln(\mathbb{E}[\mathbf{d}]). \quad (3.2)$$

The Poisson regression model assumes $\boldsymbol{\eta}$ can be modelled by a linear combination of unknown parameters. Following the work of Eilers et Marx (1996), Currie *et al.* (2004), and Camarda (2008), we used a flexible nonparametric approach based on B-splines with penalties known as P-splines to estimate those unknown parameters. The P-spline method is described in more detail in the Appendix, sections 3.7.1 to 3.7.3.

From relations (3.1) and (3.2), we then have

$$\boldsymbol{\eta} = \ln(\mathbb{E}[\mathbf{d}]) = \ln(\mathbf{e} \cdot \boldsymbol{\mu}) = \ln(\mathbf{e}) + \ln(\boldsymbol{\mu}) = \ln(\mathbf{e}) + \mathbf{B}\mathbf{a},$$

where \mathbf{d} , \mathbf{e} , $\boldsymbol{\mu}$ are respectively actual death counts, exposures to risk, and force of mortality vectors for a given country, calendar year, and sex. The term $\ln(\mathbf{e})$ is commonly referred to as the *offset* in a Poisson regression setting. Furthermore, \mathbf{B} is the B-spline basis matrix and \mathbf{a} is the vector of respective regression parameters to estimate. Using the P-spline method adapted for Poisson death counts to estimate \mathbf{a} , we obtain

$$\hat{\boldsymbol{\eta}} = \ln(\mathbf{e}) + \mathbf{B}\hat{\mathbf{a}}. \quad (3.3)$$

Taking the exponential of equation (3.3) gives smoothed death counts, which are however influenced by exposure to risk \mathbf{e} and therefore not comparable from one age, sex, calendar year and/or country to another. In order to obtain comparable – i.e., standardized – age-at-death distributions, one must extract the smoothed forces of mortality from the smoothed death counts. Given equation (3.3), a smoothed trend for the force of mortality is readily obtained as

$$\hat{\mu}(x) = \exp(\mathbf{B}(x)\hat{\mathbf{a}}).$$

The corresponding smoothed survival function expressed as

$$\hat{S}(x) = \exp\left(-\int_0^x \hat{\mu}(t)dt\right)$$

can then be calculated using standard numerical integration techniques. Finally, the smoothed probability density function describing the age-at-death distribution for a given country, calendar year, and sex corresponds to

$$\hat{f}(x) = \hat{\mu}(x)\hat{S}(x). \quad (3.4)$$

We monitored changes in the various smoothed age-at-death distributions over time with their respective estimated modal age at death

$$\hat{M} = \max_x \hat{f}(x)$$

and estimated standard deviation above the mode

$$\widehat{SD}(M+) = \sqrt{\frac{\int_{\hat{M}}^{\infty} (x - \hat{M})^2 \hat{f}(x) dx}{\int_{\hat{M}}^{\infty} \hat{f}(x) dx}}.$$

The three curves $\mu(x)$, $S(x)$, and $f(x)$ are related in such way that changes in one of them will necessarily be reflected in the others (Wilmoth, 1997). In this paper, we mainly focus on $\hat{f}(x)$, but analysis based on $\hat{\mu}(x)$ or $\hat{S}(x)$ would yield consistent conclusions.

3.4. RESULTS

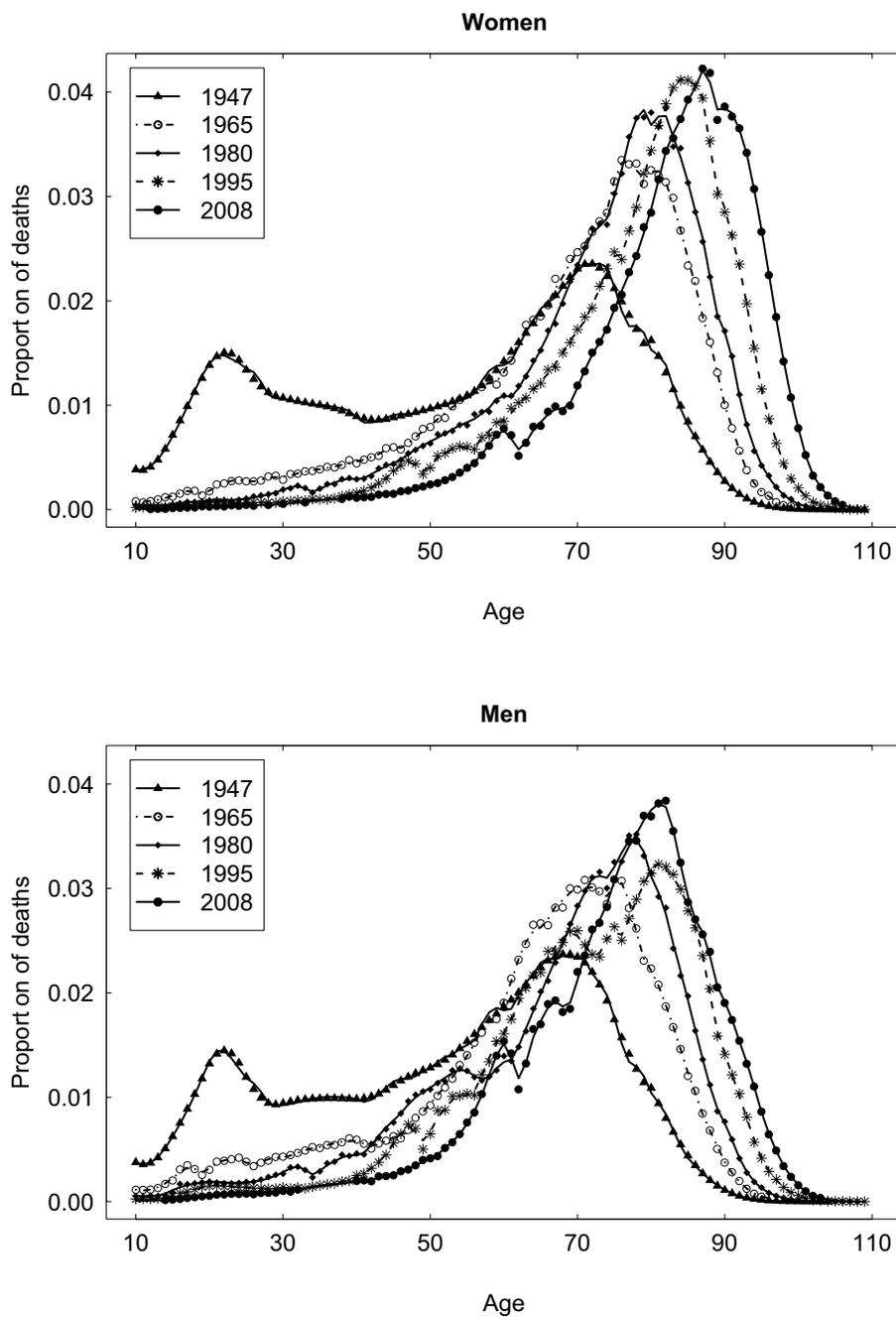
This section is divided into two parts. First, an illustration of the nonparametric Poisson P-spline smoothing approach using Japanese data is provided. An analysis of recent trends in adult mortality in Canada, France, Japan, and the USA follows.

3.4.1. Illustration of the nonparametric P-spline approach using Japanese data

Figure 3.2 shows actual and fitted death counts in Japan by sex according to model (3.3) for a sample of years between 1947 and 2008. We see that the Poisson regression model fits the actual data accurately.

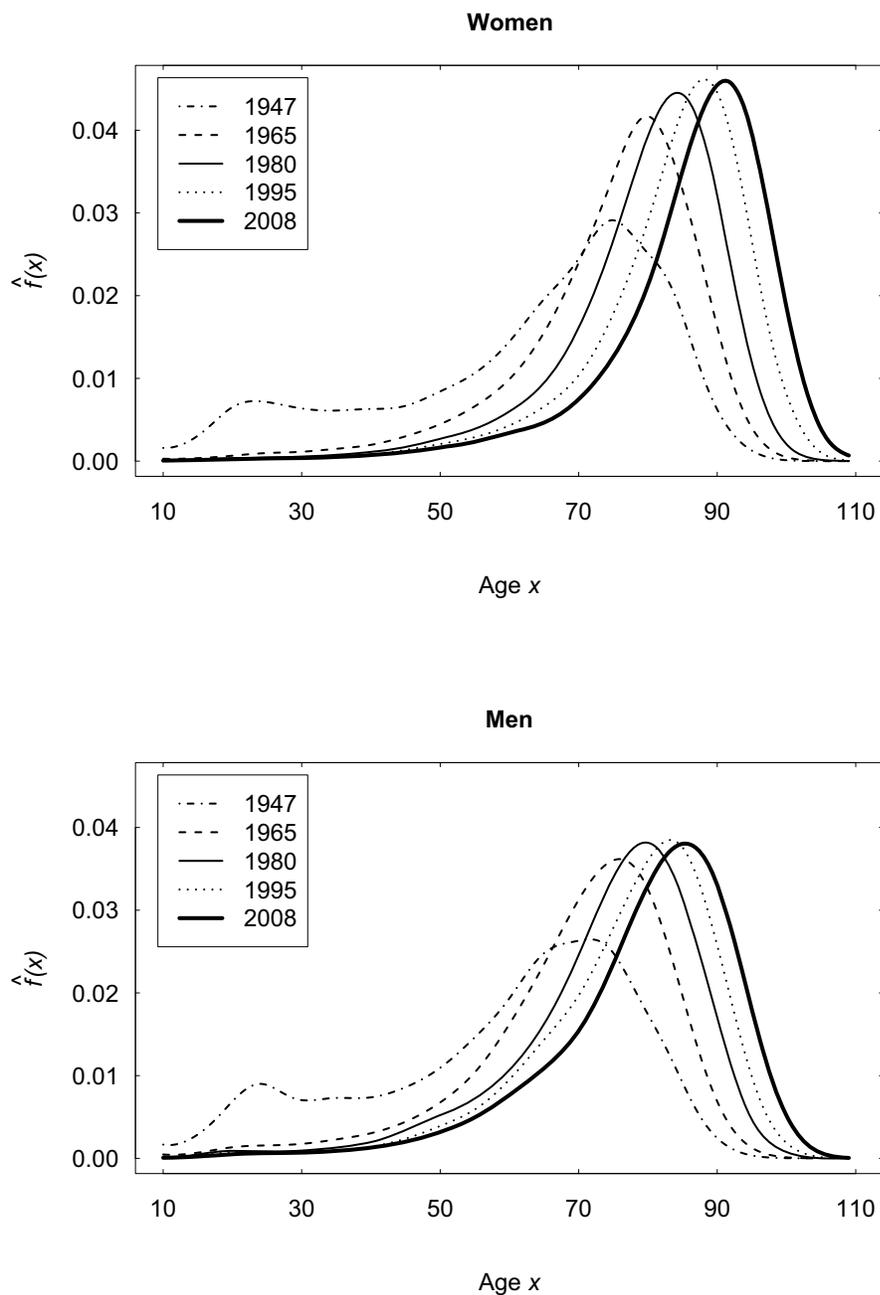
However, as discussed in section three, fitted death counts shown in this figure are influenced by exposure to risk. Extracting the smoothed forces of mortality underlying these fitted death counts and then computing the corresponding smoothed survival curves is required to obtain the smoothed density functions $\hat{f}(x)$ describing the age-at-death distributions provided in Figure 3.3. These resulting curves are therefore standardized with respect to exposure to risk. In that sense, they are comparable to life table age-at-death distributions available on the HMD for example (see Figure 3.9 in the Appendix, section 3.7.4), but they also have the following key advantages. The smoothed age-at-death distributions described

FIGURE 3.2. Comparison between actual death counts (points) and fitted death counts (lines) resulting from non-parametric P-spline estimations of Poisson regression models, Japan, 1947 to 2008



Source : As for Figure 1

FIGURE 3.3. Smoothed density functions describing the age-at-death distribution and resulting from nonparametric P-spline estimations of Poisson regression models, Japan, 1947 to 2008



Source : As for Figure 1

by $\hat{f}(x)$ in Figure 3.3 are free from any modelling structure at advanced ages that would otherwise impose a strict behaviour on deaths. They are also known numerically, which means that the level of precision is chosen by the user, and this facilitates accurate estimations of indicators \hat{M} and $\widehat{SD}(M+)$.

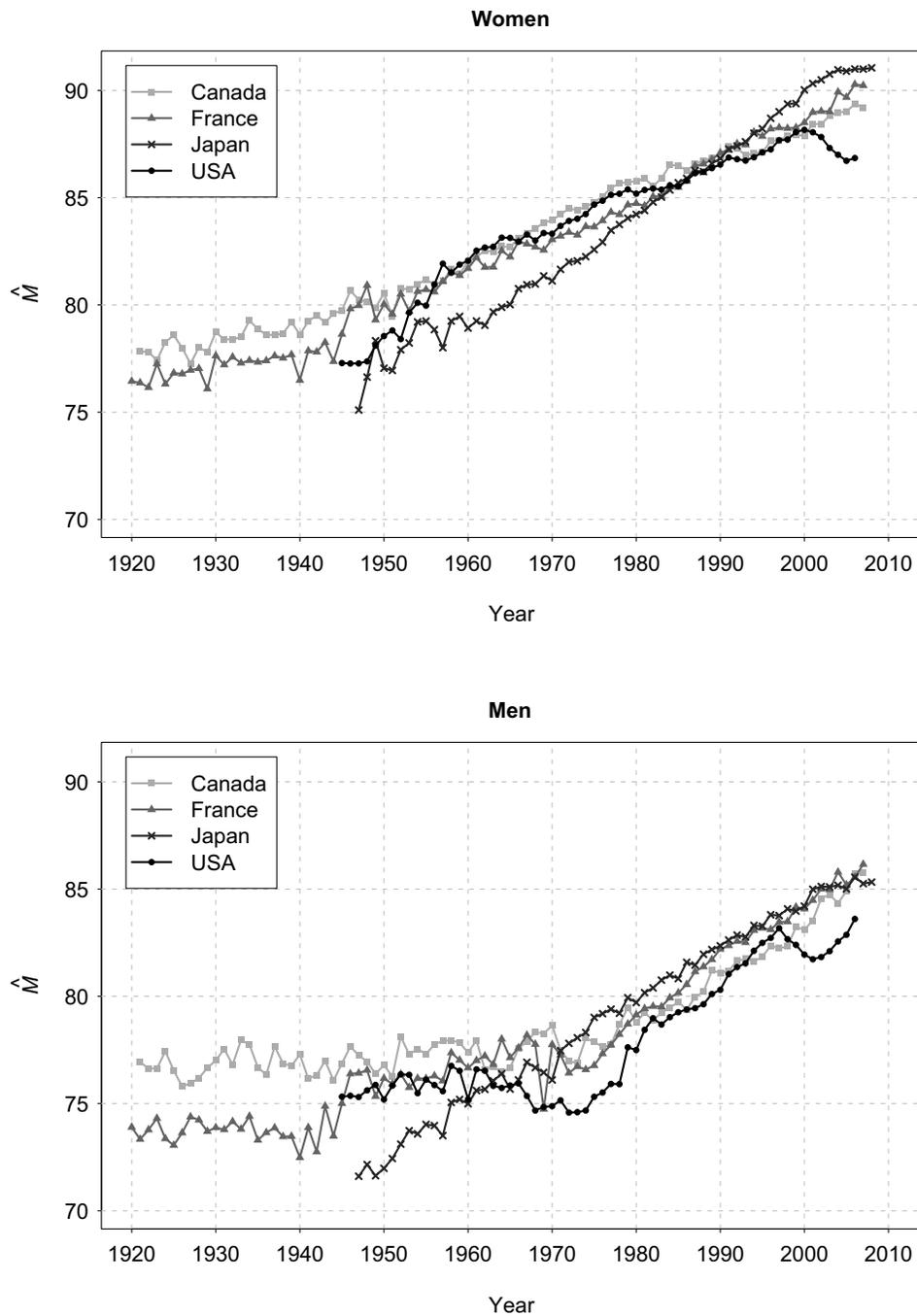
Figure 3.3 indicates that a great increase in average lifespan occurred in Japan for both women and men during the 1950-2008 period, and that it was paralleled by a substantial decrease in lifespan inequality, at least until 1980. Indeed, as time went by, the age-at-death distributions have progressively moved to higher ages and became less and less spread out. However, accurate diagnostics are difficult with such visual inspection and simultaneous comparisons with other countries quickly become confusing. Time trends of summary measures such as \hat{M} and $\widehat{SD}(M+)$ are more informative, as shown in the next subsection.

3.4.2. Adult mortality trends in four low mortality countries

3.4.2.1. *Modal age at death estimates (\hat{M})*

Figure 3.4 presents sex-specific changes in average lifespan measured by \hat{M} in each of the four countries under study. It reveals that \hat{M} increased substantially during the periods studied, although the increasing pace varied over time as well as across countries and sexes. Indeed, during the 1950s, 1960s, and 1970s, Japanese women were systematically showing the lowest modal age at death values, but their quick and steady mortality improvements eventually led them well above the others. Thanks to an average growth rate of more than 3.3 months per year between 1947 and 2000, \hat{M} reached 90 years in 2000. Such achievement in the level of the modal age at death occurred six years later among French women and has yet to be observed in Canada and the USA. However, after five decades of sustained increase in \hat{M} among Japanese women, the pace of increase slowed down and results for the most recent years even revealed an unexpected levelling off. Indeed, since 2004, their modal age at death remained almost constant at about 90.9 years.

FIGURE 3.4. Estimated modal age at death trends in four low mortality countries based on smoothed density functions



Source : As for Figure 1

Upward linear trends for \hat{M} among French and Canadian women have been recorded throughout most of the years studied, including the latest ones. The pace at which \hat{M} increased for these women might have been less spectacular than for the Japanese's – i.e., 1.9 months per year in average for French and 1.6 for Canadians – but \hat{M} still reached 90.2 years in France in 2007 and 89.2 years in Canada in 2007. Interestingly, Canadian women have enjoyed a long-lasting advantage over the French with respect to \hat{M} , but since the beginning of the 1990s, this trend has been reversed. The case of US women is somehow unique because a sudden decrease in \hat{M} occurred between 2000 and 2006, despite the fact that their prior experience was very similar to that of French and Canadian women. Indeed, in 2006, the modal age at death for US women was at 86.9 years, a level comparable to what they had achieved more than a decade earlier. Thus, based on 2006 results, US women are substantially disadvantaged compared to those of the other three countries in terms of \hat{M} .

If we focus on results for men, we see that the rapid linear progression of the modal age at death in Japan also stands out. The average growth rate for \hat{M} between 1947 and 2001 among Japanese men was above 2.9 months per year. However, a substantial slowdown occurred afterwards and their modal age at death only increased from 84.9 years in 2001 to 85.3 in 2008.

Since the beginning of the 1990s, French and Japanese men have been following each other very closely. Canadian men have also recently joined them, even though steady increases in \hat{M} only started in the early 1970s. As a matter of fact, from the 1920s to the 1960s, the modal age at death for Canadian men oscillated around 77 years, most probably because mortality reductions at ages older than \hat{M} essential for its increase were limited during those years (Canudas-Romo, 2010). Nevertheless, in 2006, the modal ages at death for men in Canada, Japan, and France were almost identical – i.e., about 85.4 years. Recall that disparities between these countries at that point in time were much greater among women. In 2007, \hat{M} was slightly higher for French men (86.2 years) and Canadian men (85.7 years) than it was for Japanese (85.3 years), but this could only be temporary.

In the USA, the upward trend for \hat{M} has been interrupted twice during the period under study : \hat{M} decreased between 1962 and 1974, and between 1997 and 2001. The latter period reveals the most unforeseen and abrupt decrease among US men. Indeed, within four calendar years, \hat{M} went from 83.2 to 81.7 years, yielding an average annual loss of 4.5 months. Fortunately, their modal age at death increased steadily afterwards and reached 83.6 years in 2006. Compared to the other three countries, US men are still lagging behind in 2006, but their delay is shorter than it was a few years ago.

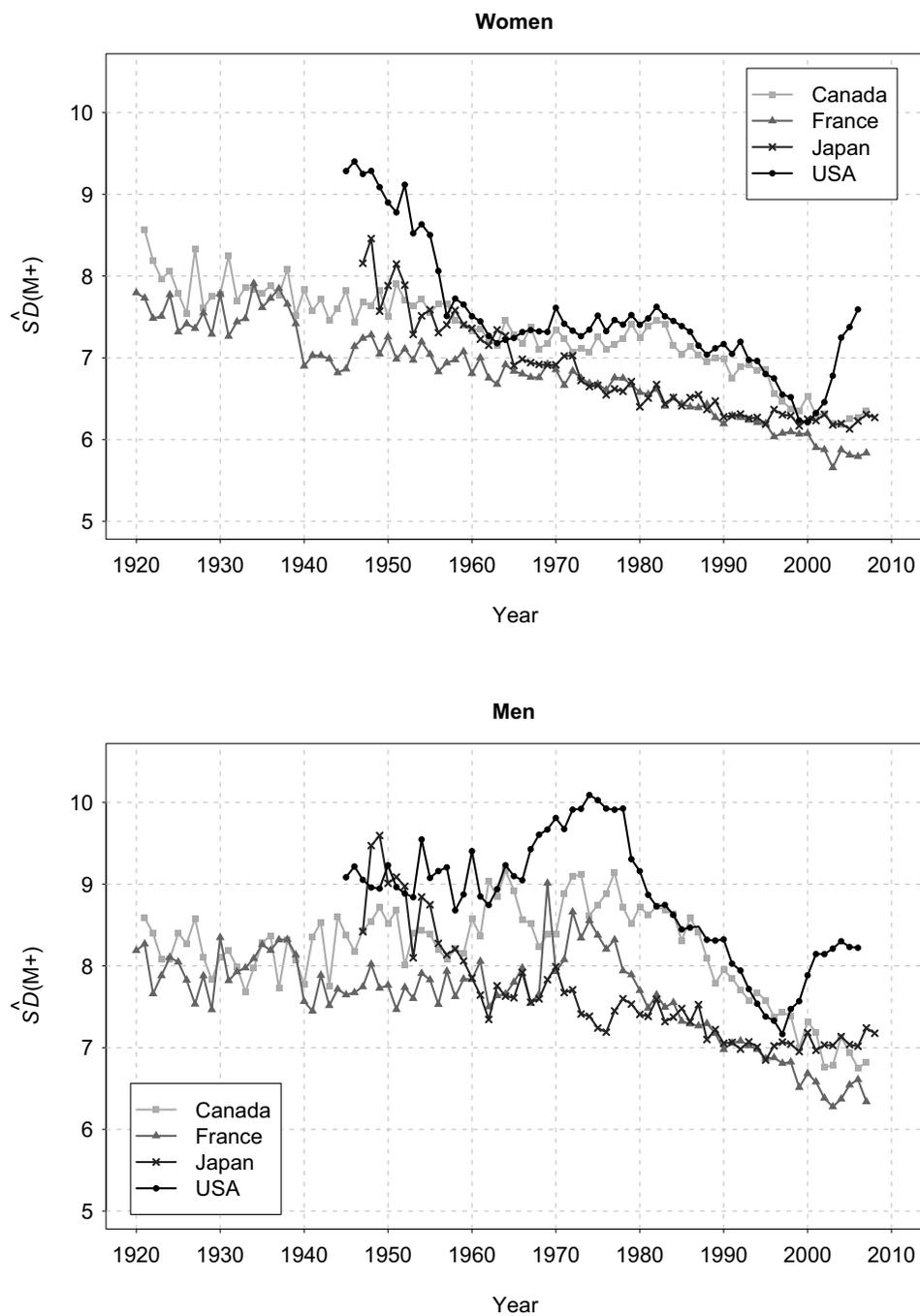
3.4.2.2. *Standard deviation of ages at death above the mode estimates* ($\widehat{SD}(M+)$)

Figure 3.5 displays country-specific trends by sex for $\widehat{SD}(M+)$. In all cases, the values at the end of the period studied were lower than at the beginning, indicating that the level of variability of deaths at older ages decreased over time. Compression of deaths above the mode thus occurred for each sex in these countries during the period covered by our study. However, the pace of decline differs across countries and between women and men.

Among Japanese women, $\widehat{SD}(M+)$ decreased rapidly between 1947 and the mid-1960s. Indeed, $\widehat{SD}(M+)$ went from 8.2 to 7.0 years in less than 20 years. Then, $\widehat{SD}(M+)$ declined at a slightly lesser pace and reached 6.3 years in 1990. Afterwards, the level of variability of age at death above \hat{M} remained more or less the same, indicating that compression of deaths above the mode has probably come to an end for these women.

French and Canadian women reveal similar declining trends in $\widehat{SD}(M+)$ over time, although the level of old-age mortality disparity was almost always higher among the latter. Between 1920 and 1960, $\widehat{SD}(M+)$ went from 7.8 to 6.8 years in France ; in Canada, it went from 8.6 years in 1921 to 7.3 years in 1960. Recall that an equivalent reduction in the level of variability of age at death above the mode was accomplished by Japanese women in one third of the time. After 1960, both French and Canadian women recorded a pause in decline for $\widehat{SD}(M+)$, but

FIGURE 3.5. Estimated standard deviation of ages at death above the mode in four low mortality countries based on smoothed density functions



Source : As for Figure 1

the decline resumed earlier in France than in Canada. Nonetheless, since the early 2000s, $\widehat{SD}(M+)$ has been stagnating at about 6.3 years among Canadian women. Old-age mortality compression might have stopped as well among French women lately because $\widehat{SD}(M+)$ remained at 5.8 years since 2001, except in 2003.

In the USA, $\widehat{SD}(M+)$ decreased very rapidly among women between 1945 and the early 1960s. Indeed, $\widehat{SD}(M+)$ went from 9.3 to about 7.2 years during that period. Afterwards, changes in variability of age at death above the mode ceased for two decades. Further decline in $\widehat{SD}(M+)$ was then observed between the early 1980s and late 1990s. In 2000, $\widehat{SD}(M+)$ reached 6.2 years among US women, a level similar to that of Japanese, French and Canadian women. Since then however, a sudden upward trend in $\widehat{SD}(M+)$ has been recorded for US women and brought $\widehat{SD}(M+)$ back up to 7.6 years in 2006. Note that this unexpected trend among US women for the latest years studied coincides with the sudden decrease in \hat{M} between 2000 and 2006 (see section 3.4.2.1).

