



## **Rapport final du projet**

*« Risques de décès aux âges extrêmes de la vie »*

Doctorante : Linh Hoang Khanh Dang  
Encadrant(e)s de thèse : France Meslé, Nadine Ouellette, Carlo Giovanni Camarda  
Financement par la Fondation d'entreprise SCOR pour la Science

**Paris 2021**

## I. Résumé du projet de thèse

Le projet de thèse a pour objectif d'estimer les risques de décès au-delà de 90 ans, à partir des données empiriques provenant de la France, de la Belgique et du Québec, ainsi que d'étudier l'évolution de ces risques de décès depuis les années 1980. Cette thèse vise non seulement à préciser le niveau des risques de décès mais aussi à apporter des éclaircissements sur la forme de la courbe de mortalité aux très grands âges, sujet qui donne encore lieu à d'importantes discussions. L'opinion est principalement partagée entre une croissance exponentielle du risque de décès avec l'âge selon la loi de Gompertz (1825) (Brouard, 2012 ; Gavrilov et Gavrilova, 2011, 2015 ; Gavrilova et Gavrilov, 2017) et une décélération de la mortalité aux âges plus avancés (Perks, 1932 ; Beard, 1959 ; Horiuchi et Wilmoth, 1998 ; Ouellette et al., 2014 ; Ouellette, 2016). La nécessité d'aboutir à des résultats robustes est d'autant plus cruciale que l'augmentation actuelle de l'espérance de vie provient essentiellement des gains obtenus grâce au recul de la mortalité aux grands âges. Or, la prévision concernant ce recul reste souvent imprécise, d'une part en raison du faible effectif des survivants aux très grands âges qui entraîne d'inévitables fluctuations dans les mesures de mortalité, et d'autre part du fait des biais possibles liés au problème d'exagération des âges. Un projet de recherche international regroupant plusieurs pays de l'OCDE a abouti à des premiers résultats sur la mortalité à partir de 110 ans (Maier et al., 2010), et de nouveaux avancements ont fait l'objet d'un second ouvrage intitulé « *Exceptional Lifespans* » (Maier et al., 2020). Pourtant, de nombreuses incertitudes persistent sur la mortalité en-dessous de 110 ans, et plus particulièrement entre 90 et 110 ans, classes d'âges qui constituent au fil des années une part de plus en plus importante de la population humaine. Le travail de thèse consiste en plusieurs étapes. Il faut d'abord s'assurer de la qualité des données sur les décès obtenus pour chaque population (française, belge et canadienne-française), puis se prononcer sur leur comparabilité, en déduire des estimations des effectifs de la population soumise au risque, et calculer précisément les risques de décès par âge. Ces résultats serviront ensuite de base aux ajustements par des modèles statistiques paramétriques traduisant des « lois » de mortalité. Finalement, l'exercice apportera un nouvel éclairage sur la forme de la courbe de mortalité aux très grands âges. Une fois l'évolution de la mortalité avec l'âge mieux connue, il sera possible de mettre à jour les tables de mortalité par génération pour la France, avec un plus grand raffinement des hypothèses sur les évolutions de mortalité aux âges avancés, le tout ayant pour but d'estimer l'espérance de vie pour toutes les générations nées au XXe siècle.

Compte tenu des objectifs scientifiques et du déroulement précédemment détaillés, ce rapport présente l'ensemble des travaux réalisés par Linh Hoang Khanh Dang (ci-après Linh Dang) sous le financement de la Fondation d'entreprise SCOR pour la Science.

## **II. État des données mises à disposition**

### **1. Qualité des données**

Avant même le début de cette thèse, l'INED disposait déjà de données de bonne qualité sur les décès survenus aux âges les plus avancés pour la France, la Belgique et le Québec. Pourtant, les sources de données, la manière dont elles ont été collectées et les procédures pour les valider sont notablement différentes entre les populations. Cela nécessite par conséquent un travail de comparaison minutieux pour s'assurer que les données mises à disposition sont comparables et que les différences existantes ne seront pas source de biais pour les analyses ultérieures d'estimation des niveaux et des évolutions de la mortalité aux âges avancés. Il s'avère ainsi indispensable de se documenter auprès des équipes qui ont effectivement mené le travail de validation pour chaque population. Ces échanges permettent de comprendre non seulement le cadre juridique auquel leur travail est soumis et les conséquences qui en résultent (e.g. délai de validation, degré de fiabilité), mais aussi les choix qui ont été faits délibérément par ces équipes de recherche et qui conditionnent par la suite les caractéristiques des populations représentées dans les bases de données.

Suite au séjour de recherche que Linh Dang a fait en compagnie de Jacques Vallin à Montpellier en novembre 2017 pour rencontrer Jean-Marie Robine, responsable de la validation des données françaises, la doctorante a effectué un séjour de recherche à Charleroi (Belgique) en août 2018 pour rencontrer Michel Poulain, responsable de la validation des données belges. Les derniers éléments d'information sur la validation des données québécoises ont été recueillis durant le séjour de recherche de Linh Dang à l'Université de Montréal (Canada) à l'été 2019.

### **2. Préparation des données pour l'analyse**

Un des atouts majeurs de ce projet tient au fait qu'il repose sur la disponibilité de données de très bonne qualité. Le tableau 1 donne une description de toutes les données disponibles sur lesquelles les analyses ont pu être menées.

À la différence des deux autres populations pour lesquelles des données individuelles sont disponibles pour les décès survenus à 90 ans et plus, les données françaises ne contiennent des données individuelles que sur les décès survenus à l'âge de 105 ans ou plus issues du RNIPP. Elles sont complétées par des données agrégées sur les décès survenus entre 90 et 104 ans révolus. Ces dernières sont issues du système d'état civil (ventilées par sexe, année de survenue du décès, année de naissance de la personne décédée et âge au moment du décès). Elles sont considérées de bonne qualité grâce à l'excellence du système d'état civil qui fonctionne sans interruption depuis la Révolution française, mais aussi suite à une procédure de validation des données de décès rigoureuse au-delà de 105 ans, où 99,7% des cas vérifiés entre 105 et 109 ans se sont révélés exacts (au jour près) et encore 92% des cas de supercentenaires (décédés à l'âge de 110 ans ou plus) (Ouellette et al., 2020).

**Tableau 1.** Caractéristiques des données disponibles dans les trois populations étudiées.

Caractéristiques	France		Belgique	Québec
Sources de données	INSEE		Registre national	Institut de la statistique du Québec
Données brutes	Nombre de décès par âge, année de naissance, année de décès et sexe	Nom et prénom, date et lieu de naissance, date et lieu de décès, sexe	Nom et prénom, dates de naissance et de décès, sexe	Nom et prénom, date et lieu de naissance, date et lieu de décès, sexe
Nombre d'observations	1 028 676		140 990	31 795
Âges	90-104	105+	90+	90+
Génération	1883-1901		1891-1904	1880-1896
Années de décès	1988-2016		1981-2015	1970-2009

L'identification des décès aux très grands âges en Belgique est faite à partir du Registre national. Il s'agit d'un système d'enregistrement des changements affectant le statut des individus où tous les événements d'état civil au même titre que les changements de résidence sont enregistrés de manière continue. La Belgique possède par ailleurs un système d'état civil fonctionnel depuis 1792, époque à laquelle elle était rattachée à la France. Le système a ensuite été maintenu sous le régime hollandais en 1815 et enfin par l'État Belge. Les données mises à disposition pour cette analyse couvrent les décès ayant eu lieu entre 1981 et 2015 et les générations éteintes nées sur le territoire belge, entre 1891 et 1904. Entre 90 et 100 ans, le bon fonctionnement du système d'état civil belge depuis un siècle garantit la fiabilité des données et aucune validation spécifique n'a été effectuée. À partir de 100 ans, la validation se fait exhaustivement au niveau individuel. Les risques d'erreur augmentent de plus en plus rapidement avec l'âge. À partir de 105 ans, une validation approfondie s'appuie non plus seulement sur les données issues du Registre National mais est complétée par des sources annexes (contacts avec les maisons de retraite, visites rendues à la personne supposée âgée de plus de 105 ans si elle est encore en vie, articles de presse et confrontation de l'information issue des diverses sources disponibles). Tous ces éléments complémentaires sont par ailleurs confrontés aux données du Registre National pour tous les événements démographiques ayant marqué la vie des intéressés afin d'assurer la validation de l'âge exceptionnel atteint par ces personnes.

Les données sur les décès au Québec proviennent d'abord de listes annuelles nominatives, fournies par l'Institut de la statistique du Québec, qui contiennent toutes les personnes réputées décédées à 90 ans ou plus d'après le registre de l'état civil québécois entre 1970 et 2009. La validation des données a été effectuée sur la base des registres paroissiaux tenus rigoureusement depuis le XVII<sup>e</sup> siècle dans le Québec ancien. Par conséquent, la population étudiée contient exclusivement des Canadiens-français catholiques et l'authentification de leur âge au décès s'effectue par un

couplage des données figurant sur le certificat de décès et l'acte de baptême. Les recensements canadiens de 1881, de 1901 et de 1911 sont aussi utilisés comme source complémentaire pour la validation. Les résultats de plusieurs études confirment la haute qualité de ces données (pour les 100-109 ans révolus, voir Beaudry-Godin (2010) et pour les 110 ans ou plus, voir Bourbeau et Desjardins (2002) et Desjardins et Bourbeau (2010)). Depuis ce jour, la validation systématique ne s'impose qu'au-delà de 105 ans dans cette population, soit là où la plupart des erreurs avaient été trouvées.

Ces sources de données et le travail important de validation constituent la richesse en quantité et en qualité des données mises à disposition, assurant en même temps l'originalité et la fiabilité des recherches menées dans le cadre de ce projet de thèse.