From the men's perspective, country-specific trends for $\widehat{SD}(M+)$ presented in Figure 3.5 indicate that old-age mortality compression was strong in Japan from the 1950s to early 1960s. Indeed, $\widehat{SD}(M+)$ went from about 9.5 to 7.5 years during that period. $\widehat{SD}(M+)$ declined at a much lesser pace afterwards and reached 7.1 years in 1990. Since then, little change in $\widehat{SD}(M+)$ has been recorded among Japanese men, suggesting that their old-age mortality compression is essentially over.

Among French and Canadian men, the onset of decline in $\widehat{SD}(M+)$ occurred in the 1970s and roughly coincides with the onset of increase in \hat{M} (see section 3.4.2.1). Prior to the 1970s, $\widehat{SD}(M+)$ rather oscillated around a value of about 7.8 years in France and tended to increase at a very slow pace in Canada – i.e., from about 8.1 to 9.0 years in five decades. From the 1970s to 2000, $\widehat{SD}(M+)$ decreased consistently in both countries and finally reached 6.7 and 7.3 years among French and Canadian men respectively. Since then, $\widehat{SD}(M+)$ seems to be declining at a lesser pace than before, but further decline in the level of variability of age at death above the mode is expected in upcoming years on this basis.

In the USA, $\widehat{SD}(M+)$ oscillated around 9.1 years among men from 1945 to the mid-1960s. Then, up to the mid-1970s, $\widehat{SD}(M+)$ increased and peaked at 10.1 years in 1974. Between the mid-1970s and the late 1990s, $\widehat{SD}(M+)$ followed a steep declining trend, which ended at 7.2 years in 1997. For four consecutive years afterwards, $\widehat{SD}(M+)$ increased drastically and reached 8.1 years in 2001. In the last five years studied, $\widehat{SD}(M+)$ barely increased and attained 8.2 years in 2006. Therefore, except for the latest 2001-2006 years, periods during which $\widehat{SD}(M+)$ stagnated among US men corresponded to periods where \hat{M} stagnated as well (see section 3.4.2.1). Furthermore, in periods where $\widehat{SD}(M+)$ was increasing, \hat{M} was decreasing and vice-versa. Given that from 2001 to 2006, \hat{M} increase among US men while $\widehat{SD}(M+)$ stagnated, convergence towards lower levels of variability of deaths above the mode achieved by the other countries remains unsure for US men. It is however too early to admit that old-age mortality compression has ended among these men.

3.5. DISCUSSION

Transformations in the survival curve or in the age-at-death distribution over time have usually been monitored with indicators tied to various parametric statistical modelling techniques (e.g., quadratic model, normal model, logistic model). In this paper, we introduced a nonparametric approach based on P-splines to refine the monitoring of changes in the central tendency and variability of adult lifespan over time. As opposed to parametric methods, the P-spline method is free from any assumptions related to the data structure, otherwise susceptible to influence results. The flexibility of P-splines instead leads to detailed smoothed age-at-death distributions, as described by actual death counts and exposure to risk data. Modal age at death (M) estimates and standard deviation of ages at death above the mode ($SD(M+)$) estimates computed from these smoothed age-at-death distributions provide precise monitoring of changes in adult mortality over time.

In recent studies (Canudas-Romo, 2008; Cheung et Robine, 2007; Cheung *et al.*, 2008; Thatcher *et al.*, 2010), authors have argued that the general scenario

of compression of mortality, where increases in average lifespan are paralleled with decreases in variability of age at death, no longer describes appropriately recent adult mortality changes recorded in some low mortality countries. Instead, the shifting mortality regime (Bongaarts, 2005; Kannisto, 1996), where adult mortality slides proportionally to higher ages over time, provides a better description.

To investigate this further, we applied the P-spline method to a subset of four low mortality countries, namely Canada, France, Japan, and the USA. The results show that since the early 1990s, compression of deaths above the mode among Japanese women and men has stopped, while the mode continued to increase rapidly, at least until the early 2000s. In other words, Japan has been involved in the shifting mortality regime for several years now. These findings are thus consistent with those of the recent studies listed above. In particular, Cheung et Robine (2007), Cheung *et al.* (2008), and Thatcher *et al.* (2010) also relied on $SD(M+)$ to measure variability of adult lifespan and found that it levelled-off or changed only slightly in Japan since the 1990s, while M continued to increase.

However, our study is the first one to clearly reveal that the long-lasting upward trend for M slowed down substantially among Japanese men and has levelled-off at about 90.9 years among Japanese women since the early 2000s. These new results were unexpected. Could they indicate that women and men in Japan are approaching longevity limits in terms of modal lifespan? Given the long-term stability of historical increase in the modal age at death in this country, it would be prudent to continue to foresee future increases among both women and men. The preciseness of the P-spline method will be an important asset for monitoring the situation closely in Japan in upcoming years.

More evidence towards the shifting mortality regime was found among Canadian and French women recently, but their male counterparts are still involved in the compression of mortality scenario. Of all four countries studied, the USA definitely revealed the most worrying picture for the latest decade. Indeed, for several consecutive years in that timeframe, US women and men have both recorded important declines in their modal age at death. These findings are in

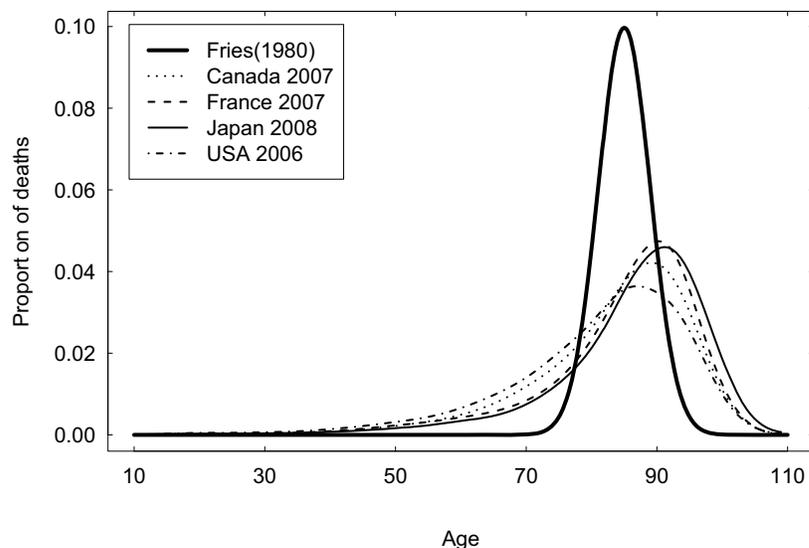
line with those of Canudas-Romo (2008) on the modal age at death in the USA for both sexes combined. Our analysis also reveals that women and men in the USA systematically exhibited higher levels of variability of age at death above the mode compared to the other countries. The explanation for these results is not simple, but the large body of literature documenting socioeconomic inequalities prevailing in the USA and major differentials with respect to health plan coverage, health care access, and health care utilization (Murray *et al.*, 2006) definitely account for some part of it. An analysis of changes in the age-at-death distribution by socioeconomic subgroup could improve our current understanding of mortality dynamics among adults in the USA.

Thus, although Canada and the USA are neighbouring countries, the findings for the former, especially over the last decade, are closer to those of France and Japan. Canada's publicly funded health care system which offers universal coverage and provides medical health care services on the basis of need rather than on the ability to pay is likely part of the multifaceted explanation for this finding.

In conclusion, when Fries (1980) presented his theory on compression of mortality, he predicted that premature mortality would eventually be eliminated and that deaths would then follow a normal distribution with an average (modal) age at death of 85 years and a standard deviation of 4 years. Thirty years later, Robine et Cheung (2009) pointed out that the current picture actually looks very different (see Figure 3.6). Indeed, the modal age at death lies well above 85 years in several low mortality countries, at least among women, and their age-at-death distributions show much greater levels of variability of age at death. It seems very unlikely that human populations will ever reach the low level of variability hypothesized by Fries. Furthermore, unlike Fries had foreseen, the end of the general scenario of compression of mortality is not an end in itself. The shifting mortality regime has indeed received support following analysis on a growing number of low mortality countries. Whether ongoing increases in longevity in these low mortality countries are paralleled with postponed diseases, functional limitations, and disability remains an open question. Mixed findings from recent studies (Christensen

et al., 2009; Robine *et al.*, 2009) calls for further research on the relationship between longevity and health.

FIGURE 3.6. Comparison between Fries's predicted age-at-death distribution and empirical smoothed age-at-death distributions among women in four low mortality countries



Source : Fries (1980) and HMD (2010)

3.6. ACKNOWLEDGEMENTS

This project was funded by the Social Sciences and Humanities Research Council of Canada. Part of the research was conducted while the first author visited the Statistical Demography Laboratory at the Max Planck Institute for Demographic Research. The authors are grateful to Jutta Gampe for her helpful suggestions and support, and to Carlo G. Camarda for sharing his R routines performing P -spline smoothing in a Poisson regression setting.

3.7. APPENDIX

In this appendix, we provide an introduction to P-splines which are used in this paper to smooth mortality data. In a nutshell, the P-spline method is a nonparametric approach which combines the concepts of B-spline and penalized likelihood. B-splines provide flexibility which leads to an accurate fit of the data, while the penalty, which acts on the coefficients of adjacent B-splines, ensures that the resulting fit behaves smoothly. Comparison between HMD life table age-at-death distributions and P-spline smoothed density functions for Japan are presented in the last section of the appendix.

3.7.1. From B-splines to P-splines

The term *B-spline* is short for basis spline and as splines in general, B-splines are made out of polynomial pieces that are joined together at points called *knots*. The degree of the B-splines (e.g., degree 1 : linear, degree 2 : quadratic, degree 3 : cubic, and so on) is set by the user and given by the degree of the polynomial pieces used to build them.

What makes B-splines attractive is that each B-spline is non-zero on a limited range of the interval over which the smoothing procedure is taking place. This also means that in any given point of the interval, only a few B-splines are non-zero, thus offering great *local* control in the resulting fit. Increasing the number of knots in the interval will increase the amount of B-splines and thus enhance the ability to capture variation in the data.

B-spline basis are well-suited for smoothing observed data points, say (x_i, z_i) , $i = 1, \dots, n$. Figure 3.7 shows an example of a B-spline basis, which includes 8 equally-spaced B-splines of degree 3, namely cubic B-splines. The basis matrix \mathbf{B} associated with this B-spline basis is defined as

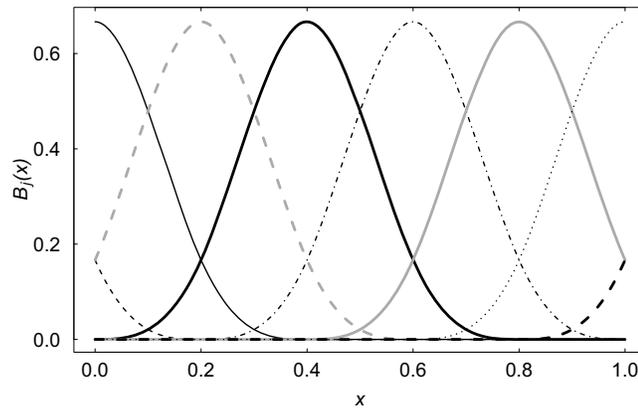
$$\mathbf{B} = \begin{bmatrix} B_1(x_1) & B_2(x_1) & \dots & B_8(x_1) \\ B_1(x_2) & B_2(x_2) & \dots & B_8(x_2) \\ \vdots & \vdots & \vdots & \vdots \\ B_1(x_n) & B_2(x_n) & \dots & B_8(x_n) \end{bmatrix},$$

where $B_j(x_i)$, $j = 1, \dots, 8$ denotes the value at x_i of the j th cubic B-spline. A fitted curve \hat{z} to observed data points (x_i, z_i) is then expressed as $\hat{z}(x_i) = \sum_{j=1}^8 \hat{a}_j B_j(x_i)$, where $(\hat{a}_1, \hat{a}_2, \dots, \hat{a}_8)^\top$ is the vector of estimated regression parameters. Each of these regression parameters is associated to one B-spline in the basis and determines its height. More generally, we have

$$\hat{z} = \mathbf{B}\hat{\mathbf{a}}, \quad (3.5)$$

meaning we are still in the well-known framework of classical regression.

FIGURE 3.7. B-spline basis containing 8 cubic B-splines with equally-spaced knots



Source : Author's calculations

The task of choosing the optimal number of knots and their position in the interval $[\min_i x_i; \max_i x_i]$ is a rather complex one. With the P-spline method, this task is avoided. Indeed, inspired by the work of O'Sullivan (1988), Eilers et Marx (1996) developed P-splines, which combines B-splines and difference penalties on the estimated coefficients of adjacent B-splines. The idea behind this approach is to use a relatively generous amount of equally-spaced knots on the interval, and to apply a penalty on adjacent regression coefficients to ensure smooth behaviour and avoid over-fitting the data.

3.7.2. The penalized likelihood function for P-splines

With the P-spline approach, the estimated vector of coefficients $\hat{\mathbf{a}}$ given in equation (3.5) results from a maximization of the following penalized log-likelihood

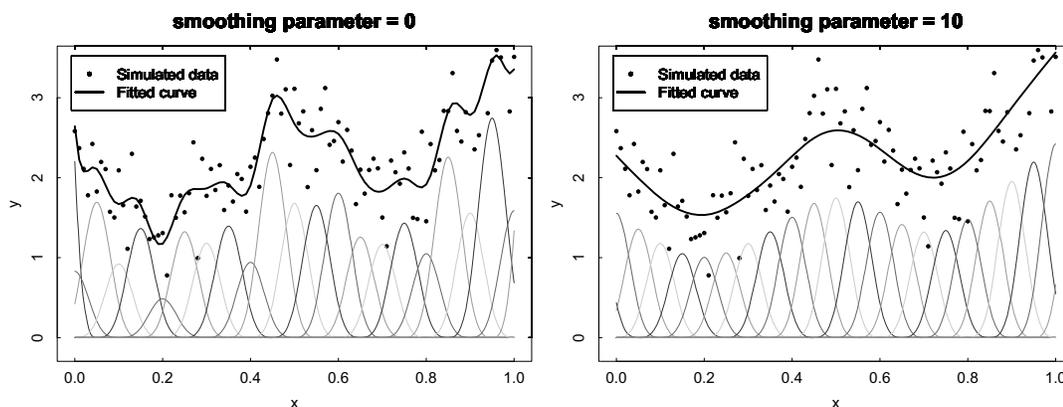
function :

$$l^* = l(\mathbf{a}; \mathbf{B}; \mathbf{z}) - \frac{1}{2} \mathbf{a}' \mathbf{P} \mathbf{a}. \quad (3.6)$$

The first term on the right-hand side of this equation corresponds to the usual log-likelihood function for a linear model. The second term is a penalty term which forces the estimated coefficients of adjacent B-splines to vary smoothly. The trade-off between parsimony – i.e., smoothness – and accuracy of the model is controlled by a smoothing parameter included in the penalty matrix \mathbf{P} .

The role of the penalty term in equation (3.6) is illustrated on simulated data in Figure 3.8. The left panel shows the results obtained when the penalty is omitted. In such case, the fitted curve is very erratic because the estimated coefficients of adjacent B-splines vary sharply from one another. Indeed, from the various B-splines shown at the bottom of the illustration, we note several sudden changes in their heights. The right panel of Figure 3.8 shows a fitted curve that is much smoother because the penalty is taken into account and this forces the estimated coefficients of adjacent B-splines to vary smoothly. Thus, the height of neighbouring B-splines can no longer change abruptly.

FIGURE 3.8. Unpenalized (left panel) and penalized (right panel) regression using simulated data and 23 equally-spaced cubic B-splines



Source : Author's calculations

3.7.3. Smoothing mortality data with P-splines

In the context of mortality data, the penalized log-likelihood function to maximize in order to find $\hat{\mathbf{a}}$ involved in equation (3.3) specifically corresponds to

$$\begin{aligned} l^* &= l(\mathbf{a}; \mathbf{B}; \mathbf{d}) - \frac{1}{2} \mathbf{a}' \mathbf{P} \mathbf{a} \\ &= l(\mathbf{a}; \mathbf{B}; \mathbf{d}) - \frac{1}{2} \lambda \mathbf{a}' \mathbf{D}' \mathbf{D} \mathbf{a}. \end{aligned} \quad (3.7)$$

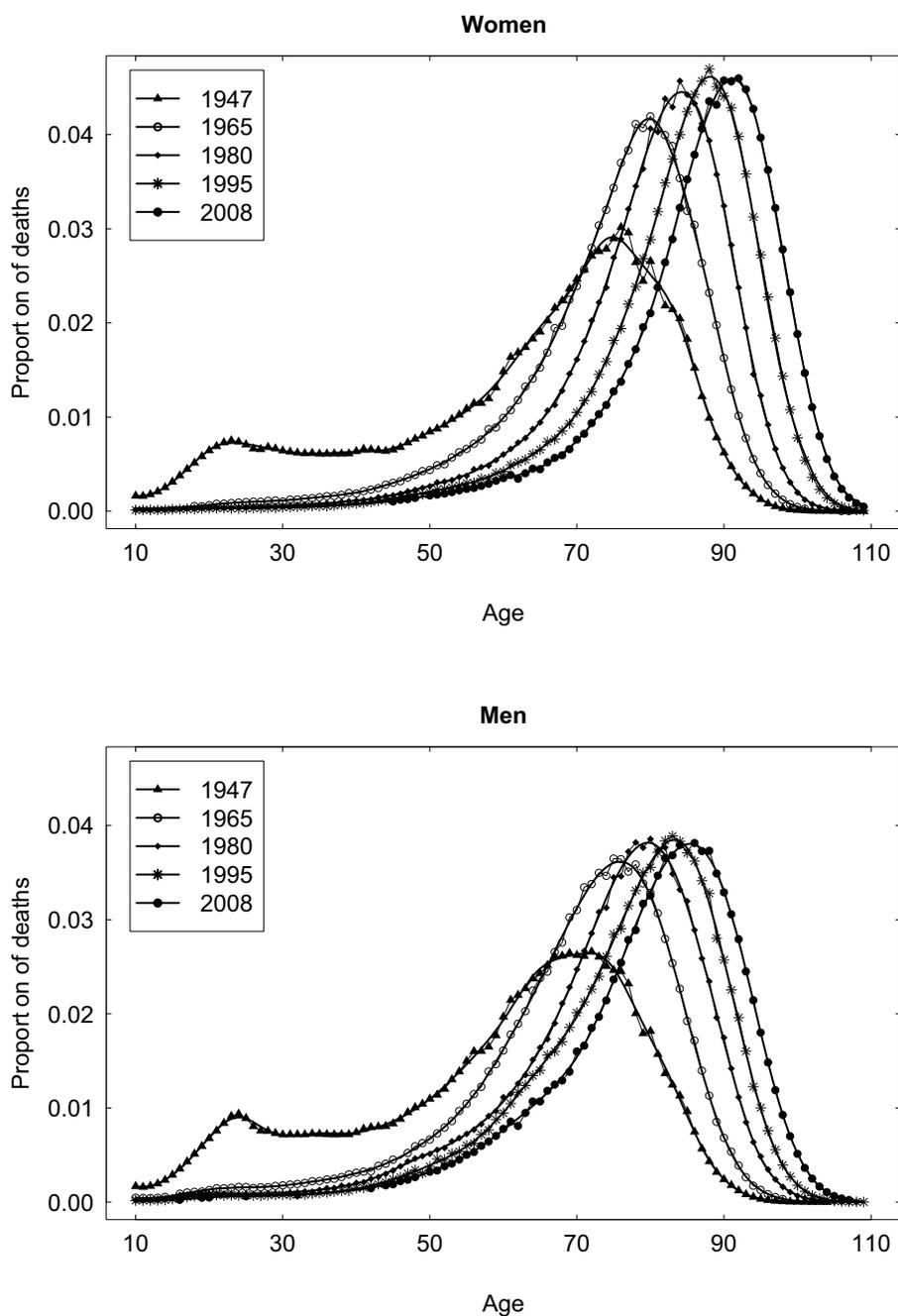
In the latter equation, $l(\mathbf{a}; \mathbf{B}; \mathbf{d})$ is the usual log-likelihood function for a generalized linear model and $\mathbf{P} = \lambda \mathbf{D}' \mathbf{D}$ is the penalty matrix. For any given smoothing parameter λ and difference matrix \mathbf{D} , the maximization of equation (3.7) can be solved with a penalized version of the iteratively reweighted least squares algorithm for the estimation of generalized linear models (Nelder et Wedderburn, 1972) (see also Currie *et al.* (2004) and Camarda (2008) for further details).

In this paper, we used a second order – i.e., quadratic – difference matrix \mathbf{D} such that $\mathbf{a}' \mathbf{D}' \mathbf{D} \mathbf{a} = (a_1 - 2a_2 + a_3)^2 + (a_2 - 2a_3 + a_4)^2 + \dots$, and the smoothing parameter λ was selected according to the Bayesian Information Criterion, as suggested by Currie *et al.* (2004). Regarding the B-spline basis, we used cubic B-splines and one knot per every five data point on the age interval.

3.7.4. Comparison between HMD life table age-at-death distributions and P-spline smoothed density functions for Japan

This last appendix section provides sex-specific comparisons between life table age-at-death distributions and smoothed density functions for Japan between 1947 and 2008 (see Figure 3.9). The life tables were extracted from the HMD and the smoothed density functions were computed with equation (3.4).

FIGURE 3.9. Comparison between life table deaths distributions (overplotted points and lines) downloaded from the HMD and smoothed density functions (plain solid lines) resulting from nonparametric P-spline estimations of Poisson regression models, Japan, 1947 to 2008



Source : As for Figure 1

Chapitre 4

Regional disparities in Canadian adult mortality : A comparative study based on smoothed mortality surfaces and age-at-death distributions

Nadine Ouellette*, Robert Bourbeau*, and Carlo G. Camarda†

*Department of Demography, Université de Montréal

†Max Planck Institute for Demographic Research, Rostock, Germany

ABSTRACT

This paper examines adult mortality differentials in Canada between 1930 and 2007 at the provincial level, thanks to the Canadian Human Mortality Database and the flexible smoothing P-spline method in two-dimensions well-suited to the study of small populations. Province-specific smoothed mortality surfaces by sex are first used to highlight converging and diverging adult mortality patterns. Old-age mortality disparities are then addressed by monitoring changes in adult smoothed age-at-death distributions in each province. Our analysis reveals that provincial disparities in adult mortality in general and among the elderly population in particular are substantial in Canada. Moreover, based on the modal age at death and the standard deviation of ages at death above the mode, provincial disparities at older ages have barely reduced over time, despite the great mortality improvements in all provinces since the early 20th century. In the last few years studied, evidence of the shifting mortality regime was found among females in most Western and Central provinces while all males were still boldly engaged in an old-age mortality compression regime.

Keywords : Adult mortality, Old-age mortality compression, Canadian provinces, P-spline smoothing

Article soumis à la revue *Canadian Studies in Population*

4.1. INTRODUCTION

As in many low mortality countries around the globe, the 20th century has been an important milestone in the evolution of epidemiological conditions in Canada. With the fall in infant mortality, infectious and parasitic diseases, and more recently cardiovascular diseases and cancer, life expectancy at birth rose considerably. Indeed, total life expectancy at birth in Canada was about 47 years in 1901 (Bourbeau *et al.*, 1997) and reached almost 81 years in 2007 (Canadian Human Mortality Database, 2010), yielding a spectacular increase of more than 30 years in about a century. Over this period, Canadian females have systematically enjoyed higher life expectancy at birth than males. However, the sex differential in life expectancy peaked at about 7 years in the 1970s and has been narrowing ever since (Trovato et Lalu, 1995, 1998; Nault, 1997; Manuel et Hockin, 2000; Zanfongnon, 2008). In terms of regional mortality conditions in Canada, all provinces and territories recorded substantial mortality progress since the beginning of the 20th century. Previous studies have shown that these mortality improvements tended to reduce disparities that prevailed among them (Field, 1980; Wilkins, 1980; Adams, 1990; Nault, 1997; Manuel et Hockin, 2000; Prud'homme, 2007). In spite of that, regional mortality differentials persist in Canada¹. For example, long-standing geographical disparities in favor to provinces in the Western part of Canada compared to those in the East continue to exist. In fact, recent findings (Manuel et Hockin, 2000; Prud'homme, 2007) reveal that since the end of the 1980s, the general east-west gradient might have clarified further into an east-central-prairies-west gradient, thus confirming that regional mortality disparities are still worth documenting in Canada.

When the probability to survive to older ages increases over time within human populations, life expectancy at birth rises, and the shape of the survival curve usually progressively becomes more rectangular. This phenomenon, well-known as the *rectangularization of the survival curve*, is associated with a reduction in the variability of age at death (Wilmoth et Horiuchi, 1999), commonly referred to as *compression of mortality*. Indeed, when compression of mortality is at work,

deaths are concentrated into a shorter age interval over time, the dispersion of the age-at-death distribution is reduced, and the downward slope of the survival curve becomes steeper, resulting in a more rectangular shape. An abundant body of demographic and epidemiology literature has already focussed on the topics of rectangularization of the survival curve and compression of mortality, providing evidence of these phenomenon in several low mortality countries and regions (Fries, 1980; Myers et Manton, 1984a,b; Nagnur, 1986; Manton et Tolley, 1991; Hill, 1993; Eakin et Witten, 1995; Nusselder et Mackenbach, 1996, 1997; Paccaud *et al.*, 1998; Wilmoth et Horiuchi, 1999; Kannisto, 2000, 2001, 2007; Lynch et Brown, 2001; Robine, 2001; Martel, 2002; Martel et Bourbeau, 2003; Cheung *et al.*, 2005, 2008, 2009; Cheung et Robine, 2007; Ouellette et Bourbeau, 2009; Thatcher *et al.*, 2010).

These studies also suggest that roughly up until the 1950s, important mortality reductions among infants, children, and even young adults led to a strong compression of the overall age-at-death distribution in the various countries and regions over time. Afterwards however, this overall compression of mortality slowed down substantially even though important mortality gains started to be recorded consistently among older adults (Kannisto *et al.*, 1994; Jeune et Vaupel, 1995). As pointed out by Thatcher *et al.* (2010), these findings have encouraged researchers to distinguish *old-age* mortality compression from *overall* mortality compression, and to study changes in the age-at-death distribution at *older ages* and over the *entire age range* separately. Moreover, the adult modal age at death² (referred to hereafter as the modal age at death) and the variability of deaths around this modal age emerged as an important set of tools to summarize and monitor changes in the age-at-death distribution at older ages over time (Kannisto, 2000, 2001, 2007; Cheung *et al.*, 2005, 2008, 2009; Cheung et Robine, 2007; Canudas-Romo, 2008, 2010; Ouellette et Bourbeau, 2009; Thatcher *et al.*, 2010). Indeed, unlike life expectancy at birth, which is highly sensitive to changes in mortality among infants and children, the modal age at death is solely influenced by adult mortality and

consequently much more sensitive to changes occurring among the elderly population (Kannisto, 2001; Horiuchi, 2003; Cheung et Robine, 2007; Canudas-Romo, 2010).