### III. Principaux résultats

#### 1. Trajectoire de mortalité la plus plausible aux âges extrêmes de la vie

À la fin de la première année de thèse, nous avons montré que les taux de mortalité observés à partir de 90 ans restent très élevés et affichent une tendance claire de poursuite de la croissance des risques de décès avec l'âge. Les premiers ajustements de quatre modèles paramétriques de mortalité « classiques » (Gompertz, Weibull, Kannisto (logistique simplifié), quadratique) sur les données de mortalité des femmes dans chaque population avaient été effectués. Dans cette première étape, une première comparaison des ajustements des données obtenues par les différents modèles a été effectuée grâce au test du Khi-deux. Lorsque le modèle est correct et les données sont fiables, pour chaque année d'âge, le Khi-deux suit une distribution à un degré de liberté. Si le modèle n'est pas correct (ou si les données ne sont pas fiables), la valeur du Khi-deux calculée sera plus élevée que celle fixée par la distribution de Khi-deux à un degré de liberté. Plus les valeurs de Khi-deux sont élevées, plus l'ajustement du modèle est défectueux. Sur le plan théorique, les statistiques de Khi-deux ont été développées pour effectuer des tests entre des modèles imbriqués (*nested models*), ce qui n'est pas notre cas, mais cette mesure peut donner des informations sur la performance relative des différents modèles (Thatcher et al., 1998). Selon cette méthode, le meilleur ajustement est produit par le modèle de Kannisto pour les Canadiennes-françaises, alors que c'est le modèle de Weibull qui fonctionne le mieux pour les Françaises et les Belges. Pour les hommes, le modèle de Gompertz fournit le meilleur résultat dans les trois populations. Quoique le choix des modèles diffère entre les populations, à ce stade de l'analyse, on peut provisoirement conclure que la mortalité continue toujours à augmenter pour les deux sexes aux âges avancés, mais à un rythme plus faible que ce que prédit le modèle de Gompertz chez les femmes.

En deuxième année, nous avons ajouté le modèle logistique de Beard qui, à la différence du modèle logistique simplifié de Kannisto, ne contraint pas son asymptote horizontale à l'unité et permet ainsi un ajustement plus flexible (voir l'ensemble des modèles utilisés et leur fonction mathématique respective en Annexe 1). Nous avons par ailleurs appliqué une batterie de critères de performance (ou de sélection) des modèles étudiés : nous avons ainsi utilisé non seulement le khi-deux, mais aussi la racine de l'erreur quadratique moyenne (*root mean square error*, RMSE), le *Akaike Information Criteria* (AIC) et le *Bayesian Information Criteria* (BIC). Chaque critère a

sa manière de définir un modèle optimal : si les deux premiers indicateurs mettent plus l'accent sur la précision de l'estimation (ou le « *closeness* » des estimations par rapport aux données observées), les deux derniers critères tiennent compte aussi de la parcimonie du modèle, pénalisant un nombre trop élevé de paramètres qui permet certes un meilleur ajustement mais alourdit les tâches d'estimation et rend la généralisation du modèle difficile. Si l'on donne la préférence à AIC ou BIC, les modèles de la famille logistique sortent toujours comme des modèles optimaux pour chaque population féminine incluse dans l'étude, confirmant le résultat trouvé à la fin de la première année de thèse. En outre, les trajectoires décrites par les différents modèles paramétriques ne diffèrent qu'après 100 ans, et dans la majorité des cas, le niveau estimé des risques de décès après l'âge de 110 ans dépasse la valeur 0,7 estimée antérieurement par Gampe (2010) comme le niveau du plateau de mortalité chez l'être humain.

**Tableau 2.** Modèle de mortalité optimal selon le critère de sélection retenu pour les femmes.

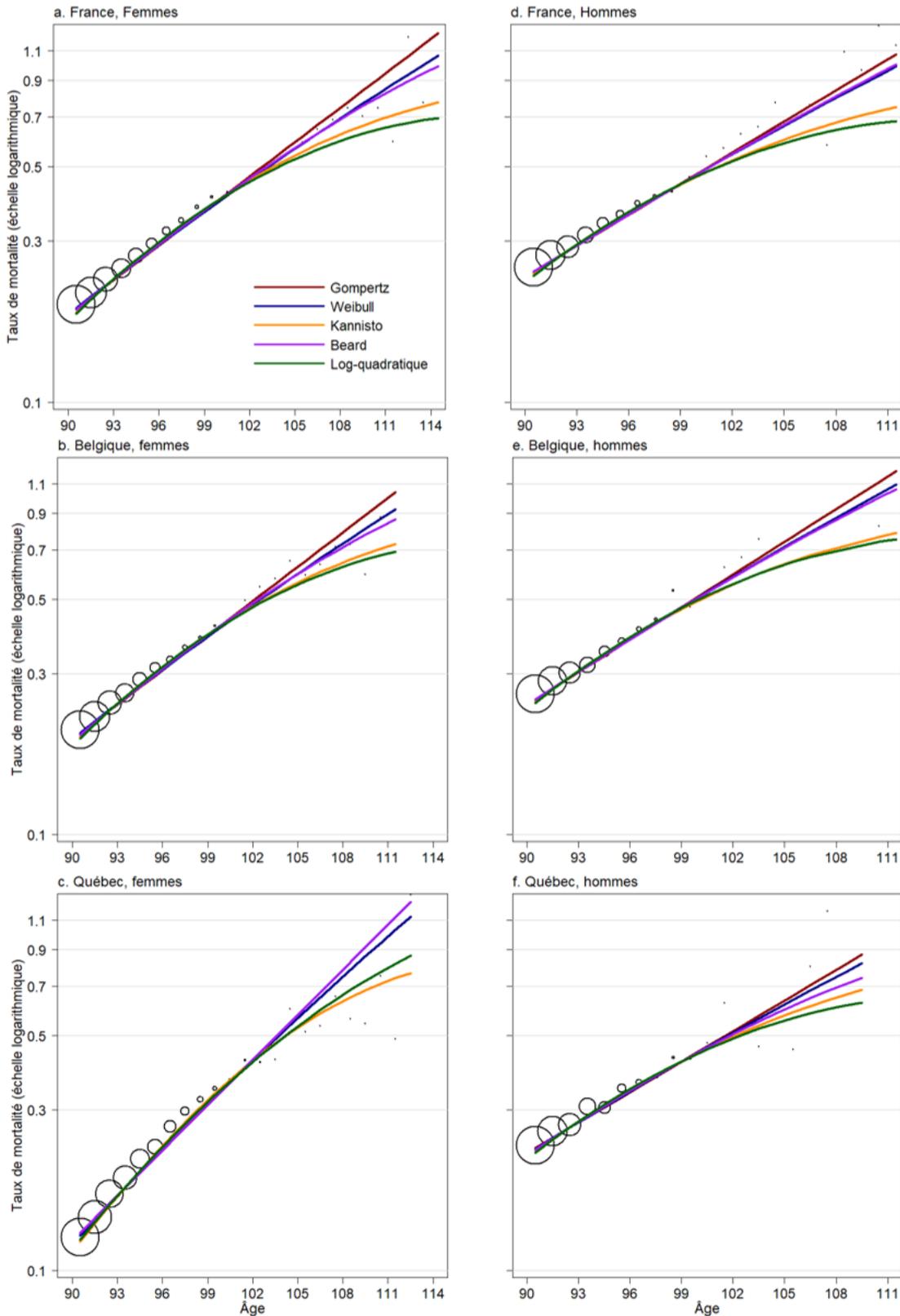
Populations (féminines)	Génération	Nb. d'obs.	Khi-deux	RMSE	AIC	BIC
Belges	1891-1904	104 661	Weibull	Gompertz	Beard	Beard
Canadiennes-françaises	1880-1896	20 917	Kannisto	Quadratique	Kannisto	Kannisto
Françaises	1883-1901	788 131	Weibull	Gompertz	Kannisto	Kannisto

L'ajustement par des modèles paramétriques « classiques » a également été réalisé pour les populations masculines (tableau 3). À la différence des femmes, la trajectoire de mortalité chez les hommes aux très grands âges est ajustée de façon optimale par une croissance exponentielle de type Gompertz. La surmortalité des hommes par rapport aux femmes peut partiellement expliquer cette différence, mais il est possible que même si la trajectoire sous-jacente de la mortalité masculine est celle de décélération, le faible nombre de survivants aux très grands âges ne soit pas suffisant pour la faire ressortir.

**Tableau 3.** Modèle de mortalité optimal selon chaque critère de sélection pour les hommes.

Populations (masculines)	Génération	Nb. d'obs	Khi-deux	RMSE	AIC	BIC
Belges	1891-1904	36 329	Gompertz	Gompertz	Gompertz	Gompertz
Canadiens-français	1880-1896	10 877	Gompertz	Gompertz	Weibull	Weibull
Français	1883-1901	240 545	Gompertz	Gompertz	Gompertz	Gompertz

**Figure 1.** Modélisation des taux de mortalité observés selon l'âge et le sexe pour les générations françaises (1883-1901), belges (1891-1904) et canadiennes-françaises (1880-1896).



## 2. La décélération de la mortalité aux très grands âges

Une des critiques récurrentes vis-à-vis des études de la mortalité aux âges extrêmes est la pratique de regrouper plusieurs générations d'individus soumises à des conditions de mortalité diverses (Gavrilov & Gavrilova, 2011 ; Brouard, 2012 ; Gavrilov & Gavrilova, 2015). Cette pratique permet d'augmenter le nombre d'observations, facilite les analyses statistiques ainsi que la fiabilité des résultats, mais introduit (ou renforce) en même temps l'hétérogénéité de la population sous étude. Ceci pose ainsi la question de savoir à quel point la décélération de la courbe de mortalité observée dans la présente étude comme dans d'autres publications mentionnées ci-dessus dans la littérature, reflète véritablement l'évolution de la mortalité humaine à ces âges, ou résulte plutôt d'un effet de changement de composition exagéré par le regroupement de plusieurs générations. Deux approches peuvent être suivies pour quantifier ce changement de composition et son impact sur la mortalité aux très grands âges, selon les caractéristiques des données mises à disposition, qu'elles soient au niveau individuel ou agrégé, qu'elles ne contiennent que des données sur les décès ou également d'autres informations sur les sujets observés. Si les données sont agrégées, quelles hypothèses pourrions-nous faire sur le caractère hétérogène non-observable qui distingue un individu (ou un groupe d'individus) d'un autre (ou des autres) ? Si les données sont disponibles au niveau individuel, comment pourrions-nous extraire les caractères hétérogènes observables et en faire une description synthétique et parlante sur la population considérée ? Telles sont les questions auxquelles la deuxième partie de la thèse cherche à répondre en mobilisant des méthodes de modélisation correspondant à chacun des cas de figure mentionnés, ceci sur la base des populations françaises, belges et canadiennes-françaises.

### 2.1. Hétérogénéité non-observable

Dans la première partie de notre analyse, l'ajustement des modèles paramétriques « classiques » de mortalité sur les données agrégées de chaque population est fait sous l'hypothèse que tous les individus suivent exactement une même trajectoire de mortalité et toute la population est homogène. Bien que largement réductrices, ces hypothèses étaient nécessaires pour avoir une vue générale sur la trajectoire de mortalité aux très grands âges et répondre à certaines critiques souvent adressées aux recherches dans ce domaine. Dans la suite des analyses, l'hypothèse de l'homogénéité de la population est supprimée mais on continue à supposer que tous les individus partagent exactement la même trajectoire de base. La modélisation lorsque l'hétérogénéité est non-observable ou observable se fonde principalement sur les modèles à risques proportionnels où la variable capturant l'hétérogénéité agit proportionnellement sur le risque de décès et où la force de mortalité individuelle est déterministe et définie par les modèles paramétriques représentant différents scénarios possibles, tels que présentés dans la première partie du présent rapport.

Lorsque l'hétérogénéité est non observable, la variable capturant l'hétérogénéité est une variable aléatoire suivant une distribution statistique prédéterminée. La force de mortalité individuelle va être le produit de deux éléments : une variable hypothétique  $z$  dont l'appellation la plus populaire est la variable de « fragilité » (Vaupel et al., 1979), et une fonction de mortalité par âge  $\mu_s(x)$  qui capture le calendrier de décès des individus dits « standards » ou de ceux dont la fragilité est à l'unité (i.e.,  $z = 1$ ). Cette notion de « fragilité » est à distinguer de ce que veut dire la « fragilité »

en gérontologie ; dans le cadre des modèles de fragilité, elle doit être interprétée comme un caractère hétérogène via lequel la composition de la population change, le plus probablement par des effets de sélection : plus l'individu est « fragile », plus il/elle sortira de la population dès les plus jeunes âges. Pour des raisons purement analytiques, la distribution statistique de la variable de fragilité  $z$  est souvent supposée être une distribution de type gamma dont l'un des deux paramètres détermine le coefficient de variation de  $z$ . La moyenne et la variance de  $z$  vont changer avec l'âge lorsque la composition de la population change, mais son coefficient de variation est constant. De fait, la fonction de la force de mortalité moyenne de la population va être la suivante :

$$\overline{\mu_x} = \bar{z}(x) \mu_s(x)$$

Où  $\bar{z}(x) = \frac{1}{1 + \gamma M(x)}$ ,  $M(x)$  étant la fonction de risques cumulés de l'âge  $x_0$  jusqu'à l'âge  $x$  considéré, et  $\gamma$  étant le coefficient de variation au carré de la variable  $z$ . La moyenne de  $z$  à l'âge  $x_0$  est souvent fixée à l'unité dans les modèles de fragilité,  $\gamma$  mesure de fait la variance de la variable  $z$  et peut être interprété comme le degré de l'hétérogénéité de la population en question.

L'objectif de cette sous-partie est double, d'une part, elle vise à voir comment l'introduction de l'hétérogénéité dans les modèles paramétriques ajuste les données de décès observées aux âges extrêmes de la vie, et d'autre part elle vise à démêler le rôle que joue le facteur de l'hétérogénéité dans la caractérisation de la forme décélérée de la mortalité moyenne de la population et à examiner de plus près la forme la plus plausible que pourrait prendre la mortalité individuelle, élément le plus prometteur pour avoir une idée de ce que c'est la « vraie » trajectoire de la mortalité humaine. Pour ces objectifs, dans le cadre de ce projet de thèse, nous avons choisi trois fonctions de mortalité individuelle représentatives de trois scénarios différents : la fonction de Gompertz pour décrire une force de mortalité individuelle qui augmente en continuité avec l'âge, la fonction de Kannisto pour décrire une force de mortalité individuelle décélérée et éventuellement plafonnée à l'unité, et la fonction log-quadratique dont la trajectoire de la mortalité individuelle peut prendre une forme de croissance, de décélération, ou même de diminution, selon les estimations des paramètres qui la constituent. Il faut noter que la fonction log-quadratique de la force de mortalité individuelle est à distinguer de celle de la force de mortalité de la population déjà employée dans la première partie. L'utilisation de la fonction log-quadratique dans les modèles de fragilité a été proposée en premier par Horiuchi (2003), avec des paramètres contenant en soi des bases d'interprétation gérontologique, qui lui donnent un certain pouvoir d'explication contrairement au modèle log-quadratique purement descriptif proposé par Coale et Horiuchi (1990).

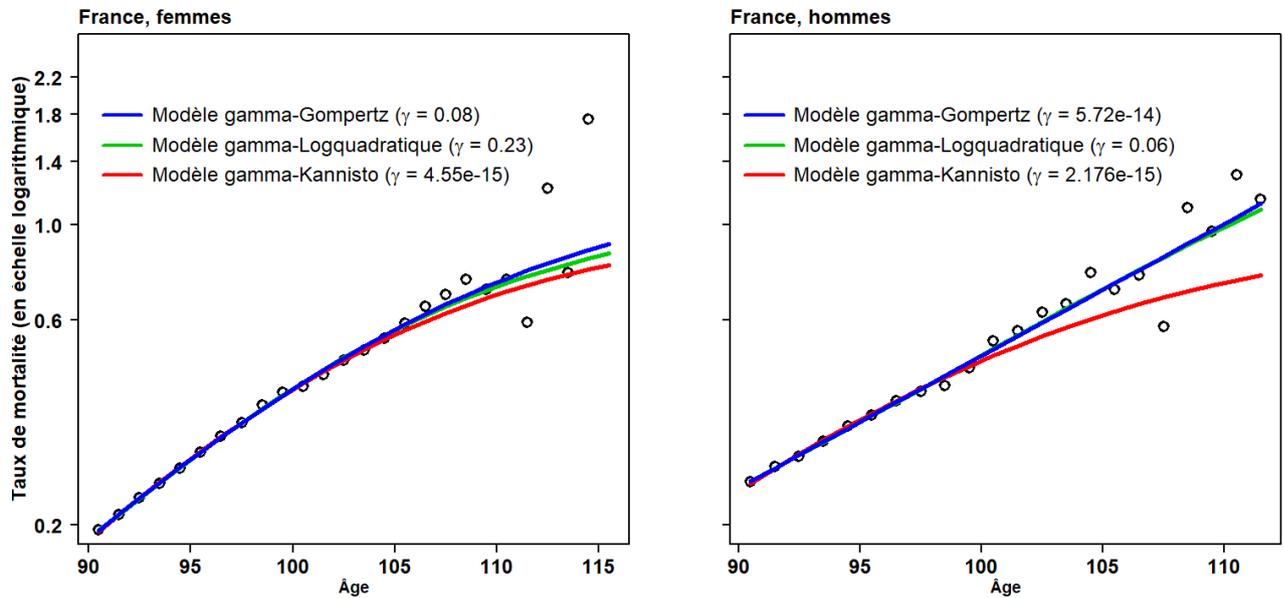
Le choix des modèles est fait encore une fois en fonction des critères de performance AIC et des intervalles de confiance. Dans ce rapport, seuls les résultats obtenus en s'appuyant sur les critères AIC seront présentés. On se reportera à la version finale de la thèse pour ceux basés sur les intervalles de confiance. Les fonctions de force de mortalité moyenne de la population correspondant à chaque fonction de mortalité individuelle mentionnée ci-dessus sont détaillées dans l'annexe 2 et les bases gérontologiques sur lesquelles le modèle gamma-log-quadratique a été développé peuvent être trouvées dans l'annexe 3 de ce rapport.

Les courbes ajustées par les modèles de fragilité étudiés sont présentées pour chacun des deux sexes pour la population française, belge et canadienne-française dans les figures 4, 5 et 6,

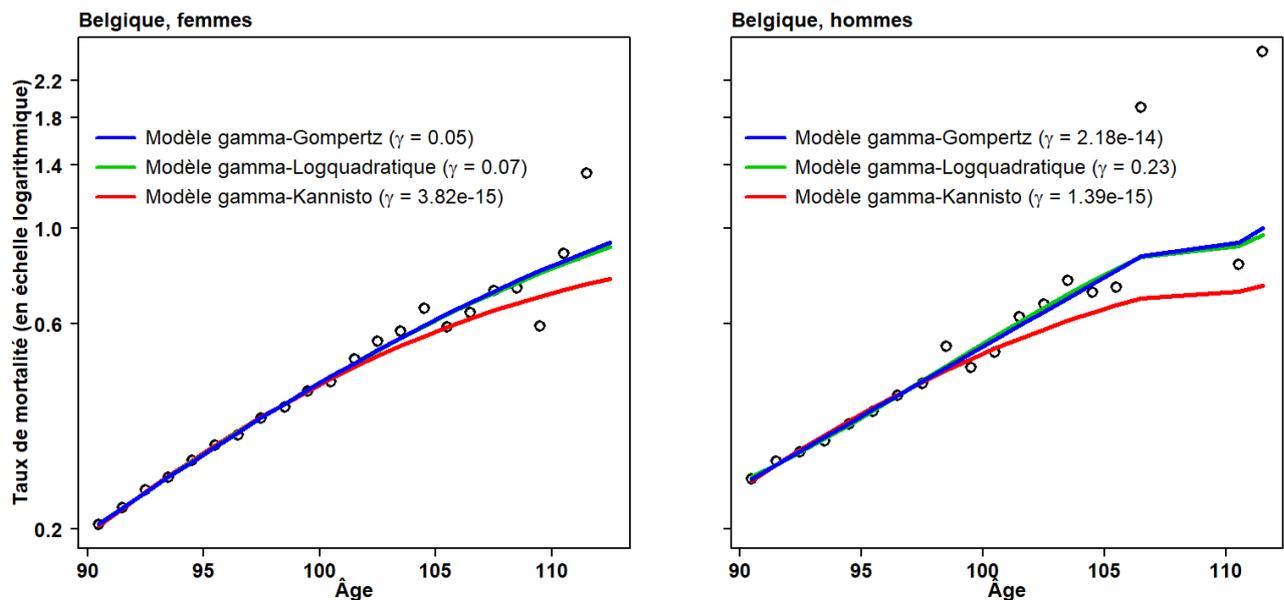
respectivement. La décélération de la force de mortalité est bien présente dans toutes les populations féminines mais reste moins prononcée chez les hommes. Pour les femmes, les estimations du modèle gamma-Gompertz sont toujours suivies de très près par les estimations du modèle gamma-log-quadratique. Les écarts sont parfois si petits, que l'on a l'impression que les courbes se superposent, notamment dans le cas des femmes belges et canadiennes-françaises (figure 5 et 6).