Rectangularization of the survival curve and compression of mortality in Canada has already been explored in a few studies. Indeed, the work by Nagnur (1986) focuses on the Canadian mortality experience from 1921 to 1981 and reveals that rectangularization of the survival curve did occur in this country during these six decades. Martel (2002) and Martel et Bourbeau (2003) take a close look at the situation in the province of Quebec between 1921 and 2000. They demonstrate that rectangularization of the survival curve and overall compression of mortality was underway during that period, although the latter occurred at a slower pace since 1960. Ouellette et Bourbeau (2009) focus on changes in adult age-at-death distributions and old-age mortality compression in the following low mortality countries : Canada (1921–2006), the U.S. (1945–2006), France (1920–2007), and Japan (1947–2007). They find that similarly to Japanese and French females, old-age mortality compression no longer seems to be occurring among Canadian females for the most recent years. Instead, the female adult age-at-death distribution in Canada has been shifting to higher ages over time while maintaining an intact shape recently, thus providing additional support to the *shifting mortality scenario* first described by Kannisto (1996) and Bongaarts et Feeney (2002, 2003), and furthered by Bongaarts (2005). Canadian males were however still involved in reducing variability of deaths at advanced ages.

4.2. STUDY OBJECTIVES

Given the absence of studies on regional variations in Canada with respect to the rectangularization of the survival curve and the compression of mortality, the present work aims notably at shedding light on these topics at the level of Canadian provinces. In that sense, it can therefore be seen as a continuation of previous work by Nagnur (1986), Martel (2002), Martel et Bourbeau (2003), and Ouellette et Bourbeau (2009), especially the latter since our focus here is

on adult mortality differentials. Furthermore, our approach also rests upon the flexible smoothing P-spline method : precisely, we provide a generalization of the approach introduced by Ouellette et Bourbeau (2009) in order to make it particularly well-suited to the study of small populations. Thus, the first objective of this paper is to highlight general converging and diverging adult mortality patterns among Canadian provinces using sex-specific smoothed mortality surfaces for each province. This will allow us to gather high-level information on regional disparities among Canadian adults. Then, the second objective is to investigate disparities among the elderly populations of the various provinces and to compare their old-age mortality compression situations. Two-dimensional smoothed adult age-at-death distributions by sex will thus be used to monitor changes in the modal age at death and in the variability of deaths above this modal age over time in each Canadian province.

4.3. DATA

The data for this study were taken from the Canadian Human Mortality Database (CHMD, 2010), which gathers detailed Canadian mortality and population data at the provincial and territorial level. This database is a unique source of information to study regional mortality patterns in great detail and accuracy in Canada. Indeed, as a satellite project of the well-known and widely used Human Mortality Database - HMD (2010) which includes 37 countries and areas, the CHMD also rests upon sophisticated demographic methods and high quality standards³.

About data collection in Canada, there are currently thirteen systems of vital events registration in the country, that is one for each province and territory. Most of them have been registering births and deaths since the 1920s, except for Newfoundland and Labrador which joined the Canadian Confederation in 1949 and the territories where registration coverage before 1950 tends to be incomplete. While the various provincial and territorial administrations are in charge of collecting the data, Statistics Canada is responsible for publishing them.

Although data included on the CHMD relies on thirteen different administrations, each with their own set of principles and practices subject to change over the years, there is no evidence of systematic discrepancies large enough to influence trends over time or between provinces and territories (Bourbeau *et al.*, 2010). Indeed, registration problems are rather punctual, well known, and have been taken care of in the CHMD. Furthermore, missing or incorrect information on deaths also arises, but concerns less than 1% of deaths each year (Bourbeau et Ouellette, 2010). In spite of that, Kannisto *et al.* (1994) classified Canada as a country where data were of inferior quality compared to the Nordic countries, several European countries and Japan given data availability and age misstatement. However, Kannisto's analysis of Canadian data rested upon published data exclusively⁴, which notably included opened age intervals starting at age 90 (Bourbeau et Lebel, 2000). These data were insufficient to draw the entire true picture in terms of data quality and mortality measurement in Canada. The study by Bourbeau et Lebel (2000) uses more detailed data and suggests that data quality in Canada is quite good until at least age 100. Further research on the validation of ages at death among French-Canadian centenarians born in Quebec demonstrates that even at the oldest ages, data quality is very good (Beaudry-Godin, 2010).

For the purpose of this paper, we extracted from the CHMD observed deaths counts and exposure data by sex, single year of age, and single calendar year for each Canadian province. Data below age ten were not extracted given our aim to focus on adult mortality. Furthermore, data covers the 1930-2007 period⁵, except for the province of Newfoundland and Labrador for which data starts in 1949. The Canadian territories, namely Yukon, Northwest Territories and Nunavut, were excluded from our analysis notably because of their very small population size. Indeed, there are so few deaths occurring each year for these territories⁶ that even with a smoothing method well-suited for handling small populations, it would have been difficult to conduct an analysis by sex on these data. Furthermore, above the population size limitation, Canadian territories have systematically exhibited important mortality differentials with Canadian provinces over the years,

mainly because of their distinct socioeconomic and cultural environment. Mortality and health disparities between Canadian territories and provinces may be substantially lower today than in the 1950s but remain nonetheless worrying (Prud'homme, 2007; Veugelers *et al.*, 2001). Canadian territories would thus deserve to take part in a study that would specifically address their mortality developments over the second half of the 20th century.

4.4. METHODS

4.4.1. Smoothing mortality data with P-splines

Recently, Ouellette et Bourbeau (2009) introduced the use of a one-dimensional smoothing approach, usually known as the P-spline method (Eilers et Marx, 1996), to monitor with great precision changes in the adult age-at-death distribution over time in low mortality countries. In the present paper, we rely on the two-dimensional version of this method in order to obtain sex-specific smoothed mortality surfaces and adult age-at-death distributions for each of the ten Canadian provinces. The P-spline method in two-dimensions is particularly useful when dealing with small populations such as Canadian provinces because it models mortality change over age and over time simultaneously. Therefore data information on neighboring ages and years is consistently used by the model, and punctual variations due to small numbers of deaths or exposures are less likely to distort the outcome (Camarda, 2010).

In comparison to other methods for modelling mortality over age and over time, such as the Brass method (Brass, 1971), the age-period-cohort model (Clayton et Schiffers, 1987), and Lee-Carter approaches (Lee et Carter, 1992; Brouhns *et al.*, 2002), the P-spline method does not make any rigid assumption about the functional form of the mortality surface (Camarda, 2008). This leads to a very fine expression of the underlying mortality patterns over age and time described by the data.

In a nutshell, the P-spline method is a nonparametric approach which combines the concepts of B-spline and penalized likelihood. The idea behind the

method is that B-splines provide flexibility which leads to an accurate fit of the data, while the penalty, which acts on the coefficients of adjacent B-splines, ensures that the resulting fit behaves smoothly. The term *B-spline* is short for basis spline and as splines in general, B-splines are made out of polynomial pieces that are joined together at points called *knots*. The degree of the B-splines (eg. 1 : linear, 2 : quadratic, 3 : cubic, etc.) is given by the degree of the polynomial pieces used to build them. In this paper, we used cubic B-splines. What makes B-splines attractive is that each B-spline is nonzero on a limited range of the interval over which the smoothing procedure is taking place. This also means that in any given point of the interval, only a few B-splines are nonzero, thus offering great *local* control in the resulting fit. Increasing the number of knots in the interval will increase the amount of B-splines and enhance the ability to capture variation in the data. With the P-spline method, knots are equally spaced over the entire interval and we use a relatively large number of them, knowing that the penalty will prevent overfitting of the data by ensuring a smooth fit. The details on the P-spline method in the context of mortality data are described in Appendix A (Section 4.9).

We used the `MortalitySmooth` package written by Camarda (2009) for the statistical programming environment R (R Development Core Team, 2010) to obtain smoothed two-dimensional forces of mortality by sex and province. Indeed, this package is precisely intended to perform P-spline smoothing in one and two dimensions for mortality data. The sex-specific smoothed death rates by age and time obtained for each province will be presented on shaded contour maps (Vaupel *et al.*, 1997) in the upcoming Results section. These graphical representations will also be specifically referred to as *mortality surfaces* in the remainder of the text. Meanwhile, the next subsection explains how smoothed two-dimensional age-at-death distributions are computed and employed in this paper.

4.4.2. Computation and use of smoothed age-at-death distributions

The force of mortality, the survival function and the density function are three particularly useful functions for describing mortality distributions. They

also share the following interesting property : if one of these three functions is known, the remaining two can be uniquely determined. Thus, from sex-specific smoothed forces of mortality for each province, we can compute the corresponding smoothed density functions describing the age-at-death distributions in these populations.

Let $\mu(a, y)$ denote the force of mortality, a continuous function of age a and time y in a given population. Similarly, let $S(a, y)$ and $f(a, y)$ be respectively the survival function and density function, which are also continuous functions of age and time. Given the usual correspondence between these three functions (Klein et Moeschberger, 1997, Chap. 2), we have

$$\begin{aligned} f(a, y) &= \mu(a, y) S(a, y) \\ &= \mu(a, y) \exp\left(-\int_0^a \mu(u, y) du\right). \end{aligned} \quad (4.1)$$

Therefore, if the smoothed force of mortality $\hat{\mu}$ is known, then the corresponding smoothed density function \hat{f} describing the two-dimensional age-at-death distribution can be computed using equation (4.1) and standard numerical integration techniques.

In order to monitor changes in the central tendency and old-age dispersion of adult deaths over time and across provinces, we used the following summary measures respectively : the modal age at death and the standard deviation of individual life durations above the modal age at death. Inspired by Lexis's concept of normal life durations (Lexis, 1877, 1878), Kannisto (2000, 2001) suggested this set of measures to focus specifically on changes in the age-at-death distribution occurring at older ages. This approach with emphasis on the modal age at death has drawn much attention and was elaborated further in several recent studies (Cheung *et al.*, 2005, 2008, 2009; Cheung et Robine, 2007; Canudas-Romo, 2008; Ouellette et Bourbeau, 2009; Thatcher *et al.*, 2010).

From the smoothed density function \hat{f} describing the two-dimensional age-at-death distribution, we first computed the modal age at death using

$$\hat{M}(y) = \max_a \hat{f}(a, y),$$

and then the standard deviation of individual life durations above the modal age at death (referred to hereafter as the standard deviation above the mode) using

$$\widehat{SD}(M+)(y) = \sqrt{\frac{\int_{\hat{M}(y)}^{\infty} (a - \hat{M}(y))^2 \hat{f}(a, y) da}{\int_{\hat{M}(y)}^{\infty} \hat{f}(a, y) da}}.$$

Since \hat{f} is a function of ages and years, \hat{M} and $\widehat{SD}(M+)$ are functions of years.

Given that provinces in Canada are very distinct in population size, 99% confidence intervals for $\hat{M}(y)$ and for $\widehat{SD}(M+)(y)$ were also computed for each year $y \in \{1930, 1931, \dots, 2007\}$ to facilitate comparisons across provinces. The details of the residual bootstrap method addressed by Koissi *et al.* (2006) and used in the present study to construct these confidence intervals are described in Appendix B (Section 4.10).

4.5. RESULTS

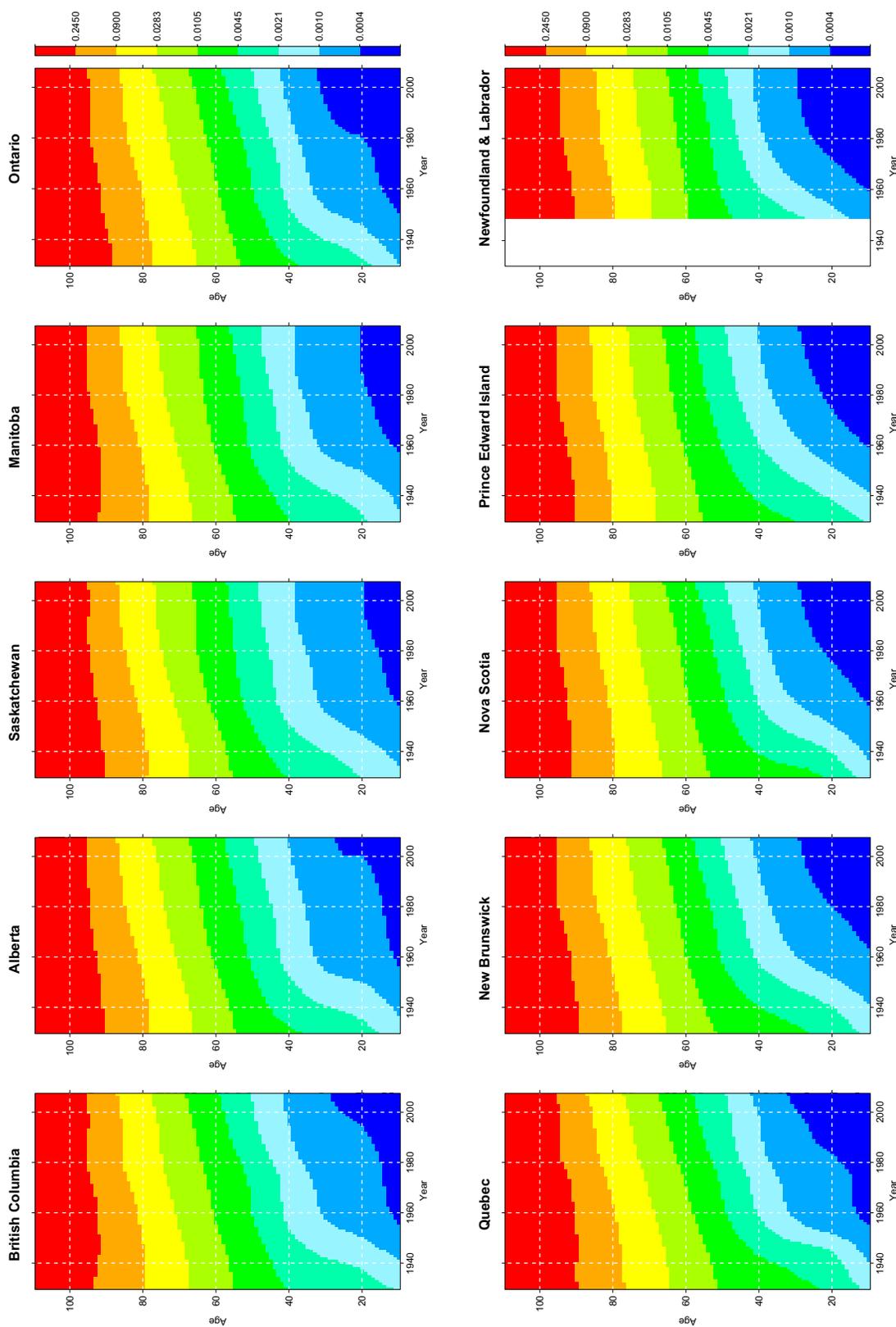
4.5.1. Uncovering provincial variations via smoothed mortality surfaces

Figures 4.1 and 4.2 show province-specific smoothed mortality surfaces for females and males respectively. To our knowledge, these are the very first smoothed mortality surfaces for Canadian provinces to be published. Each of these shaded contour maps displays smoothed death rates over age and time, and thus summarizes a great amount of information on a single graphic representation. We will see shortly that mortality surfaces provide interpretation at a glance and in the context of mortality comparisons between Canadian provinces, they are very

useful. The scales on the right-hand side of Figures 4.1 and 4.2 show how smoothed death rates were first partitioned into nine groups for each sex. Death rates belonging to the same group were then assigned the same color. For example, female death rates below 0.0004 were all plotted in dark blue on the female mortality surfaces and they represent the lowest level of mortality on the surface. Among males, this group includes all death rates below 0.001. Likewise, female death rates higher than 0.245 correspond to the highest level of mortality on the surface and were all plotted in red. Among males, this group also starts at 0.245.

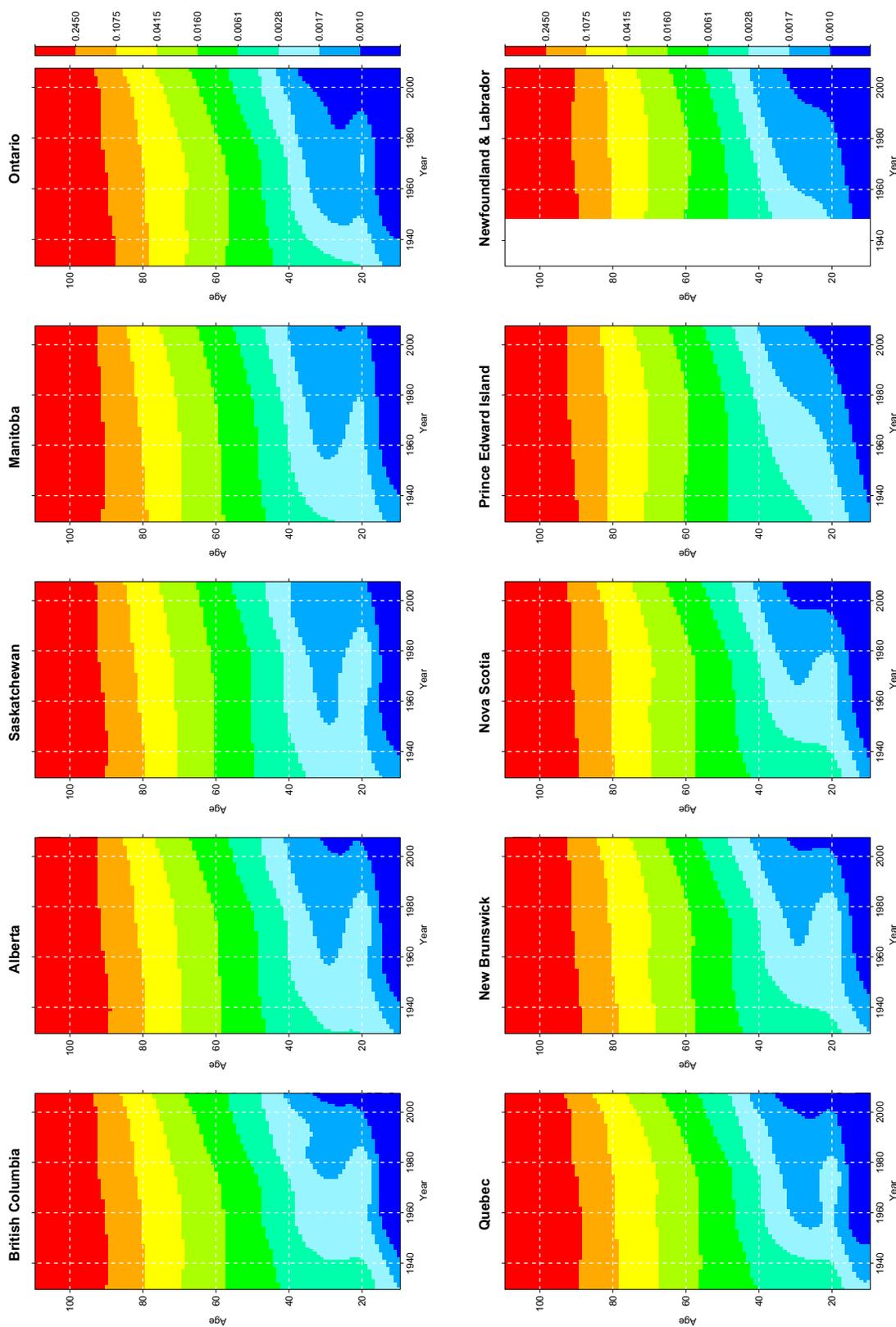
From Figures 4.1 and 4.2, mortality improvements between 1930 and 2007 at most adult ages and across all provinces for both sexes are evident. Indeed, the upward trend followed by most color bands indicates that low levels of mortality are progressively associated to a wider range of ages as the years go by. We also note that reductions in mortality at higher ages, say above 50 years old, roughly began in the 1970s among males while they were already apparent in the 1930s among females. Furthermore, in geographical terms, provinces that belong to the same great Canadian region, that is Western provinces (British Columbia, Alberta, Saskatchewan, Manitoba), Central provinces (Ontario, Quebec), and Atlantic provinces (New Brunswick, Nova Scotia, Prince Edward Island, Newfoundland and Labrador), tend to reveal similar mortality patterns. Nevertheless, substantial variations exist within these broad regions and are discussed next.

FIGURE 4.1. Province-specific smoothed mortality surfaces based on fitted death rates with two-dimensional Poisson P-splines, Females, Canada, 1930 to 2007



Note : Extra-Poisson variation (overdispersion) was detected in Ontario and accounted for in the smoothing procedure.

FIGURE 4.2. Province-specific smoothed mortality surfaces based on fitted death rates with two-dimensional Poisson P-splines, Males, Canada, 1930 to 2007



Note : Extra-Poisson variation (overdispersion) was detected in Ontario and accounted for in the smoothing procedure.

Starting with Western provinces, the situation in Saskatchewan and Manitoba differs from the one in British Columbia, and to a lesser extent in Alberta. Between 1930 and 1960, mortality at almost every adult age in Saskatchewan was the lowest of all Western provinces for each sex. However, mortality improvements have been relatively slow since the 1960s, especially among females in the last two decades where mortality reductions below age 40 are no longer perceptible on the smoothed mortality surface. The pictures for females and males in Manitoba are highly similar to those of Saskatchewan, although slightly more moderate. In British Columbia, female mortality below age 40 during the 1930s was higher than in the other Western provinces, but mortality reductions at these ages occurred at a very rapid pace until 1960. This initial mortality disadvantage at young adult ages in the 1930s in British Columbia is even more pronounced among males, but had almost completely disappeared by 1950. Furthermore, strong reductions in mortality at ages over 40 have been consistently recorded among males in British Columbia since the 1970s. From various points of view, the female smoothed mortality surface for Alberta looks like the one for British Columbia. This applies to males as well, except at ages below 40 where the smoothed mortality surface for Alberta is more similar to the ones for males of Manitoba and Saskatchewan.

Within Central provinces, the level of mortality in the 1930s was clearly higher in Quebec than in Ontario at every adult age among females and below age 60 among males. Thanks to an intensive phase of reductions in mortality at ages below 50 during the 1930s and 1940s, plus a sustained high rate of progress at ages above 50 throughout the study period, Quebec females have now almost caught up with those of Ontario. Strong reductions in mortality at younger adult ages during the 1940s also occurred among Quebec males and led them closer to their male counterparts in Ontario. Finally, similarly to British Columbia males, Quebec and Ontario males have been equally good at lowering substantially mortality at ages above 50 since the 1970s.

Among Atlantic provinces, the smoothed mortality surfaces for New Brunswick and Nova Scotia are very similar for each sex. Still, New Brunswick shows

a slightly higher female mortality above age 50 compared to Nova Scotia from 1930 to the 1970s. Similarly, males in New Brunswick experienced higher mortality above age 70 in the first decades of the period under study, but managed to record lower mortality levels than Nova Scotia at these ages during the 1970s. Prince Edward Island seemed to enjoy the lowest female and male mortality of all Atlantic provinces at most adult ages between 1930 and 1960, but the scarce information (due to the small size of this population) on which these results are based calls for caution here. Reductions in mortality afterwards, particularly at ages above 60, are much slower than in New Brunswick and Nova Scotia for both sexes. Newfoundland and Labrador also shows low rates of mortality progress in the most recent decades at ages above 60 compared to New Brunswick and Nova Scotia, especially among females.

Thus, adult mortality differentials do exist among Canadian provinces and have been roughly revealed by the smoothed mortality surfaces described above. The next subsection delves one step further into these differentials and examines changes in smoothed province-specific adult age-at-death distributions over time. Given the summary measures favored to monitor these changes, the modal age at death and the standard deviation above it, focus is on the elderly population specifically rather than on adults in general.