En comparant les valeurs des critères AIC données par modèle de fragilité, systématiquement pour chaque combinaison de population et sexe, on retrouve un même classement du niveau de performance dans les trois populations et pour les deux sexes. En premier, le modèle gamma-Gompertz apparaît comme le meilleur modèle, suivi par le modèle gamma-log-quadratique, et enfin par le modèle gamma-Kannisto (sauf dans le cas des femmes françaises où les valeurs de AIC du modèle gamma-Gompertz et du modèle gamma-log-quadratique ne diffèrent presque pas). Ce résultat n'est pas surprenant, car un développement complet du modèle gamma-Gompertz mène en fait à l'équation du modèle de Beard, modèle paramétrique de la famille logistique. La performance des modèles appartenant à cette famille dans l'ajustement des données s'est bien manifestée dans la première partie des résultats. Si l'on s'en remet aux critères AIC, c'est donc le modèle gamma-Gompertz qui doit être retenu. Cependant, à la simple lecture des courbes ajustées (figure 4), on voit que les estimations de la force de mortalité données par le modèle gamma-log-quadratique sont très proches de celles du modèle gamma-Gompertz. Dans le cas des femmes françaises, aux âges les plus élevés, les deux modèles s'éloignent des données observées et les courbes s'incurvent vers le bas. Bien que le modèle gamma-log-quadratique présente quelques désavantages (plus grand nombre de paramètres et absence d'une solution analytique pour la fonction de risques cumulées), avec l'amélioration actuelle des capacités de calcul, sa bonne performance relative, sa flexibilité dans la génération de courbes concaves comme convexes et les raisonnements biologiques qui le sous-tendent et lui donnent une capacité d'interprétation font de ce modèle un candidat prometteur dans l'ajustement des données de décès aux âges extrêmes de la vie.

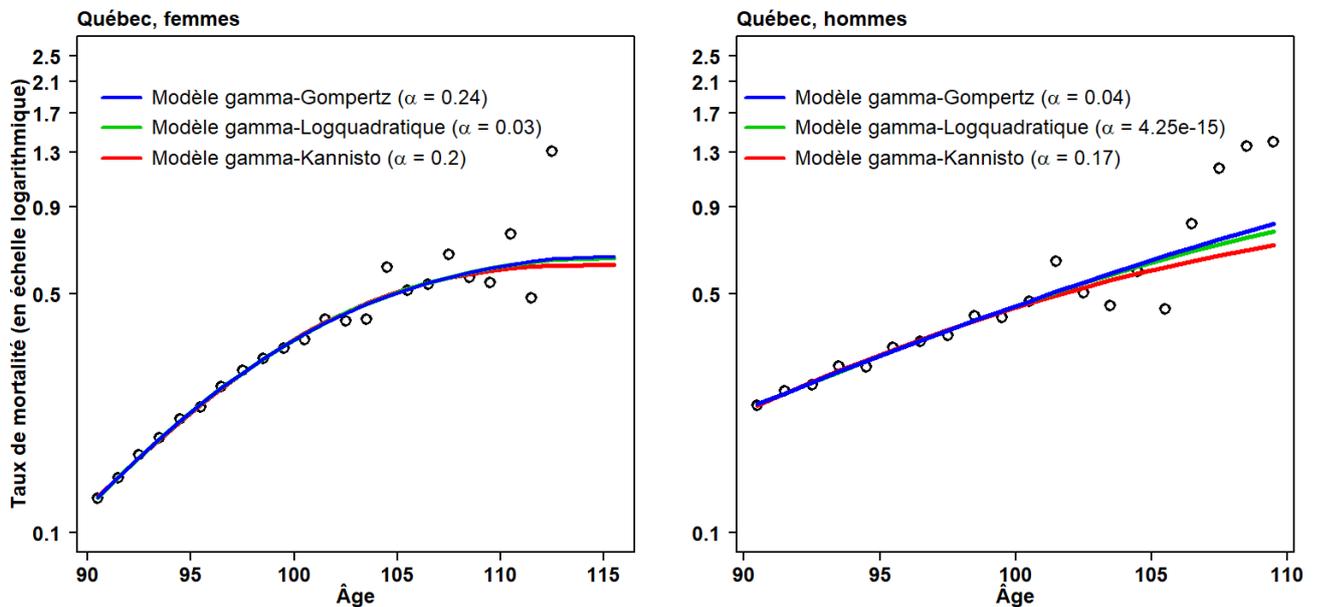
**Figure 4.** Trajectoire de mortalité modélisée par les modèles de fragilité sur la population française.



**Figure 5.** Trajectoire de mortalité aux très grands âges modélisée par les modèles de fragilité sur la population belge.



**Figure 6.** Trajectoire de mortalité aux très grands âges modélisée par les modèles de fragilité sur la population française-canadienne.



La comparaison de trois modèles de fragilité ayant pour base des fonctions de mortalité individuelle différentes nous a permis ainsi de confirmer la validité du modèle de Gompertz dans la caractérisation de la forme de la courbe de mortalité au niveau individuel sur les données actuelles, avant que l'agrégation de différentes trajectoires individuelles, ajustée par une distribution statistique gamma, aboutisse à une décélération nettement visible à partir d'un certain âge. Plus la variation de l'hétérogénéité dans la population est grande, et donc plus l'effet de sélection s'avère important, plus la courbe de mortalité agrégée adopte une forme log-concave, avec une tendance au plafonnement. Effectivement, en comparant l'estimation du paramètre  $\gamma$  obtenue lors de l'ajustement du modèle gamma-Gompertz, le meilleur modèle selon les critères AIC dans toutes les populations et pour tous les sexes, on trouve que plus la variance  $\gamma$  de la distribution de la fragilité est grande, et donc plus le niveau d'hétérogénéité dans la population est important, plus la forme log-concave de la courbe de mortalité devient apparente. La courbe de mortalité des Canadiennes-françaises, dont l'estimation de la variance de la variable capturant l'hétérogénéité est de  $\gamma = 0.239$ , plafonne beaucoup plus que celle des Françaises ( $\gamma = 0.082$ ) et celle des Belges ( $\gamma = 0.053$ ). Ce résultat rejoint en fait les remarques faites par Horiuchi et Wilmoth (1997, 1998) lorsqu'ils étudiaient les trajectoires de mortalité des causes de décès majeures. Dans leurs études, les auteurs trouvent que plus les maladies sont liées fortement à des facteurs de risques (telles que le cancer, la cirrhose du foie ou l'AVC hémorragique) et donc liées à des effets de sélection plus importants, plus leur trajectoire de mortalité est log-concave. Les maladies à moindre sélection telles que la pneumonie, la gastro-entérite ou l'insuffisance cardiaque congestive génèrent des trajectoires de mortalité plus log-convexes. Selon les auteurs, plus l'effet de la sélection est fort, que ce facteur de sélection soit environnemental ou génétique, plus la courbe de mortalité correspondant est log-concave (ou plafonne plus aisément). Ces trajectoires différentielles des causes de décès suggèrent que « les profils d'âge du risque total de

mortalité des individus, qui ne sont pas faussés par la sélection, pourraient être encore plus log-convexes que ceux des taux de mortalité au niveau agrégé provenant de maladies moins sélectives » (Horiuchi, 2003, traduction par la doctorante). Dans le cadre de ce projet de thèse, il n'a pas été possible de faire une comparaison avec les courbes de mortalité par causes de décès aux très grands âges. On se contente de voir que la forme de décélération de la courbe de mortalité omniprésente aux âges extrêmes dans les trois populations féminines est due principalement à l'hétérogénéité de la population qu'elle représente.

Une remarque mérite aussi d'être soulignée est que lorsque la fonction de risque individuelle est déjà marquée par une décélération de la mortalité, le rôle joué par la variable de « fragilité » capturant l'hétérogénéité dans la population diminue, voire devient presque négligeable. Lorsque la variance de la variable de l'hétérogénéité est presque nulle, le dénominateur de la fonction de mortalité moyenne de la population va s'approcher de l'unité et la fonction de la mortalité moyenne de la population équivaut à celle donnée à la mortalité individuelle. Ce n'est que chez les Canadiennes-françaises que la variance de la variable de « fragilité » dans le cadre du modèle gamma-Kannisto atteint une valeur non négligeable, quoique toujours plus faible que celle estimée pour les modèles gamma-Gompertz et gamma-log-quadratique. Or, pour cette population québécoise, une tendance au plafonnement est observée très clairement pour n'importe quel modèle de fragilité utilisé. Rappelons-nous aussi que dans la première partie des résultats, lorsque les modèles paramétriques classiques sont ajustés aux données observées et leur performance est classée selon le critère AIC, le modèle de Kannisto est choisi comme le modèle optimal pour les Canadiennes-françaises.

D'autre part, même si l'excellente performance du modèle gamma-Gompertz dans l'ajustement des données observées s'est bien maintenue aux âges les plus avancés, elle n'implique pas qu'un plateau de mortalité existe. L'existence d'une asymptote horizontale est une propriété mathématique intrinsèque du modèle gamma-Gompertz. Sa bonne performance tient d'abord en sa capacité de laisser baisser le rythme avec lequel les risques de décès s'accroissent avec l'âge grâce à l'inclusion d'un point d'inflexion, et peut-être moins nécessairement parce qu'il fixe un plafond que la mortalité humaine pourrait atteindre. La performance moins adéquate du modèle gamma-Kannisto parmi les trois modèles considérés a aussi démontré qu'un plafond fixé à l'unité pour la mortalité individuelle demeure trop bas, la trajectoire tracée par le modèle gamma-Kannisto dans les populations de femmes françaises et belges a dû s'éloigner beaucoup des points d'observation aux âges les plus élevés pour respecter l'asymptote qui lui a été dicté (figure 4 et 5). Ce n'est que dans le cas des femmes canadiennes-françaises qu'un plafonnement semble se dessiner assez clairement pour les trois modèles de fragilité considérés. Cette population est pourtant constituée d'un plus faible nombre d'observations parmi les trois populations considérées, avec un niveau d'hétérogénéité estimé plus élevé que celui des deux autres populations. La raison de telles différences exige bien encore d'autres recherches plus en profondeur sur les caractéristiques propres à chaque population qui dépassent pour le moment le cadre de cette thèse.

La validité du modèle de Gompertz pour la mortalité individuelle semble se manifester plus clairement chez les hommes. Une comparaison de performance des modèles basée sur les critères AIC affiche le modèle gamma-Gompertz comme leur meilleur candidat dans les trois populations.

De plus, une augmentation des risques est visible même jusqu'aux âges les plus élevés, et encore une fois, cette trajectoire peut être observée même au niveau agrégé des données. Une tendance à la croissance continue de la force de mortalité est de fait bien présente. Notons qu'une certaine discontinuité peut être observée pour les courbes ajustées dans la population belge masculine. Ceci est dû au fait qu'aucun décès masculin n'est enregistré entre 106 et 111 ans chez les hommes belges. Encore une fois, le nombre faible de survivants masculins à ces âges lance un appel à la prudence quant aux conclusions que l'on peut tirer sur la trajectoire de mortalité masculine. La faible décélération de la mortalité avec l'âge chez les hommes, voire même son absence apparente, quand on applique des modèles qui ont été spécifiquement développés pour capturer la décélération, mérite certainement un ré-examen dès que les données sur les décès masculins seront plus riches.

## **2.2. Hétérogénéité observable**

La critique sur la pratique de regrouper différentes générations, mentionnée au début de cette partie du rapport, a refait surface récemment avec la publication de l'article de Gavrilov et Gavrilova (2019) sur la « Gompertzialisation » de la courbe de mortalité par génération. Dans cet article reposant sur les décès survenus entre 80 et 105 ans au sein des générations américaines 1880-1899, les auteurs ont montré que la décélération (exprimée par le modèle logistique simplifié de Kannisto) est plus fréquemment trouvée dans les générations anciennes où la qualité des données pourrait être douteuse, alors que le modèle de Gompertz a une meilleure performance dans les générations les plus récentes. Dans nos résultats de la partie précédente, on a vu que lors de l'application des modèles de fragilité, dont la variable capturant l'hétérogénéité suit une distribution gamma, une mortalité individuelle gompertzienne se qualifie toujours comme le meilleur candidat pour décrire la trajectoire d'une mortalité décélérée avec l'âge dans toutes les populations étudiées. Il serait tentant de se demander quel résultat on obtiendrait en agrégeant les données pour chaque génération et non plus pour toute la population (i.e., ensemble des générations). La trajectoire décélérée obtenue par combinaison des trajectoires individuelles de mortalité gompertzienne finirait-elle pas devenir gompertzienne aussi ? L'idée est intéressante mais il ne faut pourtant pas oublier que le nombre d'observations aux très grands âges reste encore très faible et constitue une source importante de fluctuations non souhaitables dans les ajustements des modèles de mortalité, d'où la nécessité en premier lieu d'agréger les données de décès de plusieurs générations. Ainsi sur le plan statistique, l'ajustement des modèles de fragilité sur les données de chaque génération ne suggère pas encore de résultats suffisamment fiables. Cependant, l'idée de vérifier si la génération à laquelle l'individu appartient peut jouer un rôle dans sa trajectoire de mortalité mérite d'être poursuivie en s'appuyant sur une autre approche méthodologique.

En juin 2018, dans un article publié dans la revue *Science*, Barbi *et al.* déclaraient avoir pu fournir la meilleure preuve de l'existence d'un plateau de mortalité aux âges extrêmes chez l'être humain, à partir des données italiennes. Linh Dang, en collaboration avec ses encadrants ainsi qu'avec Jean-Marie Robine et Jacques Vallin, a appliqué les mêmes modèles paramétriques à risques proportionnels (*proportional hazard models*) sur les données individuelles des décédés français à

plus de 105 ans, pour voir si le même résultat pouvait s'observer dans cette population. Ces analyses ont ensuite été étendues aux deux autres populations.

À partir de ces données individuelles, l'analyse de survie demeure la méthode la plus répandue pour étudier l'évolution de la force de mortalité avec le temps. Les modèles de risques proportionnels y sont les plus fréquemment appliqués. Bien qu'elle soit devenue une méthode « standard », la faible utilisation de l'analyse de survie dans les travaux sur la mortalité aux très grands âges est principalement due au manque de données convenables et suffisamment fiables aux âges les plus extrêmes. Nos données soigneusement validées permettent d'envisager l'application de cette méthode dont les bases théoriques ont été très solidement assurées.

Dans cette nouvelle approche, on se situe toujours dans le cadre des modèles de risques proportionnels, où le risque de décès de chaque individu est le produit entre la fonction de mortalité individuelle « standard » et la variable capturant l'hétérogénéité. Cependant, lorsque l'hétérogénéité est observable, il ne s'agit plus d'une seule variable aléatoire à distribution prédéterminée comme c'était le cas des modèles de fragilité, et l'hétérogénéité de la population peut dès lors être capturée par les variables observées sur chaque individu. L'objectif poursuivi est double. D'abord, il s'agit de vérifier l'existence d'un plateau de mortalité dans nos populations à partir de 105 ans (ce seuil d'âge est contraint par la disponibilité des données individuelles dans chaque population), puis de savoir si d'autres facteurs, tels que la génération à laquelle appartient l'individu, son sexe, ou la population dont l'individu est issu, peuvent avoir des impacts sur sa mortalité.

Pour étudier l'hypothèse d'un plateau de mortalité aux âges extrêmes, l'idée est de comparer les modèles qui se distinguent par leur fonction de mortalité de base (*baseline*), l'une présumant l'existence d'un plateau de mortalité et l'autre non, tout en incluant d'autres variables dont on dispose au niveau individuel telles que l'année de naissance de l'individu et son sexe. La conclusion se fait par un test d'hypothèse, adossé à une comparaison du niveau de performance des modèles par le critère AIC. Le modèle caractérisant le risque de décès individuel a pour forme complète la fonction suivante :

$$h(t_i) = h_0(t_i) \{ \exp(\beta_1 C_i + \beta_2 M_i) \} = a \exp(bt_i) \exp(\beta_1 C_i + \beta_2 M_i),$$

où  $C_i$  est l'année de naissance de l'individu centré en 1891,  $M_i = 0$  lorsque l'individu est masculin et 1 lorsque l'individu est féminin, et  $t_i$  est la durée de survie de chaque individu après l'âge de 105 ans.

Les analyses ont d'abord été faites sur chaque population, puis sur l'ensemble des trois populations réunies. Lorsque les données des trois populations sont combinées, le risque de décès individuel a pour fonction :

$$h(t_i) = h_0(t_i) \{ \exp(\beta_1 C_i + \beta_2 M_i + \beta_3 P_i) \} = a \exp(bt_i) \exp(\beta_1 C_i + \beta_2 M_i + \beta_3 P_i),$$

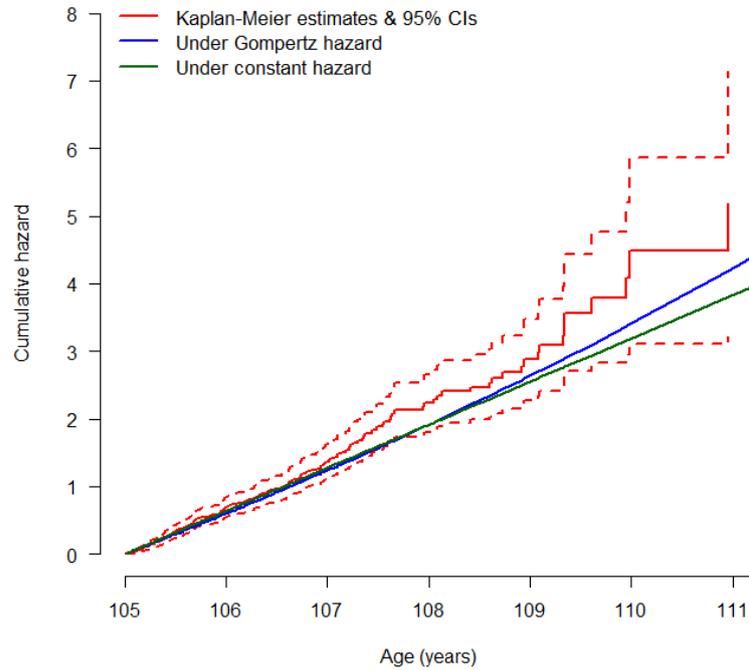
où  $P_i = 0$  si l'individu est issu de la population française, 1 si l'individu est belge et 2 si l'individu est canadien-français. Par conséquent,  $\beta_1$  capture l'effet de génération,  $\beta_2$  capture l'effet de sexe et  $\beta_3$  capture l'effet de population à laquelle l'individu appartient. La fonction de mortalité de base,  $h_0(t_i)$ , peut être interprétée comme le risque de décès des femmes françaises nées en 1891. Un test d'hypothèse pour étudier la validité du paramètre  $b$  de Gompertz dans la fonction de mortalité de base ( $H_0: b = 0$  vs.  $H_1: b \neq 0$ ), et les tests de significativité des paramètres capturant différents effets sur la mortalité peuvent être conduits avec peu de complexité.

Les tests d'hypothèses mentionnés sont effectués systématiquement pour chaque combinaison possible de la fonction de mortalité de base et des variables incluses dans le modèle. La valeur-p est rapportée pour chaque test et la valeur du critère AIC l'est pour chaque modèle estimé. En tout, on compare 8 modèles pour chaque population concernée et 16 modèles lorsque les données des populations françaises, belges et canadiennes-françaises sont réunies. Une inspection visuelle en comparant des courbes de risques cumulés estimées par une approche d'abord non paramétrique avec la méthode de Kaplan-Meier et ensuite paramétrique avec chacune des fonctions de mortalité de base, est aussi effectuée pour chaque jeu de données. Dans ce rapport, les résultats sont présentés en détails pour la population française et les résultats issus des autres populations et des données *poolées* le sont plus brièvement (un plus grand détail figurera dans la thèse).