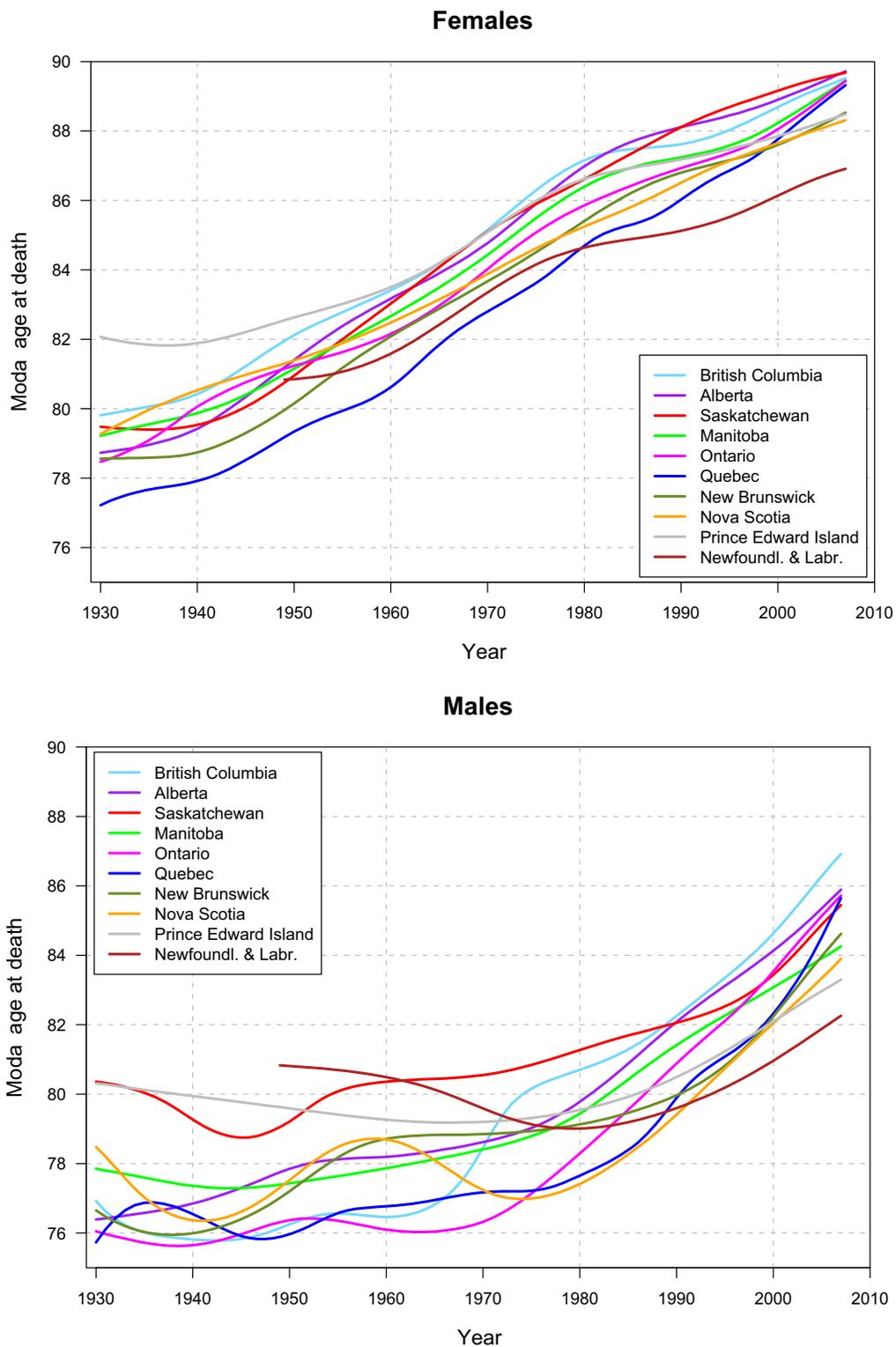
4.5.2. Summarizing changes in smoothed age-at-death distributions over time across provinces

Figure 4.3 presents estimated time-trends in the modal age at death by province for each sex computed from their respective two dimensional age-at-death distribution. Among females, all provinces showed an upward trend throughout most of the period under study. These upward trends were in fact almost linear between 1940 and 1980, and in the last two decades. In the early 1930s, female modal age at death estimates for most provinces lied between 78 and 80 years⁷. At the turn of the 21st century, the female modal age at death values had increased by about 10 years in most provinces and ranged from 87 to 90 years. Among

Western provinces, females displayed very similar results from 1930 to 2007. Indeed, 99% confidence intervals for their modal age at death estimates reveal that significant differences mainly occurred between Saskatchewan and Manitoba during the 1990s (see Table 4.1 in Appendix B, Section 4.10, for 99% confidence intervals results). Throughout the period under study, females in Western provinces enjoyed high modal age at death outcomes. In fact, from the 1960s until very recently, the confidence intervals yield that female modal age at death in British Columbia, Alberta and Saskatchewan was significantly higher than in both Central provinces and most Atlantic provinces. After 2000 however, modal age at death differences between Western provinces and Central provinces are no longer statistically different. This is particularly interesting for Quebec females, who consistently recorded significantly lower modal age at death results than most of their female counterparts between 1930 and 1990. In contrast, the situation among females in Atlantic provinces has deteriorated over time. Moreover, in the last few years under study, the female modal age at death in these provinces was significantly lower than in any other Western or Central province.

Among males, sustained upward trends in the modal age at death did not start much before the late 1960s or even early 1970s in Canadian provinces. Indeed, from 1930 up to the 1960s, the modal age at death trend for males either stagnated or rippled in the high 70s range, probably because mortality reductions at ages above the mode, essential for its increase (Canudas-Romo, 2010), were limited during that period of time. Afterwards however, substantial increases in the male modal age at death were recorded in all provinces and 2007 estimates lied between 82 and 87 years. The case of British Columbia males stands out because in the early 1960s, their modal age at death estimates were still among the lowest, but since 1990, they have consistently been the highest. According to the 99% confidence intervals for male modal age at death estimates in Table 4.1, British Columbia results have been significantly higher than in each Atlantic province since about 1980 and Central provinces between 1990 and 2005. Within Western provinces, significant differences are scarce. Indeed, they mainly occurred between Saskatchewan and Manitoba during the 1960s and 1970s, and between

FIGURE 4.3. Modal age at death by province and sex estimated from two-dimensional smoothed age-at-death distribution, Canada, 1930 to 2007

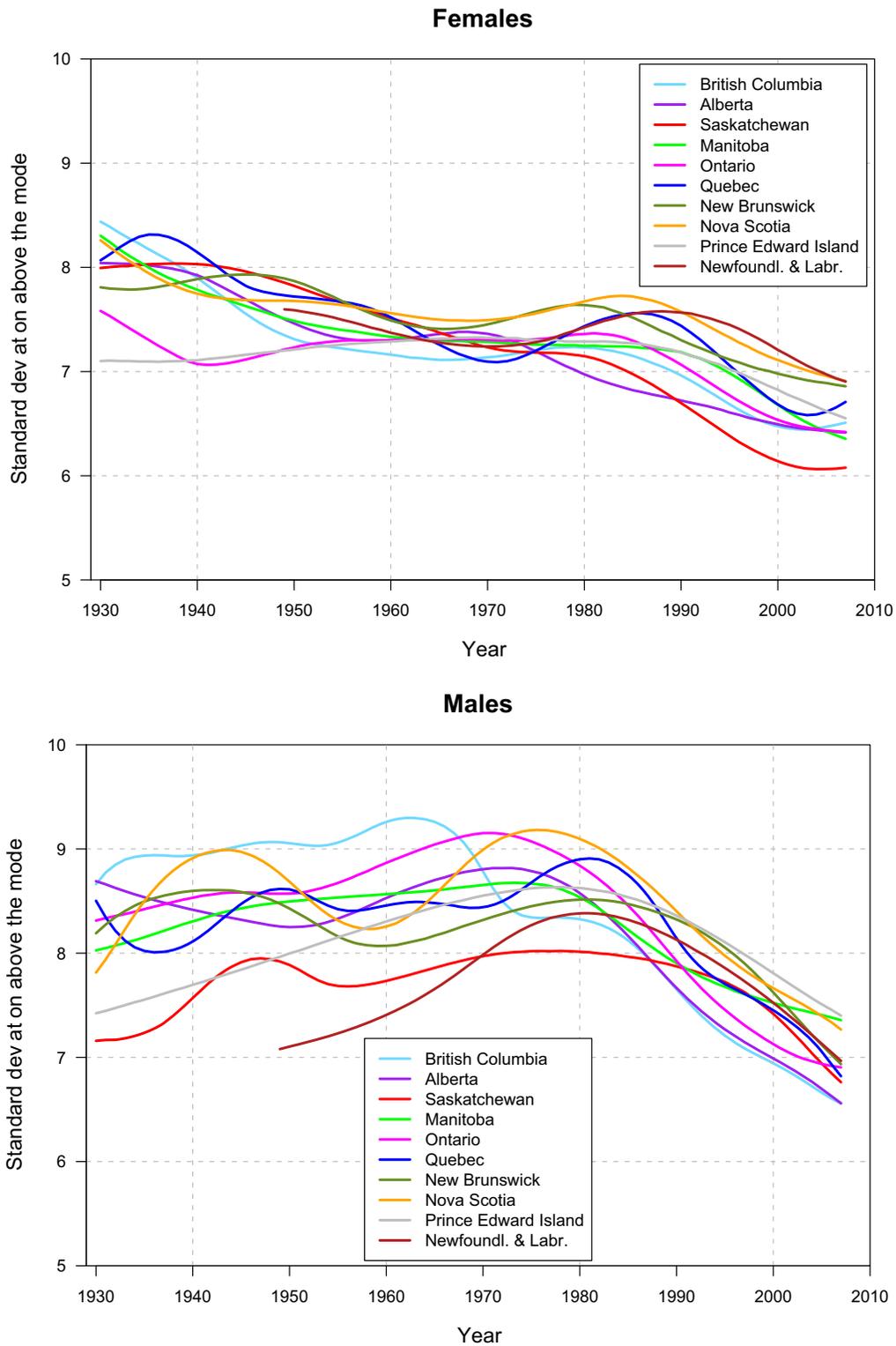


British Columbia and Manitoba in the latest years. Similarly to British Columbia, the Central provinces were also displaying some of the lowest modal age at death estimates for males between 1930 and the 1960s and made impressive progress afterwards. Among Atlantic provinces, despite increases in the male modal age at death since the 1970s, Newfoundland and Labrador, Prince Edward Island, and Nova Scotia have been showing the lowest results of all Canadian provinces since 2000. However, differences with New Brunswick, Quebec, Manitoba, and Saskatchewan were not systematically significant.

Figure 4.4 displays sex- and province-specific time-trends in the standard deviation above the modal age at death computed from two dimensional age-at-death distributions. Among females, each province recorded lower results at the end of the period covered by this study than at the beginning. In other words, old-age compression of mortality has occurred between 1930 and 2007 among these females. However, the trends did not decline steadily over the entire period. Indeed, the 1960s and 1970s were years of slow decline for most provinces. Furthermore, since 2000, the female standard deviation above the mode in Western provinces (except Manitoba) and Central provinces has either decreased at a much slower pace than in previous years, stopped decreasing or even increased slightly. Evidence from Atlantic provinces instead shows that old-age compression of mortality continues among these females because their standard deviation above the mode did not stop declining in recent years. This is perhaps related to the fact that the level of female old-age mortality compression has remained higher in most Atlantic provinces than in other Canadian provinces since 2000. Indeed, the 99% confidence intervals for the female standard deviation above the mode estimates (see Table 4.2) reveal that while significant differences across provinces were rather scarce between 1930 and 1970, they became more common afterwards, especially between Atlantic and Western provinces.

Among males, we notice that the province-specific standard deviation above the mode estimates were not necessarily lower in 2007 than in 1930. In other words, unlike among Canadian females, old-age compression of mortality did not

FIGURE 4.4. Standard deviation above the modal age at death by province and sex estimated from two-dimensional smoothed age-at-death distribution, Canada, 1930 to 2007



occur in all ten provinces over the period as a whole for males. However, if we limit ourselves to the period starting with the onset of the modal age at death increases for Canadian males – that is around 1970 – then it generally coincides with episodes of old-age mortality compression in all provinces. Indeed, male standard deviation above the mode estimates decreased quite steadily over time after 1970. Even in recent years, the pace of decline did not slowdown, except perhaps in Ontario. Since 1990, male standard deviation above the mode estimates in Atlantic provinces have been among the highest in Canada. In contrast, Western provinces generally enjoyed the lowest estimates for males, especially British Columbia and Alberta. Indeed, for several years in the 1990s and early 2000s, male standard deviation above the mode estimates in Prince Edward Island, and Nova Scotia were significantly higher than in British Columbia and Alberta (see Table 4.2). Differences between New Brunswick and British Columbia were also significant during most of the 1990s.

4.6. DISCUSSION

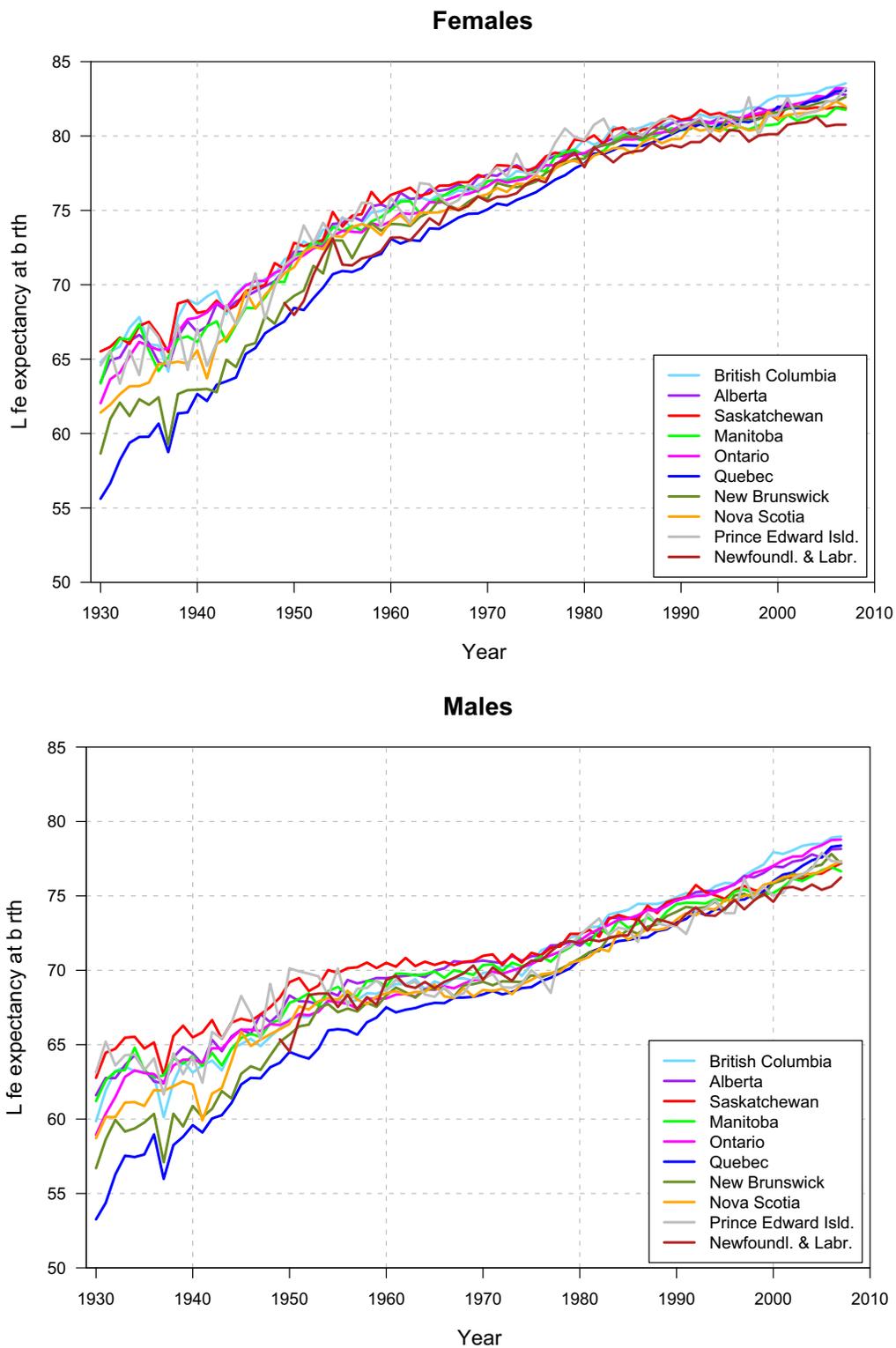
This paper explored differences in adult mortality among Canadian provinces between 1930 and 2007, and focussed notably on old-age mortality compression, a topic that has not been investigated yet at the regional level in the country. Our analysis revealed that provincial disparities in adult mortality in general and among the elderly population in particular are substantial in Canada. Moreover, these disparities have barely reduced over time, despite the great mortality improvements recorded within each of these provinces since the beginning of the 20th century.

Province-specific smoothed mortality surfaces by sex (see Figures 4.1 and 4.2) illustrated that in most recent years particularly, adult mortality decline in Newfoundland and Labrador, Prince Edward Island, Saskatchewan (males especially), and Manitoba have been slow compared to the other Canadian provinces. In contrast, females and males in British Columbia, Ontario and Quebec have been by far the most successful at lowering mortality at almost every adult age in

the last two to three decades. These findings are consistent with those of previous studies on mortality disparities in Canada (Manuel et Hockin, 2000; Prud'homme, 2007) where the relatively poor recent mortality improvements in Saskatchewan were underlined and the rise of Quebec in the provincial ranking emphasized. These studies also identified British Columbia as the new leader and reference in terms of mortality in Canada.

Time-trends results in modal age at death by province and sex (see Figure 4.3) were also in line with these observations and they further helped us appreciate the magnitude of the disparities across the provincial elderly populations. These modal ages at death trends are in fact a great addition to the usual life expectancy at birth trends (see Figure 4.5) often reported to discuss regional mortality differentials in Canada. Disparities in life expectancy at birth across Canadian provinces have been largely reduced over the 20th century because all provinces made remarkable progress in terms of infant mortality and infectious, as well as parasitic diseases during that period. However, Figure 4.5 hides the fact that old-age mortality disparities have barely been reduced among Canadian provinces between 1930 and 2007. This part of the story rather emerges from Figure 4.3 because unlike life expectancy at birth, which is very sensitive to improvements in mortality among infants and children, the modal age at death is sensitive to mortality changes occurring among the elderly population. Indeed, Figure 4.3 shows that in the last decade of our study, as the modal age at death was increasing very rapidly in Quebec and Ontario, these Central provinces were actually catching up with Western provinces, while the lag of Atlantic provinces was becoming more and more pronounced. The situation in Newfoundland and Labrador is particularly worrying and deserves to be carefully monitored in upcoming years. Thus, the long-standing geographical mortality disparities favoring provinces in the Western part of Canada compared to those in the East holds among the elderly population. However, this might be subject to change as Central provinces could perhaps take the lead in Canada in the near future.

FIGURE 4.5. Life expectancy at birth by province and sex, Canada, 1930 to 2007



The level of variability of age at death among older individuals is another great source of provincial mortality disparity in Canada (see Figure 4.4). Again, reductions in these provincial differentials over time were very scarce between 1930 and 2007. Indeed, at the end of the period under study, Canadian provinces still showed a broad palette of levels of old-age mortality dispersion. The highest levels of standard deviation above the mode tended to be recorded among provinces belonging to the Eastern part of Canada, while provinces with the lowest levels rather belonged to the Western part of Canada. Males in Manitoba are perhaps the most obvious exception, as their level of old-age mortality dispersion remains high compared to the other Western provinces.

When Kannisto (2000, 2001) introduced the modal age at death and the standard deviation above the mode to summarize and monitor changes in the age-at-death distribution at older ages, the rationale behind his set of indicators was very clear. Episodes of modal age at death increases paralleled with decreases in the standard deviation above the mode suggest that the distribution of ages at death at old ages is not solely moving towards higher ages. As Kannisto (2001) stated, “Instead, its right-hand slope is being flattened vertically as if it was meeting and invisible wall” (p.169). On the other hand, when the mode increases and the standard deviation above the mode no longer declines, the distribution of ages at death at old ages is simply sliding towards higher ages without changing its shape. Thus, each of these scenarios, respectively known as the *old-age compression of mortality scenario* and the *shifting mortality scenario*, may provide valuable indications on human longevity and limits to human life.

In the present study, evidence of the shifting mortality regime was recorded among females in most Western and Central provinces but males in all provinces were still boldly engaged in an old-age mortality compression regime. These results concur with those of Ouellette et Bourbeau (2009) where Canadian females, all provinces and territories gathered, were found to be heading towards a shifting mortality regime in upcoming years and males did not show such evidence. These findings call for further research on the health situation of the elderly populations

in Canadian provinces showing evidence of the shifting mortality regime lately. Under such regime, it is assumed that adult mortality shifts to higher ages (Bongaarts, 2005). However, whether morbidity and disability episodes also shift to higher ages or rather occupy a larger proportion of the typical lifespan remains unclear. Research on Canadian provincial disparities in terms of modal age at death and dispersion above it by causes of death should be helpful in that sense.

4.7. NOTES

¹The fact that Canada consists of ten provinces and three territories that are disparate geographically, economically, politically, culturally, and socially explains part of the regional mortality differentials recorded in the country (Trovato et Lalu, 2001; Prud'homme, 2007).

²The adult modal age at death corresponds to the adult age at which the largest number of deaths occur. In other words, it is the most common age at death among adults.

³To date, the CHMD is the only regional database to operate under the umbrella of the Human Mortality Database. The CHMD project is in fact a collaborative achievement of the *Mortality and Longevity* research team at the Department of Demography, Université de Montréal, together with demographers at the Max Planck Institute for Demographic Research and the Department of Demography, University of California at Berkeley.

⁴Access to Canadian data unpublished by Statistics Canada generally requires a special authorisation from each province and territory. For example, deaths by single year of age over the entire age range for Canada are not published by Statistics Canada, neither are deaths by cohort.

⁵Given that some minor irregularities may have occurred in the beginnings of vital registration in Canada, data from the 1921-1929 period were not included in our analysis.

⁶In 2007 for example, Yukon, Northwest Territories, and Nunavut each recorded less than 200 deaths in total (Statistics Canada, 2010, p. 20).

⁷The female modal age at death estimates for Prince Edward Island between 1930 and 1960 are higher than in any other Canadian province. However, due to the small population size of Prince Edward Island, its 99% confidence intervals for modal age at death estimates are wide

(see Table 4.1 in Appendix B, Section 4.10). Thus, despite appearances, the female modal age death for Prince Edward Island in the 1930s is not significantly higher than in British Columbia.

4.8. ACKNOWLEDGEMENTS

This project was funded by the Social Sciences and Humanities Research Council of Canada. The authors are grateful to the Council for its support.

4.9. APPENDIX A : THE P-SPLINE METHOD IN THE CONTEXT OF MORTALITY

In this appendix, we provide an overview of the P-spline smoothing method in the context of mortality data. We start by describing how the procedure works when the aim is to perform one-dimensional smoothing over ages, as discussed by Ouellette et Bourbeau (2009). Then, we present the P-spline method in two dimensions which is used in the present paper to smooth mortality data over ages and years simultaneously.

4.9.1. Smoothing in one dimension

Let d_i and e_i denote respectively the observed deaths counts and exposure data by age i for a given year in the population under study. Also, let μ_i be the force of mortality at age i .

Assuming that the force of mortality is a piecewise constant function, meaning that it is constant within each single age interval such as $\mu(x) = \mu_i$ for all $x \in [i, i + 1)$, then $d_i \sim \text{Poisson}(e_i \cdot \mu_i)$. Thus, in order to estimate μ_i , we use a Poisson regression model such that

$$\begin{aligned} \ln(\mathbb{E}[\mathbf{d}]) &= \ln(\mathbf{e} \cdot \boldsymbol{\mu}) \\ &= \ln(\mathbf{e}) + \ln(\boldsymbol{\mu}), \end{aligned}$$

where \mathbf{d} , \mathbf{e} , and $\boldsymbol{\mu}$ respectively correspond to observed deaths, exposure, and force of mortality vectors (each one including all the age-specific information). A smooth estimate of $\boldsymbol{\mu}$ is obtained by the P-spline method (Eilers et Marx, 1996) :

$$\ln(\hat{\boldsymbol{\mu}}) = \mathbf{B}\hat{\boldsymbol{\alpha}},$$

where \mathbf{B} is the B-spline basis regression matrix and $\hat{\boldsymbol{\alpha}}$ is the vector of estimated coefficients associated to each B-spline included in the basis. The vector of coefficients, $\boldsymbol{\alpha}$ is estimated according to a maximum likelihood procedure where the penalized log-likelihood function to maximize corresponds to

$$l^* = l(\boldsymbol{\alpha}; \mathbf{B}; \mathbf{d}) - \frac{1}{2}\boldsymbol{\alpha}'\mathbf{P}\boldsymbol{\alpha}. \quad (4.2)$$

The first term on the right-hand side of this equation corresponds to the usual log-likelihood function for a generalized linear model. The second term is a penalty term (\mathbf{P} is a penalty matrix), which forces the estimated coefficients of adjacent B-splines to vary smoothly. The trade-off between smoothness and fidelity of the model to the observed data is tuned by a smoothing parameter, which is selected according to the Bayesian Information Criterion (see Currie *et al.* (2004) and Camarda (2008) for further details on the penalty term and on the smoothing parameter in the context of mortality data).

4.9.2. Smoothing in two dimensions

To move from one- to two-dimensional Poisson P-spline smoothing, a new B-spline basis adapted for two-dimensional regression is required. Let \mathbf{B}_a denote the B-spline basis regression matrix for ages. Similarly, let \mathbf{B}_y be the B-spline basis regression matrix for calendar years. The new regression matrix, \mathbf{B} to be used for two-dimensional Poisson P-spline smoothing is the following :

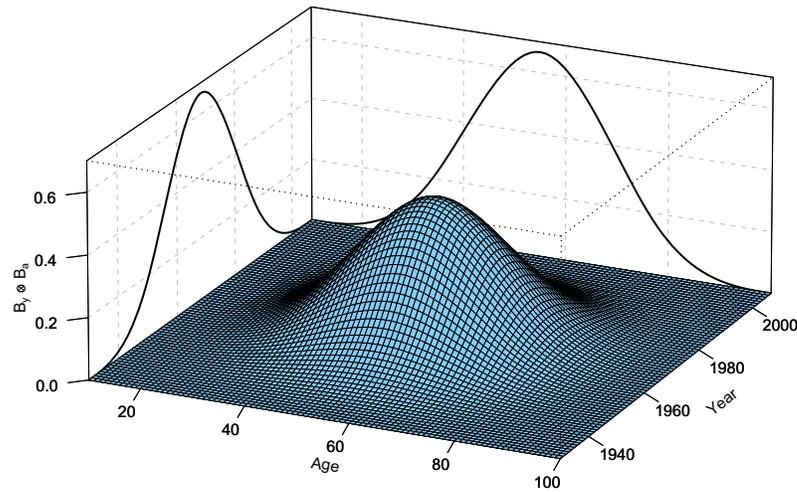
$$\mathbf{B} = \mathbf{B}_y \otimes \mathbf{B}_a, \quad (4.3)$$

where the symbol \otimes represents the Kronecker product.

As shown in Figure 4.6, the Kronecker product of two B-splines (one along the year dimension and one along the age dimension) gives rise to a hill. Thus,

the Kronecker product of the two B-spline basis, \mathbf{B}_y and \mathbf{B}_a in equation (4.3) will populate the age-year grid with several overlapping and regularly spaced hills such as the one of Figure 4.6. Indeed, the complete illustration of \mathbf{B} (not shown here) includes about 300 overlapping hills, which provide great flexibility in the modelling process.

FIGURE 4.6. Two-dimensional Kronecker product of two cubic B-splines



Let matrices \mathbf{D} and \mathbf{E} denote respectively deaths and exposure data for the population under study, where rows refer to ages and columns to calendar years. In other words, matrices \mathbf{D} and \mathbf{E} both include as many rows as there are ages considered and as many columns as there are calendar years considered. For the purpose of regression, these data are arranged into column vectors $\mathbf{d} = \text{vec}(\mathbf{D})$ and $\mathbf{e} = \text{vec}(\mathbf{E})$; the *vec* operator *vectorizes* a given matrix by stacking its columns. Also, let $\boldsymbol{\mu}$ be the force of mortality, similarly arranged into a column vector.