Pour la population française, le test d'hypothèse rejette systématiquement l'hypothèse nulle selon laquelle le paramètre  $b$  de Gompertz est nul pour chaque combinaison de modèles : lorsque le test est mené entre les différents modèles sur la mortalité de base sans les variables de génération ni de sexe (valeur-p = 1.314e-07), lorsque seul l'effet de sexe est inclus (valeur-p = 8.202e-08), lorsque seul l'effet de génération est inclus (valeur-p = 1.322e-07), et lorsque les deux effets sont inclus (valeur-p = 8.18e-08). Ces tests d'hypothèse confirment la significativité statistique du paramètre  $b$  et la validité de la fonction de Gompertz comme fonction de mortalité de base, face à l'alternative d'une fonction de mortalité à valeur constante, suggérant ainsi que les risques de décès continuent d'augmenter, même au-delà de 105 ans. Il est aussi important de noter que notre estimation du paramètre  $b$  de Gompertz est plus élevée que celle donnée par Barbi et al. (0.062 vs. 0.013). Sur ce sujet, un travail de simulation par Camarda (2019) a aussi démontré que, lorsque la vraie trajectoire de mortalité est supposée gompertzienne, plus le paramètre  $b$  a une valeur élevée plus la probabilité de détecter la vraie trajectoire sera élevée.

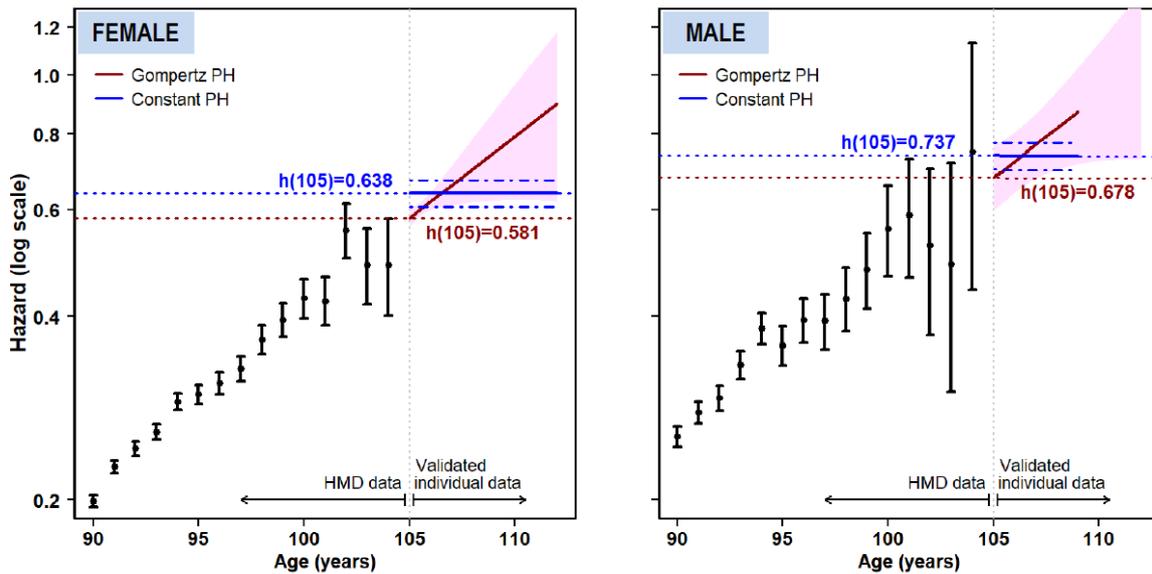
Une comparaison des courbes de risques de décès cumulés, l'une estimée par le modèle de base de mortalité gompertzienne, l'autre estimée par le modèle de base de mortalité constante, et la courbe de risques de décès cumulés générée par la méthode de Kaplan-Meier présentée dans la figure 7 suggère aussi un meilleur ajustement des données observées par le modèle de base gompertzienne.

**Figure 7.** Risques de décès cumulés au-delà de 105 ans pour la génération des femmes françaises nées en 1891 estimées par les approches non-paramétriques et paramétriques



Une croissance des risques au-delà de 105 ans s’inscrit aussi dans la continuité de la tendance de la mortalité aux âges plus jeunes d’après les données de HMD (figure 8).

**Figure 8.** Taux de mortalité observés et risques de décès ajustés des Françaises nées en 1891



De plus, un classement de performance de différents modèles de combinaison entre la fonction de mortalité de base et différentes variables peut être établi grâce à la comparaison des valeurs du critère AIC calculé pour chaque modèle (tableau 4). On retient là aussi au premier rang de performance le modèle de mortalité de base gompertzienne incluant la variable capturant l'effet de sexe. Un test d'hypothèse confirme aussi la significativité du paramètre  $\beta_2$ , suggérant en plus une surmortalité masculine prononcée qui persiste même jusqu'aux âges les plus élevés.

**Tableau 4.** Estimations des paramètres des différents modèles de combinaison de variables pour les individus français nés entre 1883 et 1901, ayant survécu au-delà de 105 ans et ensuite suivis jusqu'à leur décès.

Paramètres	Estimations	95% IC	Log-vraisemblance	AIC	Rang
<i>Modèle de mortalité de base constante, avec des effets de génération et de sexe</i>					
$a$	0.638	[0.614; 0.664]	5449.384	10904.77	6
$\beta_1$	-0.00017	[-0.00635; 0.006015]			
$\beta_2$	0.144	[0.026; 0.261]			
<i>Modèle de mortalité de base Gompertz, avec des effets de génération et de sexe</i>					
$a$	0.580	[0.550; 0.612]	5435.002	10878.00	2
$b$	0.062	[0.040; 0.084]			
$\beta_1$	0.00028	[-0.00590; 0.006460]			
$\beta_2$	0.156	[0.038; 0.273]			
<i>Modèle de mortalité de base constante, sans effet</i>					
$a$	0.645	[0.625; 0.666]	5452.158	10906.32	7
<i>Modèle de mortalité de base Gompertz, sans effet</i>					
$a$	0.589	[0.562; 0.617]	5438.235	10880.47	3
$b$	0.061	[0.039; 0.083]			
<i>Modèle de mortalité de base constante, avec l'effet de sexe</i>					
$a$	0.638	[0.617; 0.659]	5449.385	10902.77	5
$\beta_2$	0.144	[0.027; 0.261]			
<i>Modèle de mortalité de base Gompertz, avec l'effet de sexe</i>					
$a$	0.581	[0.554; 0.609]	5435.006	10876.01	1
$b$	0.062	[0.040; 0.084]			
$\beta_2$	0.155	[0.038; 0.273]			
<i>Modèle de mortalité de base constante, avec l'effet de génération</i>					
$a$	0.645	[0.621; 0.670]	5452.152	10908.3	8
$\beta_1$	-0.00035	[-0.00654; 0.00583]			
<i>Modèle de mortalité de base Gompertz, avec l'effet de génération</i>					
$a$	0.588	[0.559; 0.619]	5438.235	10882.47	4
$b$	0.061	[0.039; 0.084]			
$\beta_1$	0.00008	[-0.00610; 0.00626]			

Notes : IC est l'intervalle de confiance et AIC est Akaike Information Criterion.

Les mêmes analyses sont mises en place systématiquement pour la population belge, la population canadienne-française, et pour l'ensemble des trois populations. Tenant compte du nombre d'observations important de la population française, les mêmes analyses sont aussi menées sur les données belge et québécoise regroupées pour vérifier la robustesse des résultats.

Lorsque l'on considère séparément la population de la Belgique ou du Québec, le paramètre  $b$  de Gompertz est non significatif statistiquement et le test d'hypothèse signale qu'on ne peut pas rejeter l'hypothèse nulle. De plus, aucun effet significatif n'est retrouvé dans chacune des deux populations. Lorsque les données belges et québécoises sont regroupées, on retrouve de nouveau la significativité du paramètre  $b$  de Gompertz, avec un désavantage masculin, et le meilleur modèle suggéré par les critères AIC est celui de mortalité de base gompertzienne avec un effet sexe et population. Ces différences en matière de résultats obtenus peuvent s'expliquer principalement par le faible nombre d'observations dans chacune des populations belges et québécoises prises séparément, qui ne permet pas de faire ressortir le phénomène sous-jacent. Ceci change dès que les données de ces deux populations sont combinées, avec des régressions toujours effectuées au niveau individuel. Lorsque les données des trois populations sont regroupées, la significativité du paramètre  $b$  de Gompertz est justifiée systématiquement par le test d'hypothèse mené pour chaque combinaison de variables, suggérant que les risques de décès continuent à croître après l'âge de 105 ans. Aucun effet significatif n'est détecté pour les variables de génération et de population : la génération et la population auxquelles appartient l'individu n'ont pas d'effet sur ses risques de décès aux très grands âges de la vie. Par contre, un désavantage masculin significatif est toujours retrouvé, un homme a 1,188 fois plus de risques qu'une femme de décéder à un âge donné. Le classement de performance par comparaison des critères AIC suggère que le modèle le plus plausible est celui de base de mortalité gompertzienne, incluant la variable capturant l'effet de sexe.

Ces résultats diffèrent de ceux obtenus par Barbi et collègues (2018), sans forcément les remettre en cause, et témoignent possiblement d'une réelle variabilité des cas de figure que l'on peut observer selon les jeux de données provenant de populations diverses avec des caractéristiques distinctes. Les résultats obtenus dans cette partie montrent à quel point une réponse universelle à la question de l'existence d'un plateau de mortalité humaine est (encore) loin d'être possible. Il y a par contre à présent des fondements empiriques qui suggèrent une poursuite de la croissance des risques de décès au-delà de 105 ans. La mortalité aux âges extrêmes de la vie ne dépend guère de la génération à laquelle les survivants aux grands âges appartiennent, ni de leur pays. Une seule différence persiste, celle entre homme et femme, qui caractérise l'évolution de la mortalité à ces âges.

#### **IV. Valorisation des résultats de recherche**

Au cours des trois années financées par le projet Scor, les résultats de la recherche ont été largement valorisés avec plusieurs communications dans les conférences scientifiques internationales : deux posters présentés à la réunion annuelle de la *Population Association of America* (PAA) du 26 au 28 avril 2018 à Denver (États-Unis) et du 10 au 13 avril 2019 à Houston

(États-Unis), ainsi que de nombreuses présentations orales délivrées respectivement au Congrès de l'Union internationale pour l'étude scientifique des problèmes de population (UIESP) du 29 octobre au 3 novembre 2017 au Cap (Afrique du Sud), à la conférence de la *European Association for Population Studies* (EAPS) du 6 au 9 juin 2018 à Bruxelles (Belgique), à la conférence de la *Asian Population Association* (APA) du 11 au 14 juillet 2018 à Shanghai (Chine), à la conférence de l'Association Internationale des Démographes de Langue Française (AIDELF) du 28 au 31 août 2018 à Louvain-la-Neuve (Belgique) et au colloque de l'Association des démographes du Québec le 6-7 mai 2019 à Montréal (Canada). Une autre communication orale a été acceptée pour la séance « *Puzzles and Paradoxes in Mortality* » pour la conférence PAA 2020 mais à la suite de la crise sanitaire causée par la Covid-19, la séance a dû être annulée.

En outre, à l'invitation de Razvan Ionescu, Linh Dang a eu aussi l'opportunité de présenter certains résultats de sa recherche lors d'une réunion interne SCOR en juillet 2018.

Deux articles de recherche ont été rédigés et soumis pour publication dans des revues savantes en parallèle de la rédaction de la thèse. Le premier, soumis à la Revue Quetelet, est accepté pour publication sous quelques conditions de révision. La version révisée sera renvoyée aux éditeurs en février 2021. Le second, soumis à *Demographic Research* est en cours d'évaluation. Deux autres articles de recherche sont en cours de rédaction.

Le financement de la Fondation SCOR a également permis à Linh Dang de suivre une formation portant sur « *Life Table and Decomposition Analysis for Population Health Research* » assurée par Alyson van Raalte et Marília Nepomuceno en juin 2018 au Max Planck Institute for Demographic Research à Rostock (Allemagne), dont le contenu a apporté de nouveaux éléments à la formation de la doctorante et à ses réflexions de thèse.

En 2018, Carlo Giovanni Camarda, chercheur à l'INED, a rejoint le comité de suivi de thèse de la doctorante en tant que tuteur INED et troisième encadrant. L'expertise bien connue de ce chercheur en statistique et en programmation R a permis des avancées plus rapides et plus profondes dans les travaux d'ajustement des modèles de mortalité de Linh Dang. Elle a non seulement renforcé les travaux déjà faits mais elle s'est aussi engagée dans d'autres directions de recherche dont le niveau technique s'avère plus avancé, le tout sous la supervision rigoureuse de France Meslé et Nadine Ouellette.

Cependant, quel que soit le soutien solide que Linh Dang a reçu au cours de ses années de thèse et la diligence avec laquelle ses travaux de thèse ont été menés, la rédaction d'une thèse dans un format traditionnel de monographie en français, langue étrangère de la doctorante, combinant rigueur technique et écriture fluide, n'a pas été une tâche facile pour Linh Dang. D'autre part, les complexités sociales et sanitaires que traversent la France et le monde entier depuis les derniers mois de 2019 sont venues ajouter un poids non négligeable à ses charges tant professionnelles que personnelles. Ceci a rendu le calendrier initial de la thèse impossible à tenir. La thèse a donc dû être prolongée jusqu'à sa cinquième année. Après l'épuisement de la bourse accordée par la Fondation SCOR, pour financer ces deux dernières années de thèse, Linh Dang a reçu les soutiens financiers de la part de l'INED, institution à laquelle elle est affiliée, ainsi que plusieurs bourses de l'Université de Montréal grâce à l'appui de Nadine Ouellette. L'ensemble de ces financements a finalement permis de mener ce travail de thèse à son terme et la soutenance est prévue au plus tard pour septembre 2021.

## V/ Perspectives à l'issue du projet de thèse

### 1. Perspectives de recherche

La recherche sur la trajectoire de la mortalité aux très grands âges est un terrain de recherche qui reste encore largement à défricher mais elle ouvre aussi sur des questions fondamentales comme les conséquences sociales et économiques du vieillissement de la population qui est d'ores et déjà un phénomène démographique majeur dans de nombreux pays.

Le projet de thèse présenté initialement à la fondation SCOR proposait deux types de modélisation de la trajectoire de la mortalité aux très grands âges : une approche paramétrique et une approche non-paramétrique. Finalement, compte tenu des contraintes de temps, l'attention a été essentiellement portée sur l'approche paramétrique pour mieux mettre à jour les potentiels générés par des données individuelles d'excellente qualité, pour pouvoir ré-évaluer la validité des modèles de mortalités « classiques », dont l'usage est déjà très répandu dans les études s'intéressant aux âges extrêmes de la vie. Tout le champ de l'approche non-paramétrique reste donc à explorer et mérite bien évidemment de l'attention.

Quant au volet de la modélisation paramétrique lui-même, plusieurs stratégies de modélisation n'ont pas encore été abordées, telles que les développements récents à partir des modèles de fragilité classiques comme le « *extended Gamma-Gompertz* » (Salinari et De Santis, 2019), permettant d'assouplir certaines hypothèses jusqu'à présent très contraignantes mais nécessaires du fait de la relative pauvreté des données et de l'absence d'outils de calculs suffisamment puissants. En fait, les méthodes d'analyse utilisées dans ce projet de thèse sont issues d'un univers d'analyse statistique foisonnant dont cette thèse n'a considéré que quelques cadres parmi les plus simples pour pouvoir exploiter la richesse des données disponibles pour nos trois populations dans une limite de temps donnée. Le cours « *Life Table and Decomposition Analysis for Population Health Research* », que Linh Dang a eu l'opportunité de suivre en juillet 2018 au MPIDR, lui a fourni les bases mathématiques pour comprendre comment les changements de composition de la population pourraient agir sur la force de mortalité (moyenne) de la population, non pas sous une logique des modèles à risques proportionnels, mais du point de vue de divers exercices de décomposition (Vaupel et Missov, 2014 ; Vaupel et Zhang, 2010 ; Vaupel et Canudas-Romo, 2002). Dans l'avenir, il serait ainsi intéressant de chercher à savoir si l'hétérogénéité non observable pourrait être quantifiée d'une autre manière que par les modèles de fragilité déjà exploités dans la littérature.

Outre les essais méthodologiques, un des débouchés initialement prévus de ce travail, la construction de tables de mortalité par génération pour la France, n'a pu être complètement mené à son terme. Les résultats obtenus vont toutefois offrir une base plus solide à la formulation d'hypothèses sur l'évolution de la mortalité aux grands âges dans les prochaines décennies.

Au fur et à mesure que les données sur les décès aux très grands âges deviennent de plus en plus riches et portent sur des âges de plus en plus élevés, il sera intéressant de répéter régulièrement

l'application des différents ensembles de modèles considérés pour suivre les possibles changements de la trajectoire et réduire quelques incertitudes encore persistantes à ce stade.

## **2. Perspectives de la doctorante**

Après une formation en sciences économiques, Linh Dang a pu développer ses compétences en démographie au fil des travaux menés dans le cadre de sa thèse, notamment en matière de recherche sur la mortalité. Elle a également acquis des savoir-faire en statistiques et en programmation informatique. Avant même le commencement de cette thèse, elle portait un grand intérêt aux recherches sur les différents aspects de la santé, notamment à travers des projets auxquels elle avait eu de l'occasion de participer quand elle était étudiante en master dans le cadre de la chaire d'Hospinnomics, avec une attention particulière sur la population des personnes âgées. Cet intérêt a été renforcé par ces 5 années de travail intensif sur la mortalité d'une population très spéciale, celle des survivants au-delà de 90 ans, des centenaires et des supercentenaires. Ces derniers constituent une population considérée comme chanceuse par certains, exceptionnelle par d'autres, mais, dans tous les cas, assez vulnérable et nécessitant une attention particulière de la part des pouvoirs publics et des acteurs privés pour un bon encadrement de leur qualité de vie. Linh Dang a ainsi renforcé son envie de poursuivre ses recherches dans le domaine de la santé, de la mortalité et du bien-être des personnes âgées. Ses compétences en démographie, économie et statistiques lui donnent une vision interdisciplinaire sur le sujet ainsi qu'une capacité à traiter des questions du vieillissement sous plusieurs aspects simultanément. Concrètement, Linh Dang souhaiterait travailler, entre autres, sur la prise en charge des personnes âgées, vivant seules ou avec leurs proches, admis dans les milieux hospitaliers, et reçus dans les établissements privés ou publics destinés aux personnes âgées dépendantes (les EHPAD, les maisons de la retraite, etc.). Ses recherches pourraient notamment porter sur les questions d'analyse de coût-bénéfice ou sur des études aidant à la prise de décisions lors des périodes de crise, où les décisions à forts enjeux doivent être conclues dans l'urgence. La pandémie de Covid-19 en a été une vive démonstration, mais d'autres évènements récurrents comme les canicules ou autres évènements climatiques, ont aussi des effets non négligeables sur la santé des personnes âgées. Enfin, la pauvreté, l'isolement et le capital social des personnes âgées sont aussi des terrains importants de recherche auxquels Linh Dang pourrait contribuer. Elle est ouverte à la mobilité, apte à travailler avec les équipes internationales, et prête à poursuivre une carrière de chercheuse que ce soit en milieu public ou privé.

## ANNEXE 1 – Loïs de mortalité utilisées pour l’ajustement des modèles statistiques paramétriques

- (1) Modèle de Gompertz:  $\mu_x = ae^{bx}$       (2) Modèle de Weibull:  $\mu_x = ax^b$
- (3) Modèle de Kannisto:  $\mu_x = \frac{ae^{bx}}{1+ae^{bx}}$       (4) Modèle de Beard:  $\mu_x = \frac{ae^{bx}}{1+\delta e^{bx}}$
- (5) Modèle quadratique:  $\mu_x = e^{ax^2+bx+c}$

## ANNEXE 2 – Fonctions mathématiques des modèles de fragilité employés

- (1) Modèle de gamma-Gompertz :  $\mu_x = \frac{\exp(ax^2+bx+c)}{1+\gamma \int_{x_0}^x \exp(at^2+bt+c)dt}$
- (2) Modèle de gamma-log-quadratique :  $\mu_x = \frac{\exp(abx)}{1+\frac{a\gamma}{b}(\exp(bx)-1)}$
- (3) Modèle de gamma-Kannisto :  $\mu_x = \frac{\frac{a\exp(bx)}{1+a\exp(bx)}}{1+\frac{\gamma}{b}\log(a\exp(bx)+1)}$

## ANNEXE 3 – Bases gérontologiques de la fonction log-quadratique pour décrire la force de mortalité individuelle

La fonction log-quadratique pour décrire la force de mortalité individuelle que propose Horiuchi (2003) se base sur la notion de vitalité et s’avère cohérente avec le mécanisme de prévention et de réparation des dommages subis. En suivant le modèle de Strehler et Mildvan (1960), où la vitalité d’un organisme est présumée à détériorer linéairement avec l’âge, celle-ci s’écrit sous la forme suivante :

$$v(x) = v_0 - gx,$$

où  $v_0$  est le niveau initial de la vitalité et  $g$  est le taux de détérioration physiologique.