Following the same idea as in the one-dimensional case described earlier, we assume that the force of mortality is constant within each single age-year interval. Thus, the Poisson regression setting, together with P-splines yield

$$\begin{aligned}
\ln(\hat{E}[\mathbf{d}]) &= \ln(\mathbf{e}) + \ln(\hat{\boldsymbol{\mu}}) \\
&= \ln(\mathbf{e}) + \mathbf{B}\hat{\boldsymbol{\alpha}}.
\end{aligned}
\tag{4.4}$$

The vector of coefficients $\boldsymbol{\alpha}$ is estimated by maximizing the penalized log-likelihood function given by equation (4.2), where \mathbf{B} is defined as in equation (4.3) and the penalty matrix, \mathbf{P} ensures that neighboring estimated coefficients vary smoothly over the age and year dimensions (see Currie *et al.* (2004) and Camarda (2008) for further details on the penalty matrix). The trade-off between smoothness and accuracy of the estimated model is controlled by two smoothing parameters, one in each dimension. These smoothing parameters are selected independently according to the Bayesian Information Criterion (see Currie *et al.* (2004) and Camarda (2008) for further details on the selection of the smoothing parameters in the context of mortality data), thus allowing different amount of smoothing in each dimension.

More explicitly, from equations (4.3) and (4.4), the smoothed force of mortality corresponds to :

$$\begin{aligned}
\hat{\boldsymbol{\mu}} &= \exp(\mathbf{B}\hat{\boldsymbol{\alpha}}) \\
&= \exp((\mathbf{B}_y \otimes \mathbf{B}_a)\hat{\boldsymbol{\alpha}}).
\end{aligned}
\tag{4.5}$$

4.10. APPENDIX B : 99% BOOTSTRAP CONFIDENCE INTERVALS

In this appendix, we present a brief outline of the bootstrap approach used in this paper. We also provide an excerpt of the bootstrap confidence intervals results for the modal age at death estimates (see Table 4.1) and standard deviation above the mode estimates (see Table 4.2). Further details on bootstrap methods in general can be found in Efron et Tibshirani (1993). Moreover, Koissi *et al.* (2006) recently developed a residual bootstrap approach to build confidence intervals for

forecasted life expectancy as an extension to the original Lee-Carter method. We adapted their approach to the P-splines mortality modeling setting.

In the present study, the residual bootstrap method allows us to estimate the variability of the modal age at death estimates and the standard deviation above the mode estimates for which theoretical distributions are unknown. From the fitted force of mortality $\hat{\boldsymbol{\mu}}$ given in equation (4.5), one can obtain the matrix of fitted death counts $\hat{\mathbf{D}}$. Indeed, from equations (4.4) and (4.5), we have $\hat{\mathbf{d}} = \mathbf{e} \cdot \hat{\boldsymbol{\mu}}$, where $\hat{\mathbf{d}} = \text{vec}(\hat{\mathbf{D}})$. Deviance residuals (McCullagh et Nelder, 1989, Sect. 2.4) can thus be computed since

$$\mathbf{r} = \text{sign}(\mathbf{D} - \hat{\mathbf{D}}) \cdot \sqrt{2 \left[\mathbf{D} \ln \left(\frac{\mathbf{D}}{\hat{\mathbf{D}}} \right) - \mathbf{D} + \hat{\mathbf{D}} \right]}. \quad (4.6)$$

From these deviance residuals, we sample with replacement an entire new set of residuals \mathbf{r}_b^* called the bootstrapped residuals. Replacing deviance residuals \mathbf{r} by bootstrapped residuals \mathbf{r}_b^* in equation (4.6) and rearranging this equation leads to

$$\hat{\mathbf{D}} - \mathbf{D} \ln(\hat{\mathbf{D}}) + \mathbf{r}_b^{*2} + \mathbf{D} - \mathbf{D} \ln(\mathbf{D}) = 0. \quad (4.7)$$

Equation (4.7) can be solved numerically in order to obtain a new matrix of bootstrapped deaths $\hat{\mathbf{D}}_b^*$. Together with the matrix of exposure data \mathbf{E} , these bootstrapped deaths are then used in a two-dimensional P-spline framework (see Appendix A, Section 4.9) and lead to new bootstrapped coefficients $\hat{\boldsymbol{\alpha}}_b^*$. Corresponding smoothed bootstrapped density function describing the two-dimensional age-at-death distributions and bootstrapped modal age at death estimates are then computed.

The procedure described above, starting with the residual sampling with replacement step, was repeated 5,000 times. This led to a bootstrapped distribution of $\hat{M}(y)$ and $\widehat{SD}(M+)(y)$, from which 0.995th and 0.005th empirical percentiles were extracted and used as lower and upper confidence bounds for the 99% bootstrap confidence intervals. As an example, Figure 4.7 shows the female modal age

at death estimates along with their 99% confidence intervals for British Columbia and Quebec.

FIGURE 4.7. Modal age at death estimates and corresponding 99% bootstrap confidence intervals, British Columbia and Quebec, 1930 to 2007

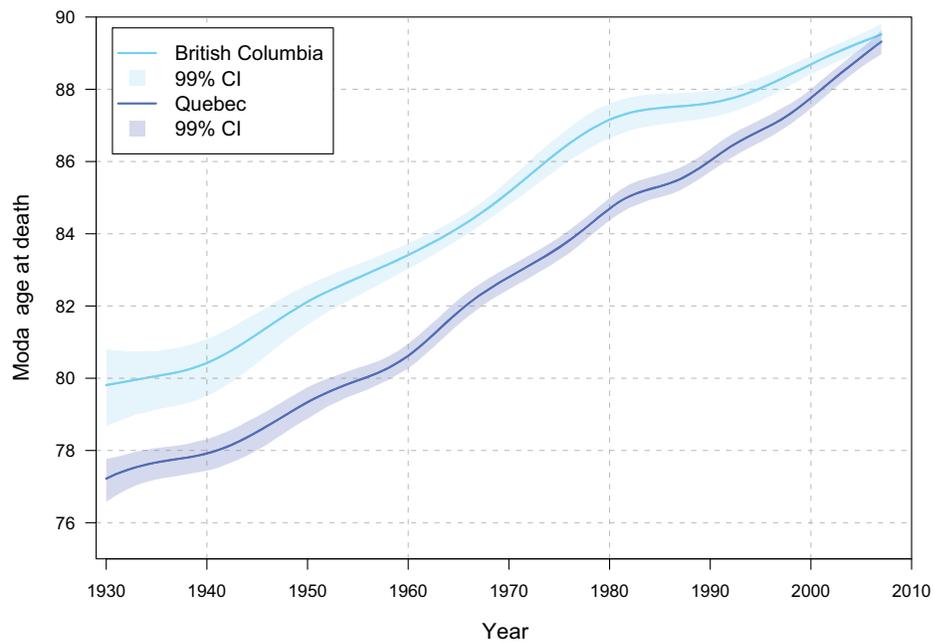


TABLEAU 4.1. Modal age at death estimates and corresponding 99% bootstrap confidence intervals by province and sex, Canada, 1930, 1950, 1970, 1990, and 2007 (Females); 1970, 1980, 1990, 2000, and 2007 (Males)

Sex	Year	BCO	ALB	MAN	SAS	ONT	QUE	NBR	NSC	PEI	NFL
Females	1930	79.8 (78.7,80.8)	78.7 (76.9,80.3)	79.2 (78.5,79.9)	79.5 (78.9,80.1)	78.5 (78.0,79.0)	77.2 (76.6,77.8)	78.6 (77.9,79.1)	79.3 (78.3,80.1)	82.1 (80.8,82.5)	
	1950	82.1 (81.5,82.6)	81.4 (80.5,82.1)	81.2 (80.7,81.5)	81.0 (80.6,81.3)	81.2 (80.4,81.3)	79.3 (78.9,79.8)	80.2 (79.7,80.5)	81.4 (80.9,81.8)	82.6 (81.7,82.8)	80.9 (80.2,81.6)
	1970	85.2 (84.8,85.5)	84.8 (84.4,85.2)	84.4 (84.0,84.8)	85.1 (84.8,85.4)	84.0 (83.8,84.2)	82.8 (82.5,83.1)	83.7 (83.3,84.0)	83.9 (83.5,84.2)	85.1 (84.3,85.3)	83.4 (82.9,83.8)
	1990	87.6 (87.2,88.0)	88.1 (87.7,88.5)	87.2 (86.8,87.6)	88.1 (87.8,88.4)	86.9 (86.6,87.1)	86.0 (85.7,86.3)	86.8 (86.4,87.1)	86.5 (86.0,86.9)	87.2 (86.5,87.4)	85.1 (84.7,85.6)
	2007	89.5 (89.2,89.8)	89.7 (89.3,90.2)	89.5 (89.1,89.8)	89.7 (89.4,90.0)	89.4 (89.2,89.6)	89.3 (89.0,89.6)	88.5 (88.1,88.9)	88.3 (87.9,88.7)	88.5 (87.7,88.9)	86.9 (86.3,87.5)
Males	1970	78.5 (76.7,80.0)	78.6 (77.4,79.6)	78.4 (77.2,79.2)	80.5 (79.6,81.5)	76.3 (75.7,77.1)	77.2 (76.0,78.0)	78.9 (77.4,79.7)	77.2 (75.9,78.4)	79.2 (77.7,79.6)	79.6 (77.9,80.6)
	1980	80.7 (80.0,81.5)	79.8 (78.8,80.8)	79.4 (78.4,80.3)	81.3 (80.3,82.1)	78.3 (77.7,78.8)	77.7 (76.7,78.3)	79.1 (78.3,79.7)	77.4 (76.3,78.3)	79.5 (78.8,80.4)	79.0 (77.7,79.9)
	1990	82.2 (81.6,82.8)	82.1 (81.2,82.8)	81.4 (80.5,82.2)	82.1 (81.0,82.8)	80.9 (80.5,81.3)	79.9 (79.3,80.6)	80.0 (79.4,80.6)	79.4 (78.5,80.2)	80.5 (79.3,80.9)	79.6 (78.8,80.3)
	2000	84.6 (84.2,85.0)	84.1 (83.6,84.6)	83.1 (82.3,83.6)	83.4 (82.4,84.1)	83.5 (83.2,83.8)	82.3 (81.7,82.8)	82.2 (81.1,83.1)	82.0 (81.1,82.8)	82.1 (81.4,82.8)	81.0 (80.2,81.7)
	2007	86.9 (86.1,87.6)	85.9 (85.0,86.9)	84.3 (83.4,85.0)	85.5 (84.5,86.4)	85.7 (85.3,86.3)	85.7 (84.8,86.6)	84.6 (83.5,85.6)	83.9 (82.6,84.9)	83.3 (82.1,83.9)	82.3 (81.2,83.1)

Note : 99% bootstrap confidence intervals results are based on 5,000 bootstrap samples.

TABLEAU 4.2. Standard deviation above the modal age at death estimates and corresponding 99% bootstrap confidence intervals by province and sex, Canada, 1930, 1950, 1970, 1990, and 2007 (Females); 1970, 1980, 1990, 2000, and 2007 (Males)

Sex	Year	BCO	ALB	MAN	SAS	ONT	QUE	NBR	NSC	PEI	NFL
Females	1930	8.44 (7.49,8.83)	8.04 (6.98,8.75)	8.30 (7.60,8.53)	7.99 (7.35,8.16)	7.58 (7.23,7.77)	8.07 (7.69,8.39)	7.81 (7.38,8.06)	8.26 (7.77,8.68)	7.10 (6.56,7.46)	
	1950	7.31 (6.94,7.50)	7.46 (6.96,7.73)	7.49 (7.13,7.63)	7.82 (7.46,7.92)	7.23 (7.14,7.56)	7.72 (7.47,7.92)	7.87 (7.55,8.02)	7.68 (7.37,7.86)	7.21 (6.83,7.43)	7.59 (7.04,7.91)
	1970	7.14 (6.86,7.25)	7.36 (7.02,7.46)	7.29 (6.97,7.44)	7.23 (6.92,7.32)	7.31 (7.20,7.42)	7.09 (6.90,7.24)	7.45 (7.14,7.58)	7.49 (7.23,7.63)	7.32 (6.92,7.54)	7.24 (6.86,7.40)
	1990	6.96 (6.74,7.12)	6.72 (6.45,6.88)	7.19 (6.91,7.37)	6.70 (6.45,6.81)	7.07 (6.92,7.16)	7.44 (7.23,7.57)	7.30 (7.00,7.43)	7.57 (7.28,7.74)	7.18 (6.82,7.42)	7.57 (7.20,7.76)
	2007	6.51 (6.26,6.70)	6.41 (6.11,6.63)	6.35 (6.05,6.56)	6.08 (5.80,6.25)	6.42 (6.25,6.51)	6.71 (6.47,6.92)	6.86 (6.52,7.09)	6.91 (6.58,7.15)	6.55 (6.11,6.93)	6.90 (6.48,7.23)
Males	1970	8.78 (8.02,9.59)	8.81 (8.29,9.36)	8.65 (8.20,9.16)	7.97 (7.47,8.35)	9.15 (8.79,9.46)	8.44 (8.01,8.99)	8.32 (7.83,8.95)	8.99 (8.38,9.56)	8.55 (8.07,8.94)	7.98 (7.37,8.62)
	1980	8.33 (7.92,8.60)	8.58 (8.05,9.02)	8.54 (8.05,8.99)	8.01 (7.55,8.43)	8.84 (8.57,9.07)	8.90 (8.54,9.35)	8.51 (8.14,8.82)	9.10 (8.60,9.57)	8.63 (8.15,8.98)	8.38 (7.83,8.88)
	1990	7.65 (7.35,7.94)	7.67 (7.30,8.05)	7.90 (7.48,8.25)	7.87 (7.45,8.29)	7.93 (7.70,8.08)	8.14 (7.81,8.40)	8.32 (7.91,8.52)	8.40 (7.95,8.80)	8.36 (7.91,8.68)	8.13 (7.68,8.42)
	2000	6.94 (6.70,7.13)	6.99 (6.73,7.20)	7.53 (7.20,7.82)	7.42 (7.03,7.89)	7.13 (7.02,7.29)	7.45 (7.21,7.71)	7.62 (7.12,8.09)	7.67 (7.21,8.06)	7.81 (7.33,8.18)	7.52 (7.09,7.82)
	2007	6.56 (6.16,6.94)	6.56 (6.03,7.00)	7.36 (6.90,7.74)	6.76 (6.21,7.19)	6.90 (6.55,7.07)	6.82 (6.34,7.22)	6.93 (6.32,7.44)	7.27 (6.70,7.87)	7.40 (6.85,7.86)	6.97 (6.46,7.43)

Note : 99% bootstrap confidence intervals results are based on 5,000 bootstrap samples.

Chapitre 5

La durée de vie la plus commune des adultes au XVIII^e siècle : l'expérience des Canadiens-français

Nadine Ouellette*, Jean-Marie Robine†, Robert Bourbeau* et
Bertrand Desjardins*

*Département de démographie, Université de Montréal

†INSERM, Démographie et Santé, Montpellier, France

RÉSUMÉ

Au cours des dix dernières années, l'âge modal au décès a pris une place prépondérante au sein des études sur la longévité humaine, notamment parce qu'il nous informe sur la durée de vie la plus commune des adultes. Nos connaissances du niveau et de l'évolution de l'âge modal au décès dans des populations du passé étant limitées, nous profitons du Registre de la population du Québec ancien, une source de données historiques reconnue pour sa fiabilité, pour apporter un nouvel éclairage sur la longévité des adultes au XVIII^e siècle. Les données utilisées, détaillées par sexe et année d'âge, couvrent la période 1740-1754 à 1785-1799. Notre analyse révèle que l'âge le plus commun au décès parmi les adultes canadiens-français a augmenté au cours de cette période. En effet, l'âge modal au décès est passé d'environ 73 ans à près de 76 ans chez les femmes et d'environ 70 ans à 74 ans chez les hommes. Les conditions de vie particulières de la population canadienne-française à cette époque pourraient expliquer cet accroissement.

Mots clés : Âge modal au décès, Registre de la population du Québec ancien, répartition des décès selon l'âge, lissage par P-splines, longévité humaine

Article à soumettre à la revue *Population*

5.1. INTRODUCTION

D'un point de vue démographique, le terme longévité peut, soit faire référence à cette capacité individuelle de survivre jusqu'à des âges très avancés, soit se rapporter à la survie de l'ensemble de la population. Lorsque conceptualisée au niveau de la population, la longévité est généralement mesurée par l'espérance de vie à la naissance, c'est-à-dire la durée de vie moyenne. La durée de vie modale, une mesure complémentaire de tendance centrale de la répartition des décès selon l'âge, s'avère toutefois particulièrement adaptée à l'étude de la longévité des adultes. En effet, contrairement à l'espérance de vie à la naissance, l'âge modal au décès est essentiellement influencé par les conditions de mortalité à l'âge adulte ; il est ainsi nettement plus sensible aux changements qui surviennent au sein de cette sous-population (Kannisto, 2001; Horiuchi, 2003).

Dès la fin du XIX^e siècle, le *Concept des durées de vie normales* émergent des travaux précurseurs de Lexis (1877, 1878) identifiait l'âge modal au décès comme étant la caractéristique la plus *centrale et normale* de la longévité humaine. Cependant, ce n'est que plus d'un siècle plus tard que Kannisto (2001) revisita les travaux de Lexis et qu'il popularisa l'utilisation de l'âge modal au décès au sein des études sur la longévité humaine. Dès lors, la durée de vie modale retient couramment l'attention des chercheurs (Canudas-Romo, 2008, 2010; Cheung *et al.*, 2005, 2008, 2009; Cheung et Robine, 2007; Kannisto, 2000, 2007; Ouellette et Bourbeau, 2010; Ouellette *et al.*, 2010; Robine, 2001; Thatcher *et al.*, 2010). Grâce à ces études, nous avons maintenant une bonne connaissance de l'évolution, depuis le milieu du XIX^e siècle, de cette durée de vie la plus commune chez les adultes dans une variété de pays à faible mortalité.

Nos connaissances relatives à l'âge modal au décès chez les adultes au sein des populations du passé demeurent quant à elles limitées, notamment en raison de la rareté de données historiques détaillées et fiables sur la mortalité avant le milieu du XIX^e siècle. Ainsi, nous ignorons si l'âge le plus commun au décès chez les adultes a augmenté, diminué, fluctué ou s'il a plutôt eu tendance à demeurer stable durant une très longue période de temps dans le passé. Sous un angle

strictement analytique, il a toutefois été démontré que si la mortalité diminue exclusivement aux âges inférieurs à l'âge modal, alors cet âge modal demeure inchangé (Canudas-Romo, 2010). Autrement dit, des gains en matière de mortalité aux âges supérieurs à l'âge modal, soit chez les personnes âgées, sont requis pour le faire augmenter. Sur cette base, nous pourrions être tentés de croire que la durée de vie la plus commune des adultes n'aurait guère changé avant le XX^e siècle.

Le Registre de la population du Québec ancien (2009), une source de données historiques unique et reconnue pour sa fiabilité, s'avère idéal pour mener une exploration empirique sur le sujet. Les données de ce Registre nous permettent effectivement de maintenir la population du Québec ancien sous observation durant une longue période de temps dans le passé. Nous proposons ainsi d'utiliser ces données dans le but de suivre l'évolution de l'âge modal au décès chez les Canadiens-français au cours du XVIII^e siècle.

Le présent article est donc organisé ainsi : à la manière de Robine et Cheung (2009), nous retraçons d'abord les changements survenus dans la répartition des décès adultes selon l'âge, en mettant l'emphase sur l'âge modal au décès, à partir des tables de mortalité déjà existantes reposant sur des données historiques des XVII^e, XVIII^e et XIX^e siècles. Cela n'a encore jamais été fait de manière aussi détaillée et ce sont les résultats de ce travail de recherche qui sont présentés dans la prochaine section. Ensuite, nous apportons un nouvel éclairage sur la longévité des adultes au XVIII^e siècle en exploitant le Registre de la population du Québec ancien. Les résultats obtenus sont situés dans le contexte particulier de la population canadienne-française à cette époque.

5.2. RÉPARTITION DES DÉCÈS SELON L'ÂGE D'APRÈS DIVERSES TABLES DE MORTALITÉ DES POPULATIONS DU PASSÉ

5.2.1. Table de mortalité de Halley : absence d'une durée de vie modale

Halley (1693) fut le premier à construire une table de mortalité à partir de données empiriques. Trente années auparavant, Graunt (1662) avait eu la brillante

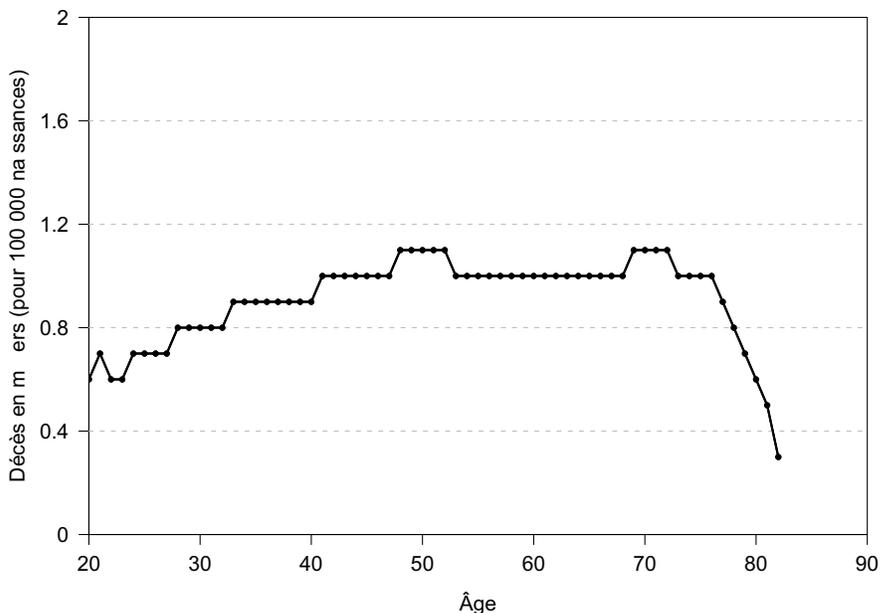
idée de réunir les variables suivantes dans un même tableau : l'âge, le nombre de décès et le nombre de survivants. Cependant, les bulletins de mortalité de la ville de Londres dont Graunt disposait ne fournissaient pas l'âge des décédés. Ainsi, pour construire sa table de mortalité, Graunt avança plusieurs hypothèses discutables et s'appuya sur une « fonction arithmétique arbitrairement définie » (Dupâquier, 1996, p.74) pour décrire l'évolution de la mortalité selon l'âge.

Les données tirées des bulletins de mortalité de la ville de Breslau, aujourd'hui connue sous le nom de Wrocław, et utilisées par Halley pour construire sa table provenaient de Neumann, un diacre de Breslau. Ce dernier avait compilé les décès par âge, sexe et mois survenus à Breslau au cours des années 1687 à 1691. Ainsi, c'est en se basant sur l'expérience de Breslau que Halley a construit sa table de mortalité complète. Il prétendait toutefois que celle-ci s'adressait plus généralement au genre humain, en raison de corrections personnelles apportées aux données de Neumann. La figure 5.1 présente un extrait de la répartition des décès selon l'âge de cette table de mortalité, c'est-à-dire à partir de 20 ans. D'après cette illustration, nous constatons qu'il survient à peu près le même nombre de décès à chaque âge entre 40 et 75 ans. De légers surplus apparaissent autour de 50 et 70 ans, mais il s'avère tout de même impossible de discerner un âge modal au décès qui correspondrait à la durée de vie la plus commune des adultes de cette époque (Robine et Cheung, 2009).

5.2.2. Tables de mortalité de Wargentín et Deparcieux : durée de vie modale autour de 70 ans

Les travaux de Halley ont inspiré plusieurs chercheurs dont Kersseboom (1742) (voir aussi la traduction en français de l'Institut national d'études démographiques (1970) d'une partie de ses oeuvres) et Struyck (1740, 1753), qui ont construit à leur tour des tables de mortalité. Néanmoins, toutes ces tables portaient sur des populations fermées et sélectionnées, précisément celle des rentiers. Il en va de même pour Deparcieux (1746) qui, dans son célèbre *Essai sur les probabilités de la durée de la vie humaine*, présentait aussi plusieurs tables sur la mortalité des rentiers et sur celle de quelques ordres religieux.

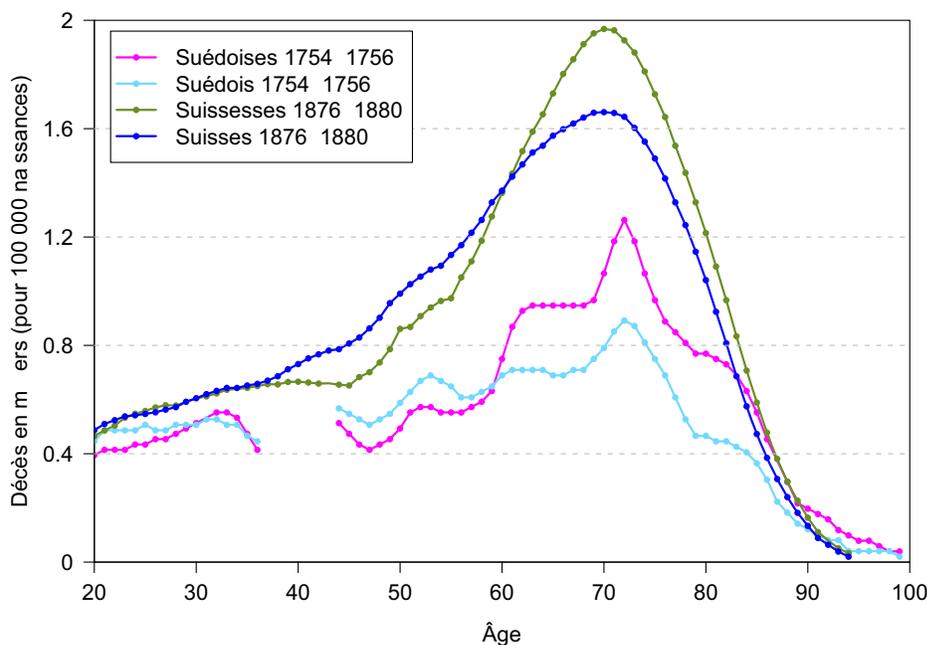
FIGURE 5.1. Répartition des décès selon l'âge chez les adultes d'après la table de mortalité complète de Halley, sexes réunis, Breslau, 1687-1691



Source : Décès calculés par les auteurs à l'aide des survivants de la table de mortalité de Halley (1693)

Il faut attendre jusqu'en 1760 pour que Deparcieux publie des tables de mortalité nationales calculées à partir des décès des années 1754, 1755 et 1756 en Suède dans son *Addition à l'Essai* (Deparcieux, 1760). Soulignons que ces tables établies selon le sexe lui avaient été communiquées par Wargentin, alors membre d'un Bureau général de Statistique en Suède. Les tables de Wargentin étant abrégées, Deparcieux a procédé à un lissage des séries de quotients quinquennaux de mortalité et en a déduit des tables de mortalité complètes, qu'il inclut aussi dans son *Addition à l'Essai*. La figure 5.2 présente un extrait de la répartition des décès par âge et par sexe de ces tables de mortalité suédoises complètes, soit à partir de 20 ans. Cette fois, les courbes indiquent que la durée de vie la plus commune chez les adultes suédois se situait probablement autour de 70 ans vers la moitié du XVIII^e siècle. En effet, l'âge auquel survient le plus grand nombre de décès se situe précisément à 72 ans, autant chez les femmes que chez les hommes.

FIGURE 5.2. Répartition des décès selon l'âge et le sexe chez les adultes en Suède, 1754-1756, et en Suisse, 1876-1880



Sources : Deparcieux (1760) et Office fédéral de la statistique suisse (2010)

Note : Les données pour la Suède à 37 ans dans la contribution de Deparcieux (1760) sont illisibles et celles entre 38 et 43 ans sont manquantes.

5.2.3. Premières tables de mortalité complètes en Suisse : la durée de vie modale n'aurait pas changé

Une attente de plus de cent ans est nécessaire avant de pouvoir confronter ces répartitions des décès selon l'âge et le sexe pour la Suède en 1754-1756 à celles d'une autre période. En effet, les prochaines tables de mortalité complètes nationales disponibles correspondent aux premières tables construites pour la Suisse. Elles concernent la période 1876-1880 et ont été établies à partir de statistiques sur les décès et la population par âge et par sexe (Office fédéral de la statistique suisse, 2010). Les répartitions des décès selon l'âge et le sexe tirés de ces tables de mortalité suisses apparaissent aussi à la figure 5.2. Nous constatons que la durée de vie la plus commune chez les adultes des deux sexes se situe à peu près au même âge que celui observé 125 ans plus tôt en Suède, c'est-à-dire autour de 70 ans.

Ces observations rejoignent celles de Robine et Cheung (2009) et nous pourrions être tentés d'en déduire que la durée de vie modale des adultes des deux sexes n'aurait pratiquement pas changé durant plus de cent ans. Cependant, une telle réflexion s'avérerait hasardeuse puisque les paires de courbes présentées à la figure 5.2 concernent deux pays distincts ; leur mortalité respective aux âges adultes aux XVIII^e et XIX^e siècles n'était donc pas forcément équivalente. À la lumière des informations provenant de la Human Mortality Database (2010), une base de données internationale fournissant des données détaillées sur la mortalité en Suède et en Suisse notamment, la mortalité chez les adultes vers la fin du XIX^e siècle était effectivement nettement plus faible en Suède qu'en Suisse. La stagnation apparente du mode pourrait donc simplement résulter du fait que la mortalité aux âges adultes en Suisse vers la fin du XIX^e atteignait tout juste un niveau observé quelque cent vingt-cinq ans plus tôt en Suède.

Pour poursuivre notre analyse de la longévité des adultes à cette époque, le Registre de la population du Québec ancien constitue ainsi une source de données idéale. En effet, cette source nous permet de suivre l'évolution d'une seule et même population sur une longue période de temps. Elle nous offre également la possibilité d'avoir le plein contrôle sur la méthodologie utilisée pour estimer la durée de vie la plus commune des adultes.

5.3. RÉPARTITION DES DÉCÈS SELON L'ÂGE DES CANADIENS-FRANÇAIS AU XVIII^E SIÈCLE

5.3.1. Registre de la population du Québec ancien

Le Registre de la population du Québec ancien (RPQA) est une base de données qui repose sur l'exploitation systématique des actes de baptême, mariage et sépulture des XVII^e et XVIII^e siècles, ainsi que sur les sépultures de la période 1800-1850 qui concernent les personnes nées avant 1750. Précisons que c'est au Programme de recherche en démographie historique (PRDH) de l'Université de Montréal que revient le mérite de la consolidation informatique de ces informations sous forme de base de données. Le RPQA concerne tous les individus

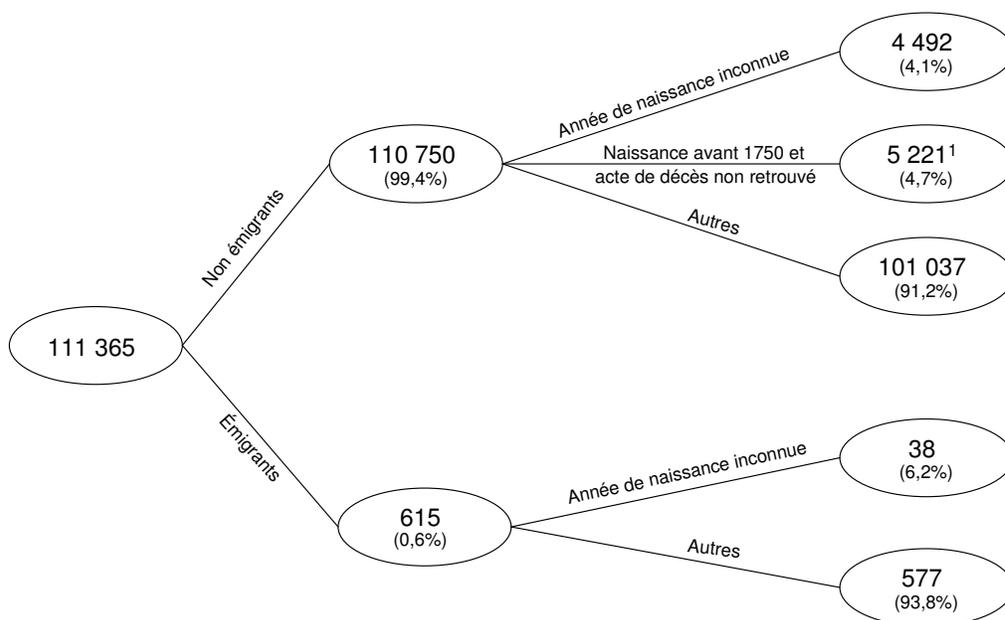
venus s'installer dans la vallée du Saint-Laurent à l'heure de la colonisation française, ainsi que leurs descendants. Cette source de données historiques a donc le grand avantage de couvrir l'ensemble du territoire occupé par la population des Canadiens-français, éliminant ainsi les problèmes de sélection et de biais qui pourraient en résulter autrement (Desjardins, 1998). De plus, les données du RPQA reposent sur des fondations très solides ; la qualité des registres a effectivement été maintes fois validée. Ces caractéristiques font de cette base de données une source historique exceptionnelle. À l'heure actuelle, le portrait transversal de la population du Québec peut être tracé depuis l'arrivée des premiers colons en 1608 jusqu'en 1799 grâce à ces données. Ajoutons que l'objectif final du PRDH vise une couverture analogue pour le XIX^e siècle.

En raison des problèmes d'observation que posent les célibataires, seuls les mariés, qui représentent la très grande majorité des adultes à cette époque, ont été retenus pour notre étude. Au sein des 126 825 individus mariés au Québec ancien, nous retirons ensuite les immigrants, les Amérindiens et les Noirs pour des raisons d'homogénéité de la population et parce que nés hors du territoire couvert par le RPQA, leur date de naissance étant généralement inconnue. Dans l'ensemble, ces retraits se chiffrent à 15 460 (12,2%) individus. Les 111 365 individus restants se partagent en 110 750 (99,4%) non émigrants et 615 (0,6%) émigrants (voir figure 5.3). Nous conservons les émigrants puisque ces derniers sont soumis au risque de décéder tant et aussi longtemps qu'ils demeurent sur le territoire du Québec ancien.

Tel que décrit à la figure 5.3, d'autres retraits s'imposent. Tout d'abord, autant chez les non émigrants que chez les émigrants, l'année de naissance est inconnue pour une minorité d'individus en raison de pertes aléatoires de registres et ces derniers ne peuvent être conservés. Ensuite, parmi les non émigrants nés avant 1750, ceux dont l'année de décès est manquante devraient aussi être retirés car tous les actes de sépulture datés entre 1800 et 1850 concernant des individus nés avant 1750 ont été relevés par le PRDH. Le fait que le destin précis de ces individus soit inconnu s'explique essentiellement par la perte ou l'absence aléatoire

d'enregistrement de leur décès (Charbonneau *et al.*, 1996), évaluée ici à 9,2%, ce qui est tout à fait exceptionnel dans un tel contexte historique. Parmi les 5 221 non émigrants nés avant 1750 touchés par cette situation, nous avons pu estimer l'année de décès de 2 720 d'entre eux grâce aux informations tirées des actes des autres membres de leur famille. Au final, notre base de données compte donc 104 334 individus, c'est-à-dire 103 757 non émigrants (101 037 + 2 720) et 577 émigrants.

FIGURE 5.3. Schéma descriptif des données sur les mariés tirées du Registre de la population du Québec ancien



Source : Registre de la population du Québec ancien (2009)

Note : ¹ Rapportés aux 56 491 non émigrants nés avant 1750, ces 5 221 cas représentent 9,2%. L'année de décès de 2 720 d'entre eux a tout de même pu être estimée.

La figure 5.4 illustre la forte croissance de la population des mariés au Québec ancien entre 1640 et 1799, reflétant la croissance de la population totale. En effet, au début du XVII^e siècle, la vallée laurentienne comptait près de 20 000 habitants au total alors qu'au tournant du XIX^e siècle, la colonie comptait environ 200 000 Canadiens-français. La figure 5.5 présente quant à elle l'évolution du nombre annuel de décès survenus au sein de cette population pour chaque sexe.

La nette augmentation des décès observée tout au long de cette période s'explique essentiellement par le peuplement progressif du territoire. Les variations annuelles, typiques de la mortalité d'ancien régime, sont le reflet des décès liés, notamment à des conditions du moment, épidémies, hivers rigoureux, mauvaises récoltes et guerres.

FIGURE 5.4. Effectifs de la population mariée retenue selon le sexe, Québec ancien, 1^{er} juillet, 1640 à 1799



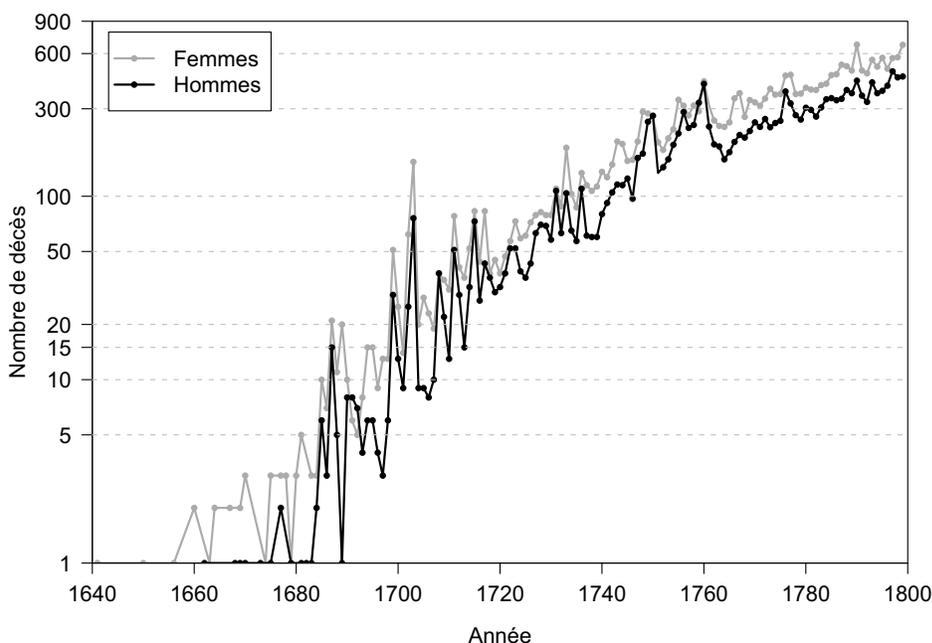
Source : Registre de la population du Québec ancien (2009)

5.3.2. Tables de mortalité complètes pour les mariés

À partir des décès et effectifs de la population par année d'âge, sexe et année de calendrier, nous pouvons construire des tables de mortalité complètes pour les mariés. Afin d'atténuer l'effet des variations annuelles typiques de la mortalité de l'époque liées à l'influence des conditions du moment, l'établissement de tables pour des périodes de 15 années s'est avéré le meilleur choix. Ces tables de mortalité qui commencent à l'âge exact de 20 ans nous offriront un aperçu des changements survenus dans la répartition des décès selon l'âge et le sexe chez les adultes mariés du Québec ancien au fil du temps. En raison des faibles effectifs observés jusqu'au

début du XVIII^e siècle (voir figures 5.4 et 5.5), la première table de mortalité porte sur la période 1740-1754. La dernière table établie repose sur les données de la période 1785-1799, de sorte que huit tables de mortalité successives couvrant plus de la moitié du XVIII^e siècle sont calculées, soit quatre pour chaque sexe.

FIGURE 5.5. Nombre annuel de décès de mariés selon le sexe, Québec ancien, 1640 à 1799



Source : Registre de la population du Québec ancien (2009)

Pour débiter, des taux de mortalité par âge, sexe et période sont obtenus en rapportant le nombre de décès survenus au cours de la période à la somme de l'exposition au risque estimée pour chaque année de la période. Ainsi, le taux de mortalité à l'âge x pour le sexe s et la période $[T, T + 14]$ vaut

$$t_x^s[T, T + 14] = \frac{\sum_{i=T}^{T+14} D_x^s[i]}{\sum_{i=T}^{T+14} E_x^s[i]},$$

où $D_x^s[i]$ représente les décès observés survenus à l'âge x pour le sexe s durant l'année i et $E_x^s[i]$ correspond à l'exposition au risque à l'âge x pour le sexe s et

l'année i . L'exposition au risque $E_x^s[i]$ est estimée par l'effectif de la population mariée d'âge x et de sexe s au 1^{er} juillet de l'année i .

Ces taux de mortalité par âge, sexe et période sont ensuite transformés en quotients de mortalité à l'aide de la méthode linéaire (aussi appelée la méthode actuarielle), c'est-à-dire sous l'hypothèse d'une répartition uniforme des décès à l'intérieur de chaque intervalle d'âge d'une année. Chez les adultes, une telle hypothèse est généralement satisfaite, sauf aux grands âges où d'autres méthodes plus sophistiquées sont habituellement utilisées (Thatcher *et al.*, 1998). Étant donné que ces tables de mortalité ne visent qu'à nous donner une première idée de la répartition des décès de mariés selon l'âge et le sexe, la méthode linéaire convient à notre démonstration.

Enfin, comme les survivants S_x et les décès d_x par âge des différentes tables de mortalité découlent directement de leurs quotients de mortalité q_x par âge respectifs, les tables de mortalité ne nécessitent pas d'être fermées.

5.3.3. Lissage des données de mortalité à l'aide de la méthode des P-splines

Afin d'obtenir des estimations précises de l'âge modal au décès selon le sexe et la période, nous utilisons la méthode de lissage par P-splines décrite par Ouellette et Bourbeau (2010), inspirée des travaux de Eilers et Marx (1996), Currie *et al.* (2004) et Camarda (2008). Étant donné le caractère typiquement erratique de la répartition des décès de la table de mortalité selon l'âge aux alentours du mode, il aurait été possible de recourir à d'autres méthodes. À titre d'exemple, nous aurions pu utiliser le modèle quadratique (Canudas-Romo, 2008, 2010; Cheung *et al.*, 2008; Kannisto, 2001, 2007; Pearson, 1902; Thatcher *et al.*, 2010) ou alors le modèle normal (Cheung *et al.*, 2009; Cheung et Robine, 2007) pour lisser les décès de la table de mortalité. Nous avons plutôt préféré lisser les décès observés, compte tenu de la population soumise au risque de décéder, à l'aide des P-splines pour davantage de flexibilité. Brièvement, la méthode par P-splines fournit une estimation non paramétrique de la fonction de densité, cette dernière décrivant la

répartition des décès selon l'âge. L'âge auquel le maximum de cette fonction de densité est atteint correspond ainsi à notre estimation de l'âge modal au décès.

Voici une description plus détaillée de la méthode de lissage par P-splines à appliquer séparément selon le sexe et la période. Par souci de clarté, les indices relatifs au sexe et à la période sont omis ici. Tout d'abord, sous l'hypothèse d'un taux instantané de mortalité μ_x (force de mortalité) constant à l'intérieur de chaque intervalle d'âge $[x, x + 1)$, les décès observés par âge D_x peuvent être vus comme les réalisations d'une loi de Poisson de moyenne $E_x \cdot \mu_x$, c'est-à-dire $D_x \sim \text{Poisson}(E_x \cdot \mu_x)$, où E_x réfère à l'exposition au risque selon l'âge.

Suivant cette dernière hypothèse, un modèle de régression de Poisson est utilisé pour prédire le nombre de décès survenus dans chaque intervalle d'âge. Les paramètres du modèle de régression de Poisson sont estimés à l'aide d'une approche non paramétrique qui combine les notions de B-splines (de Boor, 1978) et de fonction de vraisemblance pénalisée, d'où l'appellation P-splines. Ces deux notions se complètent bien puisque d'une part, les B-splines apportent beaucoup de souplesse dans le processus de modélisation et d'autre part, la vraisemblance pénalisée assure que la force de mortalité estimée sera lisse.

Ainsi, en dénotant les vecteurs de décès et d'exposition au risque par \mathbf{D} et \mathbf{E} respectivement, nous avons

$$\begin{aligned} \ln(\hat{\mathbb{E}}(\mathbf{D})) &= \ln(\mathbf{E} \cdot \hat{\boldsymbol{\mu}}) \\ &= \ln(\mathbf{E}) + \ln(\hat{\boldsymbol{\mu}}) \\ &= \ln(\mathbf{E}) + \mathbf{B}\hat{\mathbf{a}}, \end{aligned}$$

où \mathbf{B} est une matrice qui représente la base B-splines évaluée aux différents âges x et $\hat{\mathbf{a}}$ correspond aux paramètres estimés associés à chacun des B-splines inclus dans la base \mathbf{B} . Le vecteur des paramètres \mathbf{a} est estimé à l'aide de la méthode du maximum de vraisemblance et la fonction de log-vraisemblance pénalisée à maximiser est

$$v^* = v(\mathbf{a}; \mathbf{B}; \mathbf{D}) - \frac{1}{2} \mathbf{a}' \mathbf{P} \mathbf{a}.$$

Le premier terme de cette équation correspond à la log-vraisemblance habituelle d'un modèle linéaire généralisé. Le second est un terme de pénalité et il assure une certaine régularité au sein du vecteur des paramètres estimés $\hat{\mathbf{a}}$. En effet, ce terme de pénalité fait en sorte que les valeurs des paramètres estimés associés aux B-splines successifs ne changent pas trop abruptement. Il s'ensuit que la force de mortalité estimée $\hat{\mu}(x)$ résultante est lisse. Pour une description plus approfondie de cette fonction de log-vraisemblance pénalisée, voir Currie *et al.* (2004, p.282-284).

Suivant les relations d'équivalence entre la force de mortalité $\mu(x)$, la fonction de survie $S(x)$ et la fonction de densité $f(x)$, nous pouvons écrire

$$\hat{f}(x) = \hat{\mu}(x) \cdot \hat{S}(x) = \hat{\mu}(x) \cdot \exp\left(-\int_0^x \hat{\mu}(t)dt\right).$$

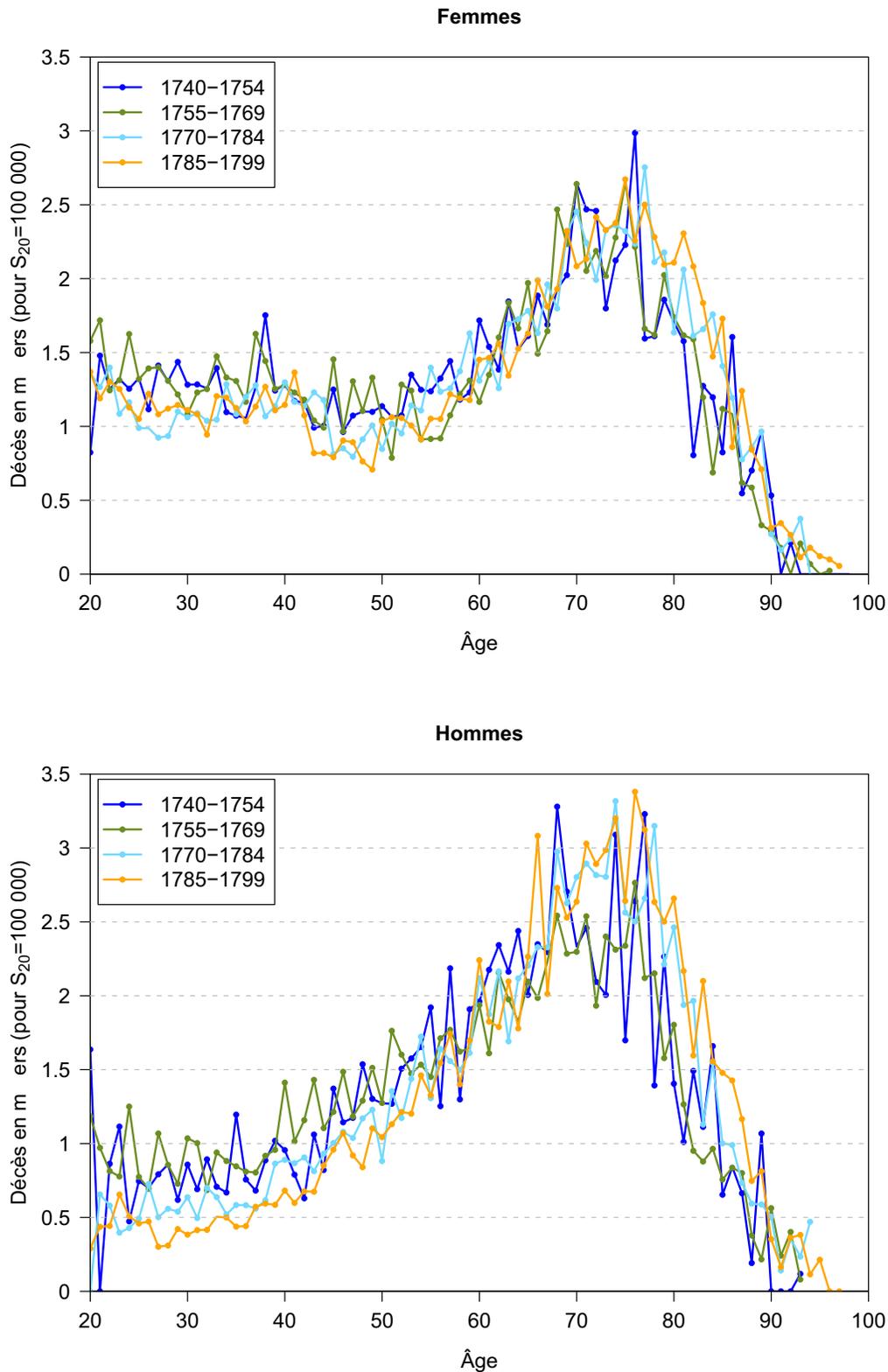
Ainsi, nous estimons la fonction de densité $f(x)$ qui décrit la répartition des décès selon l'âge à l'aide de méthodes d'intégration numérique usuelles. Ensuite, l'âge modal au décès estimé correspond à $\hat{M} = \max_x \hat{f}(x)$.

5.4. RÉSULTATS

La figure 5.6 illustre la répartition des décès selon l'âge des adultes mariés tirés des tables de mortalité complètes par sexe établies pour les périodes 1740-1754 à 1785-1799. L'allure erratique des quatre séries de décès apparaissant sur chacun de ces graphiques rend plutôt difficile la détermination d'une valeur unique à l'âge modal au décès dans ces cas. Cependant, à première vue, il semblerait que l'âge modal au décès au cours de toutes ces périodes se soit constamment situé entre 70 et 80 ans. Au fil des périodes, nous constatons également un déplacement de la répartition des décès des deux sexes selon l'âge vers la droite, c'est-à-dire vers des âges plus avancés. Il restera à confirmer si ces changements dans la répartition des décès selon l'âge pour chaque sexe se reflètent à travers la durée de vie la plus commune au cours du XVIII^e siècle au Québec ancien.

À titre d'exemple, la figure 5.7 présente la fonction de densité décrivant la répartition des décès selon l'âge estimée par la méthode de lissage par P-splines

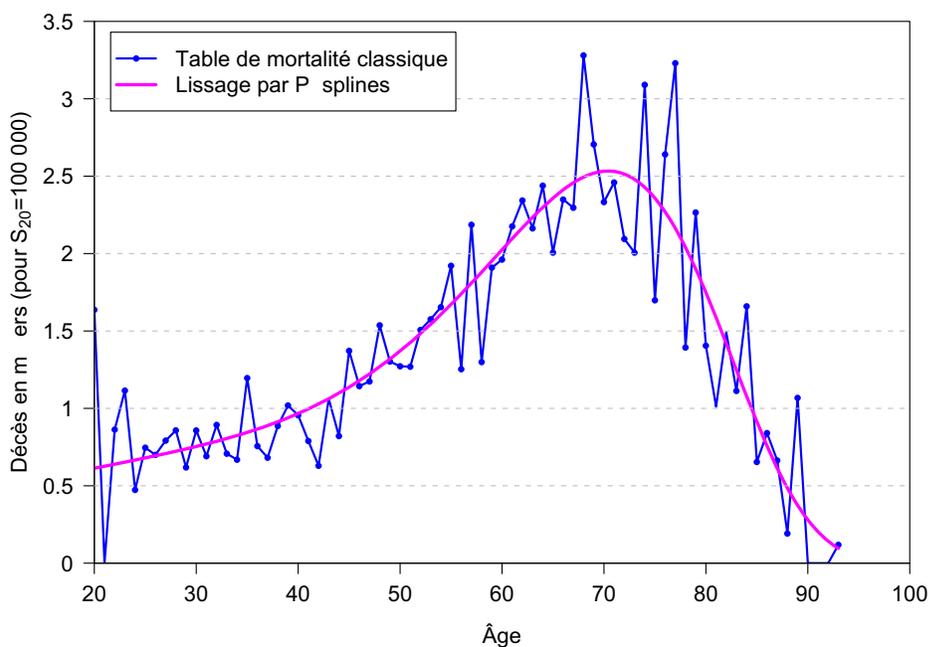
FIGURE 5.6. Répartition des décès des adultes mariés selon le sexe et l'âge d'après les tables de mortalité pour les périodes 1740-1754 à 1784-1799, Québec ancien



Source : Registre de la population du Québec ancien (2009)

chez les hommes mariés pour la période 1740-1754. Précisons que nous avons rapporté cette fonction de densité lissée à l'échelle de 100 000 survivants à l'âge 20 ans ; ceci dans le but de la comparer à la série des décès de la table de mortalité complète des hommes mariés pour la période 1740-1754 qui apparaît également sur la figure. Alors qu'il s'avère difficile de déterminer la durée de vie la plus commune à partir de la répartition des décès provenant de la table de mortalité complète, nous pouvons l'identifier précisément à l'aide de la fonction de densité. En effet, comme cette fonction atteint son maximum à 70,4 ans, nous estimons que la durée de vie la plus commune chez les hommes mariés du Québec ancien durant la période 1740-1754 se situait à cet âge.