Or lorsque l’organisme possède une capacité de se réparer des dommages subis, sa fonction de vitalité devient :

$$v(x) = v_0 - (g - r)x,$$

où  $r$  est le taux de réparation de l’organisme et  $0 < r < g$ .

Ce taux de réparation quant à lui pourrait se détériorer aussi avec l’âge. Sous l’hypothèse que la baisse en capacité de se réparer de l’organisme est linéaire,  $r(x) = r_0 - ax$ , où  $r_0$  est le niveau de la capacité initiale à se réparer et  $a$  est le taux de diminution de cette capacité. La fonction de vitalité de l’organisme devient ainsi :

$$v(x) = v_0 - (g - (r_0 - ax)x).$$

En utilisant la relation vitalité-mortalité que propose le modèle de Strehler et Mildvan, la force de mortalité individuelle a comme fonction mathématique :

$$\mu(x) = \exp(ax^2 + bx + c),$$

où  $b = g - r_0 > 0$  et  $c = -v_0 < 0$ .

Lorsque l'âge initial du processus est tel que  $x_0 \neq 0$ , alors :

$$b = g - r_0 - 2ax_0$$

$$c = -v_0 - gx_0 + r_0x_0 + ax_0^2.$$

La forme de la trajectoire de mortalité individuelle a été déjà étudiée dans la littérature. Entre d'autres, Yashin et Iachine (1997) ont estimé une trajectoire des risques de décès de jumeaux danois par le modèle de fragilité semi-paramétrique et ils arrivent à une trajectoire visiblement log-convexe. En 2002, Doblhammer et Oeppen ont publié des études sur la trajectoire de mortalité individuelle au-delà de 50 ans des nobles anglais et ils arrivent aussi à cette même conclusion. Il est de fait utile de recourir à une fonction mathématique suffisamment souple, capable de reproduire les formes log-convexes ou log-concaves plausibles en matière de mortalité individuelle.

## RÉFÉRENCES

- Barbi, E., Lagona, F., Marsili, M., Vaupel, J.W., Wachter, K.W., (2018), The plateau of human longevity: Demography of longevity pioneers, *Science*, vol. 360, n°6396.
- Beard, R.E., (1959), Note on some mathematical mortality models, in Wolstenholme and O'Connor (eds.), *The lifespan of animals*, Ciba Foundation Colloquium on Aging, Little, Brown, Boston, pp.802-811.
- 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. Département de démographie. Université de Montréal.
- Bourbeau, R., Desjardins, B., (2002), Dealing with problems in data quality for the measurement of mortality at advanced ages in Canada. *North American Actuarial Journal*, vol.6, n°3, p. 1-13.
- Böhnstedt, M., Gampe, J., (2019a), Detecting mortality deceleration: Likelihood inference and model selection in the gamma-Gompertz model, *Statistics and Probability Letters*, vol.150, pp.68-73.
- Böhnstedt, M., Putter, H., Ouellette, N., Claeskens, G., Gampe, J., (2019b), Shifting attention to old age: Detecting mortality deceleration using focused model selection, <https://arxiv.org/abs/1905.05760>.
- Camarda, G.C, Dang, L.H.K., Meslé, F., Robine, J-M., Vallin, J., (2018), 2018, RE: Premature claim of a plateau of human mortality: the role of sample size, *Science eLetters*, [science.sciencemag.org/content/360/6396/1459/tab-e-letters](http://science.sciencemag.org/content/360/6396/1459/tab-e-letters).
- Camarda, G.C, (2019), The curse of the plateau. A skeptical view on mortality estimation at extreme ages. Presented at 14<sup>th</sup> Supercentenarians seminar. 28-29 November. Paris.
- Dang, L.H.K, Meslé, F., Ouellette, N., (2018a), Risks of death at oldest ages: A birth cohort analysis in France, Belgium and Québec (Canada), poster presentation at PAA annual meeting.
- Dang, L.H.K, Meslé, F., Ouellette, N., (2018b), Risks of death at oldest ages: What do we learn from experiences of French, Belgian & Quebec centenarians?, oral presentations at EAPS conference (Belgium, June 6th-9th) and APA meeting (Shanghai, July 11th-14th).
- Dang, L.H.K., Camarda, G.C, Meslé, F., Ouellette, N., Robine, J-M., Vallin, J., (2019), Survival analysis on mortality data at oldest ages: First results on longevity pioneers in France, poster presentation at PAA annual meeting.
- Desjardins, B, Bourbeau, R., (2010), The emergence of supercentenarians in Canada, in Maier H., Gampe J., et al. (eds), *Supercentenarians*, Berlin Heidelberg, Springer (coll. Demographic Research Monographs), p. 59-74.
- Doblhammer, G. et Oeppen, J., (2002), Reproduction and longevity among the British peerage: The effect of frailty and health selection, presentation at PAA annual meeting.
- Gampe, J., (2010). Human mortality beyond age 110, in Maier H., Gampe J., et al. (eds), *Supercentenarians*, Berlin Heidelberg, Springer (coll. Demographic Research Monographs), pp.231-245.
- Gavrilov, L.A. & Gavrilova, N. S., (2011), Mortality measurement at advanced ages: A study of the Social Security Administration Death Master File, *North American Actuarial Journal*, vol.15, n°3, pp. 432 – 447.
- Gavrilova, N.S., Gavrilov, L.A., (2015), Biodemography of old-age mortality in humans and rodents, *Journals of Gerontology: Biological Sciences*, vol.70, n°1, p. -9.
- Gavrilov, L.A., Gavrilova, N.S., (2019), New trend of old-age: Gompertzialization of mortality trajectory, *Gerontology*, Online publication DOI: 10.1159/000500141.
- Gompertz, B., (1825), On the nature of the function expressive of the law of the human mortality and on a new mode of determining the value of life contingencies, *Philosophical Transactions of the Royal Statistical Society*, vol.27, n°2, pp.513-583.
- Horiuchi, S., Coale, A., (1990), Age patterns of mortality for older women: An analysis using the age-specific rate of mortality change with age, *Mathematical Population Studies*, vol.2, n°4, pp.245-267.
- Horiuchi, S., Wilmoth, J.R., (1998), Deceleration in the age pattern of mortality at older ages, *Demography*, vol. 35, n°4, pp.391-412.
- Horiuchi, S., (2003), Interspecies comparison of life span distribution: Humans versus invertebrates, dans Carey, J., et Tuljapurkar, S., *Life Span: Evolutionary, Ecological and Demographic Perspectives*, *Population and Development Review*, Special Supplement to volume 29, pp. 127-151.
- Kannisto, V., (1994), *Development of oldest old mortality, 1950 – 1990: Evidence from 28 developed countries*. Odense. Odense University Press.

- Maier, H., Gampe, J., Jeune, B., Robine, J-M., Vaupel, J. (eds), (2010), *Supercentenarians*. Berlin, Springer-Verlag.
- Maier, H., Jeune, B., Vaupel, J. (eds), (2020), *Exceptional Lifespans*. Berlin, Springer-Verlag.
- Meslé, F., & Vallin, J., (2002). Comment améliorer la précision des tables de mortalité aux grands âges? *Population*, n°4, pp. 603–631.
- Meslé, F., Vallin, J., Robine, J-M., Desplanques, G., Cournil A., (2010). Is it possible to measure life expectancy at 110 in France?, dans: Maier H., Gampe J., et al. (eds), *Supercentenarians*, Berlin Heidelberg, Springer (coll. Demographic Research Monographs), pp.231-245.
- Ouellette, N., Bourbeau, R., (2014), Measurement of mortality among centenarians in Canada, *2014 Living to 100 Monographs* (Living to 100 Symposium, Orlando FL, Jan. 8 -10), Society of Actuaries.
- Ouellette, N., Poulain, M., Bourbeau, R., (2016a), Mortality trajectories at older ages, French – Canadian & Belgian populations, 12<sup>th</sup> Supercentenarian Workshop, Estonian Institute for Population Studies, Tallinn University, June 2nd-3rd 2016.
- Ouellette, N., (2016b), La forme de la courbe de mortalité des centenaires canadiens-français, *Gérontologie et Société*, vol.38, n°151, pp.41-53.
- Ouellette, Nadine, Meslé, France, Vallin, Jacques & Robine Jean-Marie, (2020) , Supercentenarians and semi-supercentenarians in France, dans Maier Heiner, Jeune Bernard, Robine Jean-Marie and Vaupel James, *Exceptional Life Spans*. Berlin, Heidelberg, Springer.
- Perks, W., (1932), On some experiments on the graduation of mortality statistics, *Journal of the Institute of Actuaries*, vol.63, pp.12-40.
- Poulain, M., (2011), On the age validation of supercentenarians, in Maier H., Gampe J., et al. (eds), *Supercentenarians*, Berlin Heidelberg, Springer (coll. Demographic Research Monographs).
- Salinari, G., De Santis, G., (2019), One or more rates of aging? The extended gamma-Gompertz model, *Statistical Methods & Application*.
- Vaupel, J.W., Manton, K.G., Stallard, E., (1979), The impact of heterogeneity in individual frailty on the dynamics of mortality, *Demography*, vol.16, pp.439-454.
- Vaupel, J.W., Canudas-Romo, V., (2002), Decomposing demographic change into direct vs. compositional components, *Demographic Research*, vol.7, pp.1-14.
- Vaupel, J.W., Zhang, Z., (2010), Attrition in heterogeneous cohorts, *Demographic Research*, vol.23, pp.737-748.
- Vaupel, J.W., Missov, T.I., (2014), Unobserved population heterogeneity: A review of formal relationships, *Demographic Research*, vol.31, pp.659-686.
- Weibull, W.A., (1951), A statistical distribution function of wide applicability, *Journal of Applied Mechanics*, n°18, pp.293-297.
- Yashin, A.I. et Iachine, I.A., (1997), How frailty models can be used for evaluating longevity limits: Taking advantage of an interdisciplinary approach, *Demography*, vol.34, n°1, pp.31-48.