FIGURE 5.7. Comparaison de la répartition des décès de la table de mortalité et de celle résultant d'un lissage par P-splines, Hommes mariés, Québec ancien, 1740-1754

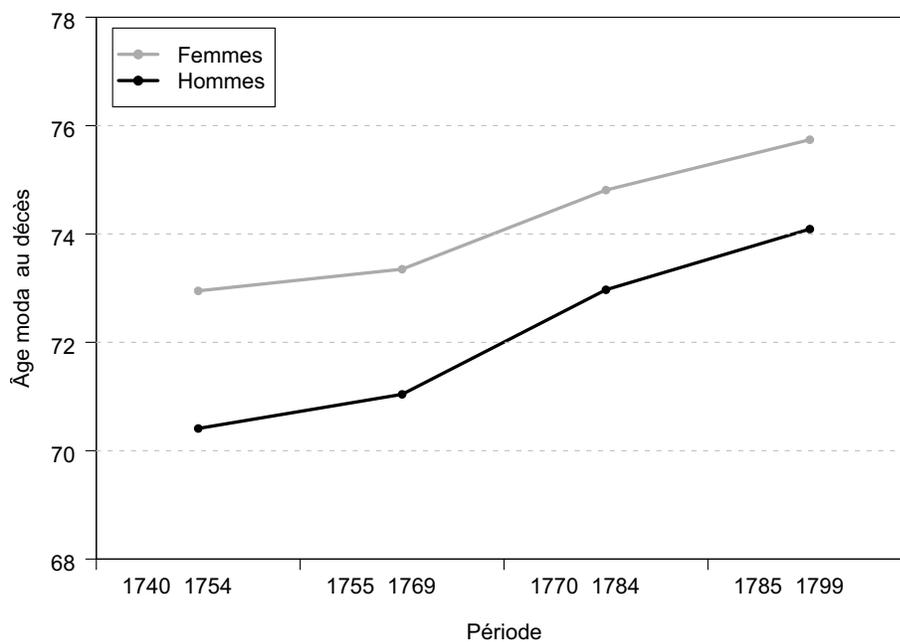


Source : Registre de la population du Québec ancien (2009)

La figure 5.8 illustre l'évolution de l'âge modal au décès estimé selon le sexe et la période au Québec ancien entre 1740-1754 et 1785-1799. Nous constatons d'emblée que la durée de vie la plus commune n'est pas demeurée inchangée au fil de ces périodes. En effet, l'âge modal au décès a constamment augmenté chez les

femmes : il est passé d'environ 73 ans en 1740-1754 à près de 76 ans en 1785-1799 (voir le tableau 5.1). Chez les hommes, la durée de vie la plus commune a aussi augmenté de manière soutenue entre 1740-1754 et 1785-1799. En 1740-1754, l'âge modal valait 70,4 ans et il dépassait 74 ans en 1785-1799. L'âge le plus commun au décès varie donc selon le sexe au Québec ancien et il est systématiquement plus élevé chez les femmes.

FIGURE 5.8. Durée de vie la plus commune chez les adultes mariés selon le sexe et la période, Québec ancien, 1740-1754 à 1785-1799



Source : Registre de la population du Québec ancien (2009)

TABLEAU 5.1. Estimations de la durée de vie la plus commune chez les adultes mariés, Québec ancien, 1740-1754 à 1785-1799

Périodes	\hat{M}	
	Femmes	Hommes
1740-1754	72,95	70,41
1755-1769	73,35	71,04
1770-1784	74,81	72,97
1785-1799	75,74	74,09

Source : Registre de la population du Québec ancien (2009)

5.5. DISCUSSION

Depuis une dizaine d'années, le concept de durée de vie modale retient de plus en plus l'attention dans les travaux axés sur le thème de la longévité humaine. Poursuivant cette tendance, la présente étude visait principalement à réduire l'incertitude relative au niveau et à la trajectoire temporelle de l'âge modal au décès au sein de populations du passé. Grâce au RPQA, une source exceptionnelle de données historiques reconnue pour sa couverture de la population du Québec ancien et sa fiabilité, nous avons pu suivre l'évolution de la durée de vie modale des Canadiens-français durant la seconde moitié du XVIII^e siècle. Les données du RPQA, généralement exploitées sous un devis longitudinal d'analyse démographique, se sont avérées idéales pour notre étude transversale de la durée de vie la plus commune chez les adultes.

Les résultats présentés à la section précédente révèlent que la durée de vie la plus commune des adultes Canadiens-français a continuellement augmenté entre 1740-1754 et 1785-1799. De plus, les femmes ont systématiquement enregistré un âge modal au décès plus élevé que les hommes durant cette période. Ces résultats sont surprenants à première vue pour deux raisons. Tout d'abord, il a été démontré analytiquement que l'âge modal au décès ne peut s'accroître en l'absence de gains au chapitre de la mortalité au-delà de celui-ci (Canudas-Romo, 2010). Sur cette base, nous étions donc tentés de croire que l'âge modal au décès n'allait guère changer dans les populations du passé. Ensuite, d'après les tables de mortalité de

Deparcieux (1760) pour la Suède en 1754-1756 et celles de l'Office de la statistique suisse (2010) pour la période 1876-1880, il n'y avait pas d'écart d'âge modal au décès entre les sexes.

Nos résultats doivent cependant être situés dans le contexte particulier de la population canadienne-française de l'époque. L'augmentation soutenue de l'âge modal au décès chez les adultes Canadiens-français au cours de la seconde moitié du XVIII^e siècle indique que les conditions de mortalité aux âges supérieurs à l'âge modal, c'est-à-dire chez les personnes âgées, se sont améliorées au Québec ancien durant cette période. Or, des travaux récents portant sur les Canadiens-français démontrent qu'une chute de la mortalité aux âges avancés est survenue dans les générations successives de la fin du XVII^e siècle et du premier quart du XVIII^e siècle (Gagnon et Mazan, 2009; Lacroix, 2009). Nos résultats rejoignent ainsi ceux de Gagnon et Mazan (2009) et Lacroix (2009) puisque les individus de ces générations constitueront essentiellement la population des personnes âgées durant seconde moitié du XVIII^e siècle.

Les années de conflit au Québec ancien ayant culminé avec la Conquête britannique en 1759 ont été particulièrement défavorables à la survie des individus, en raison des pertes au sein des miliciens et, surtout, des exactions commises par l'armée britannique. En outre, suite à la défaite française, les administrateurs et leur personnel de soutien, les marchands et les artisans, sont retournés en France. Comme ces personnes étaient essentiellement citadines, il en a résulté une hausse de la proportion de la population vivant en milieu rural chez les Canadiens-français. Comme la mortalité était plus faible à la campagne que dans les villes, ces facteurs pourraient ainsi expliquer l'accroissement de la durée de vie la plus commune des Canadiens-français entre 1740-1754 et 1785-1799. L'urbanisation massive ultérieure de la colonie, dès le début du XIX^e siècle, a pour sa part contribué à la détérioration des conditions de vie et potentiellement à la survie des individus. Cette dernière information suggère que le rythme d'augmentation de l'âge modal au décès prévalant chez les Canadiens-français durant la seconde moitié du XVIII^e siècle pourrait s'amenuiser après 1800.

Le travail de reconstitution de la population canadienne-française effectué par l'équipe du PRDH se poursuit et nous permettra éventuellement d'étendre notre analyse de l'évolution de la durée de vie modale jusqu'à la fin du XIX^e siècle. De pair avec les données provenant de l'État civil, qui débute en 1926 au Québec, un portrait en continuum de la longévité des adultes québécois pourra ainsi pratiquement être dressé avec ces nouvelles données historiques. D'ailleurs, une étude récente qui repose notamment sur les données de l'État civil québécois nous révèle que l'âge modal au décès au Québec au début des années 1930 se situait à 77,5 ans environ chez les femmes et à un peu plus de 76 ans chez les hommes (Ouellette *et al.*, 2010). Voilà qui suggère également que le rythme auquel l'âge modal au décès a augmenté durant la seconde moitié du XVIII^e siècle au Québec ancien ne sera probablement pas maintenu tout au long du XIX^e siècle. Pour l'instant, il est cependant impossible de savoir si l'âge modal continuera d'augmenter à un rythme moindre durant le XIX^e siècle, s'il stagnera, ou encore s'il diminuera.

Enfin, bien que le RPQA soit une source de données historiques exemplaire, des comparaisons avec d'autres sources permettant le suivi de populations du passé durant plusieurs années sont souhaitables. En raison des conditions de vie particulières de la population du Québec ancien durant la seconde moitié du XVIII^e siècle, le potentiel de généralisation de nos résultats relatifs à la durée de vie modale des Canadiens-français demeure effectivement difficile à évaluer. L'analyse des données de la Human Mortality Database (2010) constitue une avenue intéressante à ce sujet. Parmi les 37 pays inclus dans cette grande base de données internationale, dix d'entre eux comportent des données qui remontent jusqu'au XVIII^e siècle. La qualité de ces données historiques semble acceptable, quoique loin d'être équivalente à celle dont jouissent les données du RPQA.

5.6. REMERCIEMENTS

Ce projet a été financé par le Conseil de recherches en sciences humaines du Canada.

CONCLUSION

La baisse remarquable de la mortalité observée au cours du XX^e siècle dans toutes les régions du monde, en particulier dans les pays développés, a fait en sorte qu'une proportion de plus en plus importante de personnes atteint des âges avancés. Cette thèse portant sur le thème de la longévité des adultes s'est intéressée spécifiquement au suivi fin et détaillé des changements survenus dans la répartition des décès selon l'âge chez les personnes âgées. Pour ce faire, nous avons d'abord proposé une approche de lissage non paramétrique novatrice reposant sur les P-splines, qui s'avère parfaitement adaptée à un tel objectif. En effet, cette approche a le grand avantage de tirer entièrement profit de la régularité du profil de la mortalité selon l'âge et possiblement selon les années de calendrier également. Elle permet ainsi un relâchement des hypothèses relatives à la structure des données, autrement requises, et mène à une expression précise de la mortalité, telle que décrite par les données observées. Par conséquent, l'estimation des indicateurs de tendance centrale et de dispersion des décès résumant la distribution des décès selon l'âge, en particulier l'âge modal au décès, est réalisée dans un contexte de modélisation très flexible. Les P-splines offrent donc une excellente alternative aux méthodes statistiques paramétriques plus rigides habituellement utilisées, qui sont susceptibles d'exercer une influence sur l'estimation des indicateurs.

Grâce à cette nouvelle approche ainsi qu'aux données de la *Human Mortality Database*, nous avons pu analyser de près les changements survenus récemment en matière de répartition des décès aux grands âges au sein d'un sous-ensemble de quatre pays à faible mortalité : le Canada (1921 à 2007), les États-Unis (1945 à 2006), la France (1920 à 2007) et le Japon (1947 à 2008). À la lumière de nos

résultats, certains d'entre eux auraient depuis peu franchi l'ère de la compression de la mortalité aux grands âges, soit l'ère durant laquelle les décès au sein des personnes âgées tendent à se concentrer dans un intervalle d'âge progressivement plus court. Effectivement, depuis le début des années 1990 au Japon, l'âge modal au décès continue d'augmenter alors que le niveau d'hétérogénéité des durées de vie au-delà de cet âge demeure inchangé. Ces résultats suggèrent que dans ce pays, nous assistons depuis quelques années à un déplacement de l'ensemble des durées de vie adultes vers des âges plus élevés, sans réduction parallèle de la dispersion des décès aux grands âges. Une telle situation serait d'ailleurs en conformité avec le *shifting mortality scenario*, soit un régime où la mortalité adulte est décalée vers des âges progressivement plus avancés avec le temps, sans toutefois que le profil selon l'âge soit modifié. Ajoutons qu'en France et au Canada, les femmes affichent aussi de tels développements depuis le début des années 2000.

Ces constats rejoignent ceux des travaux parus dernièrement sur le sujet, suggérant que quelques pays à faible mortalité pourraient maintenant être en phase de transition entre le régime de compression de la mortalité aux grands âges et le *shifting mortality scenario*. Notre étude est cependant celle qui offre le suivi le plus précis de ces changements récents en matière de mortalité. À titre d'exemple, elle est la première à clairement démontrer que l'augmentation soutenue et marquée de longue date de l'âge modal au décès s'effectue à un rythme considérablement moindre chez les Japonais et Japonaises depuis les années 2000. Ces résultats seraient-ils l'indication d'une limite à l'augmentation de la durée de vie modale au Japon ? Il s'agit là d'une question importante puisque depuis quelques décennies, le Japon constitue la référence parmi les pays à faible mortalité. À notre avis, la stabilité remarquable avec laquelle l'âge modal au décès a augmenté durant la seconde moitié du XX^e siècle porte à croire que malgré le ralentissement notable enregistré ces dernières années, l'accroissement de celui-ci se poursuivra au cours des prochaines années chez les deux sexes au Japon. Le rythme auquel l'âge modal au décès augmentera s'avère toutefois très difficile à prévoir. Ces résultats démontrent bien l'utilité des P-splines pour les recherches futures. En effet, la précision offerte par la méthode des P-splines demeurera un véritable

atout pour le suivi des changements subtils, néanmoins potentiellement lourds en conséquences, qu'afficheront les pays à faible mortalité dans les années à venir.

Nos travaux ont également permis de mettre en évidence le caractère inattendu des changements survenus récemment aux États-Unis. Durant la dernière décennie, l'âge modal au décès, soit la durée de vie la plus commune des adultes, a diminué de manière importante pendant plusieurs années consécutives chez les deux sexes. L'explication de ces résultats n'est pas évidente. Plusieurs études ont déjà souligné les inégalités socio-économiques profondes qui prévalent dans ce pays, de même que les écarts majeurs relatifs aux régimes d'assurance que détiennent les individus en cas de maladie, leur accès aux soins de santé et l'utilisation qu'ils en font. Ces éléments offrent quelques pistes d'explication. La situation aux États-Unis devra donc absolument être suivie de près au cours des prochaines années. Une analyse des changements dans la répartition des décès chez les personnes âgées selon le profil socio-économique serait notamment souhaitable puisqu'elle pourrait améliorer notre compréhension actuelle de la dynamique récente de la mortalité aux grands âges dans ce pays. Une analyse menée à une échelle géographique plus fine, telle celle réalisée pour le Canada et décrite ci-dessous, pourrait également s'avérer très révélatrice.

La *Base de données sur la longévité canadienne*, une base « satellite » de la *Human Mortality Database*, nous a permis d'entreprendre une analyse plus approfondie au Canada et de mettre en lumière les disparités provinciales en matière de mortalité adulte entre 1930 et 2007. Étant donné que certaines provinces canadiennes sont peu peuplées, la méthode de lissage par P-splines selon l'âge et les années de calendrier (dite en *deux dimensions*) a été utilisée pour ces travaux. Celle-ci consiste en une généralisation de la méthode de lissage par P-splines selon l'âge (dite en *une dimension*) employée pour l'étude impliquant les quatre pays à faible mortalité décrite ci-haut ; elle est particulièrement adaptée à l'étude de populations de petite taille. La méthode de lissage par P-splines en deux dimensions nous a ainsi permis de profiter pleinement du riche potentiel de la *Base*

de données sur la longévité canadienne et d'obtenir des résultats cohérents pour chacune des dix provinces canadiennes.

Malgré la baisse notable de la mortalité enregistrée dans toutes les provinces depuis 1930, les trajectoires temporelles de l'âge modal au décès et de l'écart type des âges au décès situés au-delà du mode nous ont révélé que les disparités provinciales avaient à peine diminué au cours de la période étudiée. Ajoutons que l'évolution de la durée de vie modale s'est effectuée à un rythme différent selon la province, entraînant ainsi des changements de hiérarchie. Soulignons notamment le rattrapage des provinces du centre du Canada, le Québec en particulier, et la détérioration de la situation dans certaines provinces de l'Atlantique. Le cas de Terre-Neuve-et-Labrador est spécialement préoccupant. En ce qui concerne l'hétérogénéité des durées de vie adultes, nous avons constaté que les provinces qui affichaient les niveaux d'écart type des âges au décès situés au-delà du mode les plus élevés ces dernières années faisaient généralement partie des provinces de l'Atlantique. À l'opposé, les provinces qui comportaient les niveaux les plus faibles appartenaient plutôt à celles de l'Ouest du pays.

Alors que nos travaux portant sur quatre pays à faible mortalité ont révélé que les Canadiennes semblaient avoir franchi tout récemment l'ère de la compression de la mortalité aux grands âges, cette seconde étude, menée sous une perspective géographique plus fine, nous permet de préciser davantage ce propos. En effet, nous constatons que ce sont spécifiquement les femmes issues de provinces de l'Ouest et du centre du pays qui semblent avoir amorcé leur transition entre le régime de compression de la mortalité aux grands âges et le *shifting mortality scenario*. Parmi les femmes des autres provinces, de même que les hommes de toutes les provinces canadiennes, le scénario de compression de la mortalité aux grands âges était cependant encore valide.

Bien que l'âge modal au décès ait pris une place prépondérante au sein des études sur la longévité humaine au cours des dix dernières années, nos connaissances relatives à son niveau et à son évolution dans des populations du passé demeurent très limitées. Ainsi, nous ignorons si l'âge le plus commun au décès chez

les adultes a augmenté, diminué, fluctué ou s'il a plutôt eu tendance à demeurer stable durant une très longue période de temps dans le passé. Ces informations s'avèrent cependant essentielles à notre compréhension de la longévité des adultes. Le *Registre de la population du Québec ancien*, une source exceptionnelle de données historiques reconnue pour sa couverture de la population du Québec ancien et sa fiabilité, nous a permis d'apporter un nouvel éclairage sur la durée de vie la plus commune des adultes au XVIII^e siècle. Plus précisément, nous avons pu suivre la progression de l'âge modal au décès chez les adultes canadiens-français durant la seconde moitié du XVIII^e siècle. Les résultats de notre analyse ont révélé que l'âge le plus commun au décès parmi ces adultes a continuellement augmenté entre 1740-1754 et 1785-1799. En effet, il est passé d'environ 73 ans à près de 76 ans chez les femmes et d'environ 70 à 74 ans chez les hommes.

Ces résultats, qui peuvent s'avérer surprenants à première vue, doivent cependant être situés dans le contexte particulier de la population canadienne-française de l'époque. Rappelons, en premier lieu, que les années de conflit au Québec ancien ayant culminé avec la Conquête britannique en 1759 ont été particulièrement défavorables à la survie des individus. En second lieu, les départs occasionnés par la défaite française ont ensuite eu pour effet d'augmenter la proportion de la population vivant en milieu rural chez les Canadiens-français. Comme la mortalité était plus faible à la campagne que dans les villes à cette époque, ces facteurs pourraient expliquer l'accroissement de la durée de vie la plus commune des Canadiens-français durant la période étudiée.

Ajoutons que l'urbanisation massive ultérieure de la colonie suggère que le rythme d'augmentation de l'âge modal au décès observé chez les Canadiens-français durant la seconde moitié du XVIII^e siècle pourrait s'amenuiser après 1800. Voilà également ce que laisse entendre la comparaison des résultats pour la période 1785-1799 avec ceux issus de notre étude portant sur les provinces canadiennes ; les estimations de l'âge modal au décès au Québec du début des années 1930 se situent à 78 ans environ chez les femmes et à un peu plus de 76 ans chez les hommes. En somme, bien que le *Registre de la population du Québec ancien*

soit un véritable laboratoire de population, les conditions de vie particulières de la population Canadienne-française durant la seconde moitié du XVIII^e siècle font en sorte que le potentiel de généralisation de ces résultats relatifs à la durée de vie modale demeure difficile à évaluer. Des comparaisons avec d'autres sources de données, qui permettraient le suivi de populations du passé durant plusieurs années, seraient grandement souhaitables.

Au terme de cette thèse, nous considérons qu'en plus de donner priorité aux sujets déjà mentionnés ci-haut, les recherches futures sur le thème de la longévité des adultes auraient également avantage à considérer les trois avenues suivantes : exploiter tout le potentiel qu'offrent les P-splines, proposer des mesures de dispersion des décès aux grands âges plus raffinées et analyser les changements dans la répartition des décès chez les personnes âgées selon les causes de décès.

En effet, bien que nos recherches aient démontré la pertinence de l'approche de lissage par P-splines pour l'étude de la longévité des adultes, le potentiel de cette méthode dans le contexte de l'analyse de la mortalité n'est pas encore exploité à sa pleine capacité. Rappelons que par définition, cette méthode combine les notions de B-splines et de fonction de vraisemblance pénalisée. Par conséquent, si la pénalité utilisée dans le cadre de notre recherche avait été remplacée par une pénalité spécifiquement adaptée à l'étude de la mortalité, l'approche par P-splines se serait avérée encore plus utile pour nos travaux. Une telle pénalité nous aurait entre autres permis d'évaluer l'ampleur du compromis effectué lorsque des modèles de mortalité (ex. modèle logistique) sont utilisés plutôt que l'approche de lissage par P-splines.

Ensuite, la proposition de mesures de dispersion plus raffinées, qui permettraient de dresser un portrait plus complet de l'hétérogénéité des décès aux grands âges, s'avère pour sa part essentielle. En effet, des changements mineurs en matière d'écart type des durées de vie au-delà de l'âge modal au décès peuvent refléter des modifications énormes du nombre de personnes âgées et très âgées. Des mesures qui mettraient l'accent sur des changements dans la dispersion des décès

survenant dans un voisinage plus ou moins rapproché de l'âge modal au décès, tels les quantiles de la distribution des âges au décès, sont ainsi à privilégier.

L'analyse des changements dans la répartition des décès chez les personnes âgées selon les causes de décès est quant à elle pertinente puisque jusqu'à présent, nous ignorons si le *shifting mortality scenario* vers lequel certains pays à faible mortalité semblent se diriger s'accompagnera d'un report ou d'un allongement des épisodes de morbidité et d'invalidité. Ainsi, dans l'éventualité où la mortalité adulte serait effectivement décalée vers des âges progressivement plus avancés au fil du temps, sans modification du profil selon l'âge, les épisodes de morbidité et d'invalidité seraient-ils également décalés ou occuperaient-ils plutôt une part plus importante de la durée de vie typique des personnes âgées ?

Si les recherches futures donnent effectivement priorité à ces différents sujets, elles devraient largement contribuer à l'amélioration de notre compréhension de la dynamique des changements dans la répartition des décès chez les personnes âgées.

Bibliographie

- ADAMS, O. (1990). Life expectancy in Canada - An overview. *Health Reports*, 2(4):361–376.
- BASE DE DONNÉES SUR LA LONGÉVITÉ CANADIENNE - BDLIC (2010). Université de Montréal (Canada), University of California, Berkley (États-Unis) et Max Planck Institute for Demographic Research (Allemagne). Disponible au <http://www.bdlc.umontreal.ca> (Données téléchargées le 16 juillet 2010).
- BEARD, R. E. (1959). Note on some mathematical mortality models. In WOLSTENHOLME, G. E. W. et O'CONNOR, M., éditeurs : *The Lifespan of Animals*. Boston.
- BEARD, R. E. (1971). Some aspects of theories of mortality, cause of death analysis, forecasting and stochastic processes. In BRASS, W., éditeur : *Biological Aspects of Demography*. Taylor & Francis Ltd, London.
- BEAUDRY-GODIN, M. (2010). *La démographie des centenaires québécois : validation des âges au décès, mesure de la mortalité et composante familiale de la longévité*. Thèse de doctorat, Montréal, Université de Montréal, Département de démographie.
- BELL, F. C. et MILLER, M. L. (2005). *Life Tables for the United States Social Security Area 1900-2100*. Baltimore.
- BERTILLON, J. (1878). Discussion suite à l'article de Lexis : Sur la durée normale de la vie humaine et sur la théorie de la stabilité des rapports statistiques. *Annales de démographie internationale*, 2:460–461.
- BONGAARTS, J. (2005). Long range trends in adult mortality : Models and projection methods. *Demography*, 42(1):23–49.
- BONGAARTS, J. et FEENEY, G. (2002). How long do we live? *Population and*

- Development Review*, 28(1):13–29.
- BONGAARTS, J. et FEENEY, G. (2003). Estimating mean lifetime. *Proceedings of the National Academy of Science*, 100(23):13127–13133.
- BOOTH, H., MAINDONALD, J. et SMITH, L. (2002). Applying Lee-Carter under conditions of variable mortality decline. *Population Studies*, 56:325–336.
- BOURBEAU, R. et LEBEL, A. (2000). Mortality statistics for the oldest-old : An evaluation of Canadian data. *Demographic Research*, 2(2).
- BOURBEAU, R., LÉGARÉ, J. et ÉMOND, V. (1997). *New birth cohort life tables for Canada and Quebec, 1801-1991*. Statistics Canada (Cat. No. 91F0015MPE1997003).
- BOURBEAU, R. et OUELLETTE, N. (2010). The Canadian Human Mortality Database : A project worth the challenges! Presented at the Third Human Mortality Database Symposium, Paris, June 17–19.
- BOURBEAU, R., OUELLETTE, N. et LECOURS, C. (2010). La base de données sur la longévité canadienne : corrections apportées aux données sur les décès de 1921 à 2007. Montréal, Université de Montréal, Département de démographie.
- BRASS, W. (1971). On the scale of mortality. In BRASS, W., éditeur : *Biological Aspects of Demography*, pages 69–110. Taylor & Francis, London.
- BROUHNS, N., DENUIT, M. et VERMUNT, J. K. (2002). A Poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance : Mathematics & Economics*, 31:373–393.
- CAMARDA, C. G. (2008). *Smoothing Methods for the analysis of mortality development*. Thèse de doctorat, Universidad Carlos III de Madrid (Spain) and Max Planck Institute for Demographic Research (Germany).
- CAMARDA, C. G. (2009). *MortalitySmooth : Smoothing Poisson counts with P-splines*. R package version 1.0.
- CAMARDA, C. G. (2010). The challenge of smoothing mortality in small populations : A P-spline approach with R. Presented at the Third Human Mortality Database Symposium, Paris, June 17-19.
- CANADIAN HUMAN MORTALITY DATABASE - CHMD (2010). Université de Montréal (Canada), University of California, Berkley (USA), and

- Max Planck Institute for Demographic Research (Germany). Available at <http://www.bdlc.umontreal.ca> (Data downloaded on July 16 2010).
- CANUDAS-ROMO, V. (2008). The modal age at death and the shifting mortality hypothesis. *Demographic Research*, 19:1179–1204.
- CANUDAS-ROMO, V. (2010). Time trends and record values for longevity. *Demography*, 47(2):299–312.
- CARNES, B. A. et OLSHANSKY, S. J. (2007). A realist view of aging, mortality, and future longevity. *Population and Development Review*, 33(2):367–381.
- CARSTENSEN, B. (2007). Age-period-cohort models for the Lexis diagram. *Statistics in Medicine*, 26:3018–3045.
- CHARBONNEAU, H. (1975). *Vie et mort de nos ancêtres : étude démographique*. Les Presses de l'Université de Montréal, Montréal.
- CHARBONNEAU, H., DESJARDINS, B., GUILLEMETTE, A., LANDRY, Y., LÉGARÉ, J. et NAULT, F. (1987). *Naissance d'une population : les Français établis au Canada au XVIIe siècle*. Presses Universitaires de France, Les Presses de l'Université de Montréal, Paris et Montréal.
- CHARBONNEAU, H., DESJARDINS, B., LÉGARÉ, J. et HUBERT, D. (1996). La population française de la vallée du St-Laurent avant 1760. In COURVILLE, S., éditeur : *Atlas historique du Québec : Population et territoire*, pages 31–43. Presses de l'Université Laval, Québec.
- CHEUNG, S. L. K. et ROBINE, J.-M. (2007). Increase in common longevity and the compression of mortality : The case of Japan. *Population Studies*, 61(1):85–97.
- CHEUNG, S. L. K., ROBINE, J.-M. et CASELLI, G. (2008). The use of cohort and period data to explore changes in adult longevity in low mortality countries. *Genus*, LXIV(1-2):101–129.
- CHEUNG, S. L. K., ROBINE, J.-M., PACCAUD, F. et MARAZZI, A. (2009). Dissecting the compression of mortality in Switzerland, 1876-2005. *Demographic Research*, 21:569–598.
- CHEUNG, S. L. K., ROBINE, J.-M., TU, E. J.-C. et CASELLI, G. (2005). Three dimensions of the survival curve : Horizontalization, verticalization, and longevity extension. *Demography*, 42(2):243–258.

- CHRISTENSEN, K., DOBLHAMMER, G., RAU, R. et VAUPEL, J. W. (2009). Ageing populations : The challenges ahead. *The Lancet*, 374:1196–1208.
- CLAYTON, D. et SCHIFFLERS, E. (1987). Models for temporal variation in cancer rates. II : Age-period-cohort models. *Statistics in Medicine*, 6:469–481.
- CLEVELAND, W. S. (1979). Robust locally weighted regression and smoothing scatter plots. *Journal of the American Statistical Association*, 74:829–836.
- COMFORT, A. (1956). *The Biology of Senescence*. Routledge & Kegan Paul, London.
- COMFORT, A. (1964). *Aging : The Biology of Senescence*. Holt, Rinehart and Winston.
- CURRIE, I. D., DURBAN, M. et EILERS, P. H. C. (2004). Smoothing and forecasting mortality rates. *Statistical Modelling*, 4:279–298.
- CURRIE, I. D., DURBAN, M. et EILERS, P. H. C. (2006). Generalized linear array models with applications to multidimensional smoothing. *Journal of the Royal Statistical Society, (B)* 68:259–280.
- CZADO, C., DELWARDE, A. et DENUIT, M. (2005). Bayesian Poisson log-bilinear mortality projections. *Insurance : Mathematics and Economics*, 36:260–284.
- de BOOR, C. (1978). *A Practical Guide to Splines*. Springer, Berlin.
- DEPARCIEUX, A. (1746). Essai sur les probabilités de la durée de la vie humaine. *In Essai sur les probabilités de la durée de la vie humaine (1746) et Addition à l'essai (1760)*, pages III–CLXI. Éditions de l'Institut national d'études démographiques, 2003, Paris.
- DEPARCIEUX, A. (1760). Addition à l'essai. *In Essai sur les probabilités de la durée de la vie humaine (1746) et Addition à l'essai (1760)*, pages CLXXIX–CCXI. Éditions de l'Institut national d'études démographiques, 2003, Paris.
- DESJARDINS, B. (1998). Le registre de la population du Québec ancien. *Annales de démographie historique*, (2):215–226.
- DUPÂQUIER, J. (1996). *L'invention de la table de mortalité*. Presses Universitaires de France, Paris.
- EAKIN, T. et WITTEN, M. (1995). How square is the survival curve of a given species? *Experimental Gerontology*, 30(1):33–64.

- EDWARDS, R. D. et TULJAPURKAR, S. (2005). Inequality in life spans and a new perspective on mortality convergence across industrialized countries. *Population and Development Review*, 31(4):645–674.
- EFRON, B. et TIBSHIRANI, R. J. (1993). *An Introduction to the Bootstrap*. Monographs on Statistics and Applied Probability. Chapman & Hall, London.
- EILERS, P. H. C., CURRIE, I. D. et DURBAN, M. (2006). Fast and compact smoothing on large multidimensional grids. *Computational Statistics and Data Analysis*, 50:61–76.
- EILERS, P. H. C. et MARX, B. D. (1996). Flexible smoothing with B -splines and penalties (with discussion). *Statistical Science*, 11(2):89–102.
- EILERS, P. H. C. et MARX, B. D. (2009). The craft of smoothing. Course handouts, Irish Statistical Association Short course on “P-splines Master-class”, Dublin, October 2–3.
- EUBANK, R. L. (1988). *Spline Smoothing and Nonparametric Regression*. Dekker, New York.
- FIELD, N. C. (1980). Temporal trends and spatial patterns of mortality in Canada. In BARRETT, F. A., éditeur : *Canadian Studies in Medical Geography*, pages 32–58. Department of Geography, Atkinson College, York University, Toronto.
- FORFAR, D. O., MCCUTCHEON, J. J. et WILKIE, A. D. (1988). On graduation by mathematical formula. *Journal of the Institute of Actuaries*, 115:1–149.
- FRIEDMAN, J. et SILVERMAN, B. W. (1989). Flexible parsimonious smoothing and additive modeling (with discussion). *Technometrics*, 31:3–39.
- FRIES, J. F. (1980). Aging, natural death, and the compression of morbidity. *New England Journal of Medicine*, 303:130–135.
- FRIES, J. F. (1984). The compression of morbidity : Miscellaneous comments about a theme. *The Gerontologist*, 24:354–359.
- GAGNON, A. et MAZAN, R. (2009). Does exposure to infectious diseases in infancy affect old-age mortality? Evidence from a pre-industrial population. *Social Science & Medicine*, 68:1609–1616.
- GAVRILOV, L. A. et GAVRILOVA, N. S. (1991). *The Biology of Lifespan : A Quantitative Approach*. Harwood Academic Publishers. English translation of

- Russian edition published in 1986, New York.
- GOOD, I. J. (1969). Some applications of the singular decomposition of a matrix. *Technometrics*, 11:823–831.
- GORDON, T. J. (1980). Changing mortality patterns. In *The Future of Life Expectancy : Economic, Social and Medical Implications of Living Longer*, pages 1–23. Society of Actuaries, Chicago.
- GRAUNT, J. (1662). *Natural and Political Observations Mentioned in a Following Index and Made upon the Bills of Mortality*. T. Raycroft, London.
- HALD, A. (1990). *A History of Probability and Statistics and Their Applications before 1750*. Wiley, New York.
- HALLEY, E. (1693). An estimate of the degrees of mortality of mankind, drawn from curious tables of the births and funerals at the city of Breslaw. *Philosophical Transactions of the Royal Society of London*, 17(196):576–610.
- HELIGMAN, L. et POLLARD, J. H. (1980). The age pattern of mortality. *Journal of the Institute of Actuaries*, 107(1):49–80.
- HILL, G. (1993). The entropy of the survival curve : An alternative measure. *Canadian Studies in Population*, 20(1):43–57.
- HIMES, C. L., PRESTON, S. H. et CONDRAN, G. A. (1994). A relational model of mortality at older ages in low mortality countries. *Population Studies*, 48:269–291.
- HORIUCHI, S. (2003). Interspecies differences in the life span distribution : Humans versus invertebrates, in J. R. Carey and S. Tuljapurkar, éditeurs. *Population and Development Review*, 29(Suppl.):127–151.
- HUMAN MORTALITY DATABASE - HMD (2010). University of California, Berkeley (USA) and Max Planck Institute for Demographic Research (Germany). Available at <http://www.mortality.org> (Data downloaded on March 1 2010).
- HÄRDLE, W. (1990). *Applied Nonparametric Regression*. Cambridge University Press.
- INSTITUT NATIONAL D'ÉTUDES DÉMOGRAPHIQUES (1970). *Essais d'arithmétique politique (par Willem Kersseboom)*. Paris.

- JEUNE, B. et VAUPEL, J. W. (1995). *Exceptional Longevity : From Prehistory to the Present*. Odense University Press, Odense.
- KANNISTO, V. (1992). Presentation at a workshop on “Old age mortality” held at Odense University, Odense, June.
- KANNISTO, V. (1994). *Development of Oldest-Old Mortality, 1950-1990 : Evidence from 28 Developed Countries*, Monographs on Population Aging, 1. Odense University Press, Odense.
- KANNISTO, V. (1996). *The Advancing Frontier of Survival Life Tables for Old Age*, Monographs on Population Aging, 3. Odense University Press, Odense.
- KANNISTO, V. (2000). Measuring the compression of mortality. *Demographic Research*, 3(6).
- KANNISTO, V. (2001). Mode and dispersion of the length of life. *Population : An English Selection*, 13(1):159–171.
- KANNISTO, V. (2007). Central and dispersion indicators of individual life duration : New methods. In ROBINE, J.-M., CRIMMINS, E., HORIUCHI, S. et ZENG, Y., éditeurs : *Human longevity, individual life duration, and the growth of the oldest-old population*, pages 111–129. Springer, Netherlands.
- KANNISTO, V., LAURISTEN, J., THATCHER, A. R. et VAUPEL, J. W. (1994). Reductions in mortality at advanced ages : Several decades of evidence from 27 countries. *Population and Development Review*, 20(4):793–809.
- KERSSEBOOM, W. (1738-1742). *Verhandeling tot een proeve om te weeten de probable menigte des volks in de provintie van Hollandt en West-Vrieslandt [Essai sur une preuve visant à déterminer la taille probable de la population de la Hollande et de la Frise occidentale]*. La Haye.
- KEYFITZ, N. (1977). *Applied Mathematical Demography*. John Wiley and Sons, New York.
- KLEIN, J. P. et MOESCHBERGER, M. L. (1997). *Survival Analysis. Techniques for Censored and Truncated Data*. Springer, New York.
- KOISSI, M.-C., SHAPIRO, A. F. et HÖGNÄS, G. (2006). Evaluating and extending the Lee-Carter model for mortality forecasting : Bootstrap confidence interval. *Insurance : Mathematics and Economics*, 38:1–20.

- KOOPERBERG, C. et STONE, C. T. (1991). A study of logspline density estimation. *Computational Statistics & Data Analysis*, 12:327–347.
- KOOPERBERG, C. et STONE, C. T. (1992). Logspline density estimation for censored data. *Journal of Computational and Graphical Statistics*, 1:301–328.
- LACROIX, C. (2009). Mortalité adulte et longévité exceptionnelle au Québec ancien. Mémoire de maîtrise, Montréal, Université de Montréal, Département de démographie.
- LAPLACE, P. S. (1812). *Théorie analytique des probabilités*. Veuve Courcier, Paris.
- LE BRAS, H. (1976). Lois de mortalité et âge limite. *Population*, 31(3):655–692.
- LEE, R. C. et CARTER, L. R. (1992). Modeling and forecasting U.S. mortality. *Journal of the American Statistical Association*, 87:659–675.
- LEIBNIZ, G. W. (1680). Essay de quelques raisonnements nouveaux sur la vie humaine et sur le nombre des hommes. In ROHRBASSER, J. M. et VÉRON, J., éditeurs : *Leibniz et les raisonnements sur la vie humaine*, pages 105–123. Éditions de l’Institut national d’études démographiques, 2001, Paris.
- LEXIS, W. (1877). *Zur Theorie der Massenerscheinungen in der menschlichen Gesellschaft [Sur la théorie des phénomènes de masse dans la société humaine]*. Fr. Wagner’sche Buchhandlung, Freiburg i.B.
- LEXIS, W. (1878). Sur la durée normale de la vie humaine et sur la théorie de la stabilité des rapports statistiques. *Annales de démographie internationale*, 2:447–460.
- LYNCH, S. M. et BROWN, J. S. (2001). Reconsidering mortality compression and deceleration : An alternative model of mortality rates. *Demography*, 38(1):79–95.
- MAKEHAM, W. M. (1860). On the law of mortality and the construction of annuity tables. *Journal of the Institute of Actuaries*, 8:301–308.
- MANTON, K. G., STALLARD, E. et TOLLEY, H. D. (1991). Limits to human life expectancy : Evidence, prospects and implications. *Population and Development Review*, 17(4):603–637.

- MANTON, K. G. et TOLLEY, H. D. (1991). Rectangularization of the survival curve. *Journal of Aging and Health*, 3(2):172–193.
- MANUEL, D. G. et HOCKIN, J. (2000). Recent trends in provincial life expectancy. *Canadian Journal of Public Health*, 91(2):118–119.
- MARTEL, S. (2002). Évolution de la dispersion des durées de vie au Québec au cours du XXe siècle. Mémoire de maîtrise, Montréal, Université de Montréal, Département de démographie.
- MARTEL, S. et BOURBEAU, R. (2003). Compression de la mortalité et rectangulation de la courbe de survie au Québec au cours du XXe siècle. *Cahiers québécois de démographie*, 32(1):43–75.
- MCCULLAGH, P. et NELDER, J. A. (1989). *Generalized Linear Models*. Monographs on Statistics and Applied Probability. Chapman & Hall, London, second edition.
- MURRAY, C. J., KULKARNI, S. C., MICHAUD, C., TOMIJIMA, N., BULZACHELLI, M. T., IANDIORIO, T. J. et EZZATI, M. (2006). Eight Americas : investigating mortality disparities across races, counties, and race-counties in the United States. *PLoS Medicine*, 3(9):1513–1524.
- MYERS, G. C. et MANTON, K. G. (1984a). Compression of mortality : Myth or reality? *The Gerontologist*, 24:346–353.
- MYERS, G. C. et MANTON, K. G. (1984b). Recent changes in the U.S. age at death distribution : Further observations. *The Gerontologist*, 24:572–575.
- NAGNUR, D. (1986). Rectangularization of the survival curve and entropy : The Canadian experience, 1921-1981. *Canadian Studies in Population*, 13:83–102.
- NAGNUR, D., CURRIE, S. G. et HEATH, B. (1981). *Évaluation de la qualité des statistiques de l'état civil (Étude pilote)*.
- NAULT, F. (1997). Narrowing mortality gaps, 1978 to 1995. *Health Reports*, 9(1):35–41.
- NELDER, J. A. et WEDDERBURN, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society*, (A) 135:370–384.
- NUSSELDER, W. J. et MACKENBACH, J. P. (1996). Rectangularization of the survival curve in the Netherlands, 1950-1992. *The Gerontologist*, 36:773–782.

- NUSSELDER, W. J. et MACKENBACH, J. P. (1997). Rectangularization of the survival curve in the Netherlands : An analysis of underlying causes of death. *Journal of Gerontology : Social Sciences*, 52B:S145–S154.
- ODGEN, R. T. (1997). *Essential Wavelets for Statistical Applications and Data Analysis*. Birkhäuser, Boston.
- OFFICE FÉDÉRAL DE LA STATISTIQUE SUISSE (2010). Données téléchargées à partir de <http://www.bfs.admin.ch/bfs/portal/fr/index.html>.
- OMRAN, A. R. (1971). The epidemiologic transition : A theory of the epidemiology of population change. *The Milbank Memorial Fund Quarterly*, 49(4):509–538.
- O’SULLIVAN, F. (1988). A statistical perspective on ill-posed inverse problems (with discussion). *Statistical Science*, 1:505–527.
- OUELLETTE, N. et BOURBEAU, R. (2009). Changes in the age-at-death distribution in low mortality countries : A nonparametric approach. Presented at the XXVI International Population Conference of the International Union for the Scientific Study of Population, Marrakech, September 27 – October 2.
- OUELLETTE, N. et BOURBEAU, R. (2010). Continuous summary measures for changes in the distribution of ages at death. Presented at the Annual Meeting of the Population Association of America, Dallas, April 15–17.
- OUELLETTE, N., BOURBEAU, R. et CAMARDA, C. G. (2010). Regional disparities in Canadian adult mortality : A comparative study based on smoothed mortality surfaces and age-at-death distributions. Submitted to *Canadian Studies in Population*.
- PACCAUD, F., PINTO, C. S., MARAZZI, A. et MILI, J. (1998). Age at death and rectangularization of the survival curve : Trends in Switzerland 1969-1994. *Journal of Epidemiology and Community Health*, 52(7):412–415.
- PAQUETTE, L. (2006). Projections de la mortalité pour le Canada, les provinces et les territoires 2003-2056 : comparaison de deux méthodes. Mémoire de maîtrise, Montréal, Université de Montréal, Département de démographie.
- PEARSON, K. (1902). On the modal value of an organ or character. *Biometrika*, 1(2):260–261.

- PERKS, W. (1932). On some experiments on the graduation of mortality statistics. *Journal of the Institute of Actuaries*, 63:12–40.
- PROGRAMME DE RECHERCHE EN DÉMOGRAPHIE HISTORIQUE - PRDH (2010). Les registres paroissiaux au Québec. Université de Montréal. Disponible au <http://www.genealogie.umontreal.ca/fr/lesParoisses.htm> (Consulté le 19 novembre 2010).
- PRUD'HOMME, P. (2007). Évolution de la mortalité différentielle selon la province au Canada. Mémoire de maîtrise, Montréal, Université de Montréal, Département de démographie.
- QUETELET, A. (1835). *Sur l'homme et le développement de ses facultés ou Essai de physique sociale*. Bossange & Co, Londres.
- QUETELET, A. (1871). *Anthropométrie ou mesure des différentes facultés de l'homme*. Muquardt, Bruxelles.
- R DEVELOPMENT CORE TEAM (2010). *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- REGISTRE DE LA POPULATION DU QUÉBEC ANCIEN - RPQA (2009). Programme de recherche en démographie historique, Université de Montréal (Québec, Canada). Disponible au <http://www.genealogie.umontreal.ca> (Données obtenues le 10 décembre 2009).
- RENSHAW, A. et HABERMAN, S. (2003a). Lee-Carter mortality forecasting : A parallel generalized linear modelling approach for England and Wales mortality projections. *Journal of the Royal Statistical Society (Applied Statistics)*, (C) 52(1):119–137.
- RENSHAW, A. et HABERMAN, S. (2003b). Lee-Carter mortality forecasting with age-specific enhancement. *Insurance : Mathematics and Economics*, 33:255–272.
- RENSHAW, A. et HABERMAN, S. (2006). A cohort-based extension to the Lee-Carter model for mortality reduction factors. *Insurance : Mathematics and Economics*, 38:556–570.

- ROBINE, J.-M. (2001). Redéfinir les phases de la transition épidémiologique à travers l'étude de la dispersion des durées de vie : le cas de la France. *Population*, 56(1-2):199–222.
- ROBINE, J.-M. et CHEUNG, S. (2009). Nouvelles observations sur la longévité humaine. *Revue économique*, 59(5):941–953.
- ROBINE, J.-M., CHEUNG, S. L. K., THATCHER, A. R. et HORIUCHI, S. (2006). What can be learnt by studying the adult modal age at death. Presented at the Population Association of America Annual Meeting, Los Angeles, March 30 – April 1.
- ROBINE, J.-M., SAITO, Y. et JAGGER, C. (2009). The relationship between longevity and healthy life expectancy. *Quality in Aging*, 10(2):5–14.
- ROTHENBERG, R., LENTZNER, H. R. et PARKER, R. A. (1991). Population aging patterns : The expansion of mortality. *Journal of Gerontology : Social Sciences*, 46(2):S66–S70.
- SCHNEIDER, E. L. et BRODY, J. A. (1983). Aging, natural death, and the compression of morbidity : Another view. *New England Journal of Medicine*, 304:854–855.
- SILVERMAN, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.
- STATISTICS CANADA (2010). *Deaths, 2007*. Health Statistics Division (Cat. No. 84F0211X), Ottawa.
- STREHLER, B. L. (1975). Implications of aging research for society. *Federation Proceedings*, 34(1):5–8.
- STRUYCK, N. (1740). *Inleiding tot de algemeene geografie, beneevens eenige sterrekundige en andere verhandelingen [Introduction à la géographie générale ainsi qu'à certains essais astronomiques et autres]*. Amsterdam.
- STRUYCK, N. (1753). *Nader ontdekkingen noopens dey staet van't menschelyk geslagt [Autres découvertes à propos de l'état de la race humaine]*. Amsterdam.
- SÜSSMILCH, J. P. (1765). *L'ordre divin*. 3ème ed. Traduit par J. Hecht. Éditions de l'Institut national d'études démographiques, 1979, Paris.

- THATCHER, A., KANNISTO, V. et VAUPEL, J. W. (1998). *The force of mortality at ages 80 to 120*, Monographs on Population Aging, 5. Odense University Press, Odense.
- THATCHER, A. R. (1999). The long-term pattern of adult mortality and the highest attained age. *Journal of the Royal Statistical Society*, (A) 162 (Part 1):5–43.
- THATCHER, A. R., CHEUNG, S. L. K., HORIUCHI, S. et ROBINE, J.-M. (2010). The compression above the mode. *Demographic Research*, 22:505–538.
- THATCHER, A. R., KANNISTO, V. et ANDREEV, K. (2002). The survivor ratio method for estimating numbers at high ages. *Demographic Research*, 6(1):2–15.
- TROVATO, F. et LALU, N. M. (1995). The narrowing sex differential in mortality in Canada since 1971. *Canadian Studies in Population*, 22(2):145–167.
- TROVATO, F. et LALU, N. M. (1998). The contribution of cause-specific mortality in changing sex differences in life expectancy : Seven nations case study. *Social Biology*, 45(1-2):1–20.
- TROVATO, F. et LALU, N. M. (2001). Narrowing sex differences in life expectancy : Regional variations, 1971-1991. *Canadian Studies in Population*, 28(1):89–110.
- VAUPEL, J. W., MANTON, K. G. et STALLARD, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16:439–454.
- VAUPEL, J. W. et YASHIN, A. I. (1985). Heterogeneity's ruses : Some surprising effects of selection on population dynamics. *The American Statistician*, 39:176–185.
- VAUPEL, J. W., ZHANG, Z. et van RAALTE, A. (2009). Life expectancy and disparity. Presented at the Annual Meeting of the Population Association of America, Detroit, April 30 – May 2 2009.
- VAUPEL, J. W., ZHENGLIAN, W., ANDREEV, K. F. et YASHIN, A. I. (1997). *Population Data at a Glance : Shaded Contour Maps of Demographic Surfaces over Age and Time*, Monographs on Population Aging, 4. Odense University Press, Odense University. Available at <http://www.demogr.mpg.de/Papers/Books/Monograph4/PopData1.htm>.

- VEUGELERS, P. J., YIP, A. M. et MO, D. (2001). The north-south gradient in health : Analytic applications for public health. *Canadian Journal of Public Health*, 92(2):95–98.
- VINCENT, P. (1951). La mortalité des vieillards. *Population*, 6(2):181–204.
- VÉRON, J. et ROHRBASSER, J.-M. (2003). Wilhelm Lexis : The normal length of life as an expression of the “Nature of things”. *Population : An English Selection*, 58(3):303–322.
- WILKINS, R. (1980). *L'état de santé au Canada, 1926-1976*. Institut de recherches politiques, Montréal.
- WILMOTH, J. R. (1993). Computational methods for fitting and extrapolating the Lee-Carter model of mortality change. Technical report, Department of Demography, University of California, Berkeley.
- WILMOTH, J. R. (1997). In search of limits. In WACHTER, K. W. et FINCH, C. E., éditeurs : *Between Zeus and the Salmon : The Biodemography of Longevity*, pages 38–63. National Academy Press, Washington.
- WILMOTH, J. R. (2001). Les modèles âge-période-cohorte en démographie. In CASELLI, G., VALLIN, J. et WUNSCH, G., éditeurs : *Démographie : analyse et synthèse. I. La dynamique des populations*, pages 379–397. Éditions de l'Institut national d'études démographiques, Paris.
- WILMOTH, J. R., ANDREEV, K., JDANOV, D. et GLEI, D. A. (2007). Methods Protocol for the Human Mortality Database. Available at <http://www.mortality.org/Public/Docs/MethodsProtocol.pdf>.
- WILMOTH, J. R. et HORIUCHI, S. (1999). Rectangularization revisited : Variability of age at death within human populations. *Demography*, 36(4):475–495.
- ZANFONGNON, R. (2008). Évolution de la mortalité différentielle selon le sexe au Canada : variations régionales, 1921-2004. Mémoire de maîtrise, Montréal, Université de Montréal, Département de démographie.
- ZHANG, Z. et VAUPEL, J. W. (2009). The age separating early deaths from late deaths. *Demographic Research*, 20:721–730.