

**Mémoire présenté le :
pour l'obtention du diplôme
de Statisticien Mention Actuariat
et l'admission à l'Institut des Actuares**

Par : Monsieur Baye Salif DIACK

Titre du mémoire : Construction de lois d'expérience en assurance dépendance

Confidentialité : NON OUI (Durée : 1 an 2 ans)

Les signataires s'engagent à respecter la confidentialité indiquée ci-dessus.

Membres présents du jury de la
filière :

Signature :

Entreprise :

CNP Assurances

Nom :

Signature :



Directeur de mémoire en
entreprise

Membres présents du jury de
l'Institut des Actuares :

Signature :

Nom :

BENTAHAR OUMMA

Signature :



Invité :

Nom :

Signature :

**Autorisation de publication et de mise
en ligne sur un site de diffusion de
documents actuariels (après expiration
de l'éventuel délai de confidentialité)**

Signature du responsable

entreprise :

BENTAHAR OUMMA

Signature du candidat :





Institut de Statistique de l'Université de Paris
Sorbonne Université

Mémoire d'Actuariat :
CONSTRUCTION DE LOIS D'EXPÉRIENCE
EN ASSURANCE DÉPENDANCE

Par Baye Salif DIACK

Responsable entreprise : Oumnia BENTAHAR
Responsable pédagogique : Guillaume BIESSY

CNP ASSURANCES



Juin 2022

Table des figures

1.1	Fiche récapitulative évaluation AGGIR	6
1.2	Évolution de l'espérance de vie en France	7
1.3	Évolution comparée EV et EVSI	8
2.1	Répartition des assurés par année d'adhésion	16
2.2	Répartition des adhésions hors 2002	17
2.3	Répartition genre	17
2.4	Pyramide des âges	18
2.5	Pyramide des âges à l'entrée en dépendance	19
2.6	Passages en dépendance	19
2.7	Décès en dépendance	20
2.8	État des assurés au 03/12/2018	20
2.9	Décès connus par survenance	21
3.1	Modélisation de la dépendance	22
3.2	Durée de maintien en dépendance	26
3.3	Durée de maintien en dépendance des sinistres clos	26
3.4	Loi de maintien en dépendance par estimateur actuariel	28
3.5	Fonction de survie du maintien en dépendance par l'estimateur Kaplan-Meier	29
3.6	Comparaison estimateurs actuariel et Kaplan-Meier	30
3.7	Impact de la franchise sur le maintien en dépendance	30
3.8	Comparaison lois de maintien en dépendance par sexe	31
3.9	Lois de maintien en dépendance par classe d'âge	36
3.10	Lois de maintien en dépendance par sexe et classe d'âges	38
3.11	Taux d'incidence bruts	41
3.12	Taux d'incidence bruts en échelle logarithmique	42
3.13	Comparaison taux d'incidence bruts par sexe	43
3.14	Comparaison des lissages sur les taux d'incidence Hommes	54
3.15	Comparaison des lissages sur les taux d'incidence Femmes	55
3.16	Prolongement lois d'incidence Hommes	60
3.17	Prolongement lois d'incidence Femmes	61
3.18	Comparaison lois d'incidence avec des lois de tarification internes	61
3.19	Taux bruts de mortalité des valides par âge	64
3.20	Comparaison des lissages sur les taux de mortalité des valides	65
3.21	Comparaison loi de mortalité des valides et table THF00-02	67
3.22	Prolongement loi de mortalité des valides	68
4.1	Impat de la loi de maintien sur la PSC	70

4.2	Impact des lois d'expérience sur la PRC	72
-----	---	----

Note de synthèse

Contexte

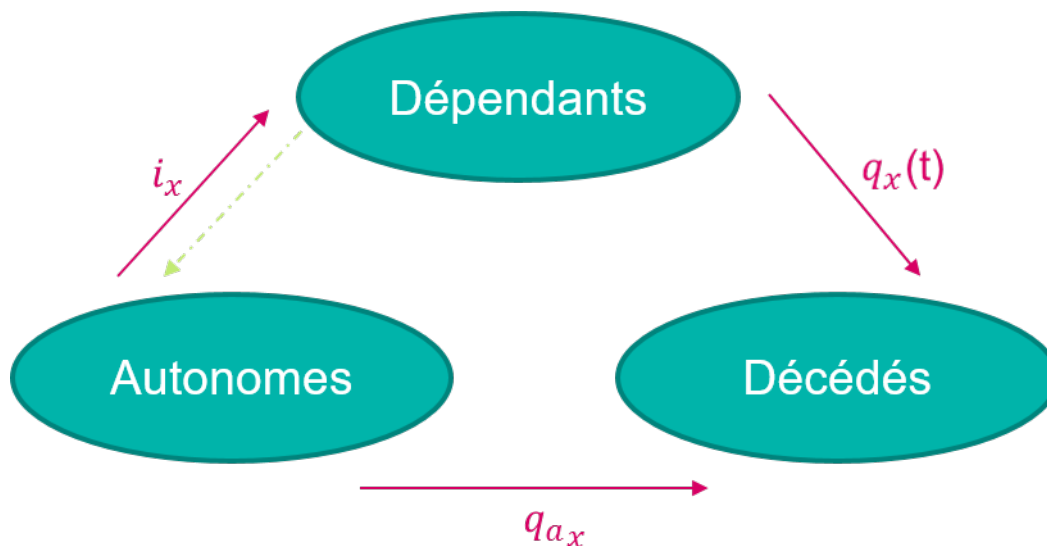
La prise en charge de la perte d'autonomie des personnes âgées ou dépendance est devenue un enjeu sociétal. Les progrès de la médecine ont permis des gains en espérance de vie non négligeable. Selon l'INSEE, entre 1950 et 2019, les femmes ont gagné 3,7 années d'espérance de vie et les hommes 6 années. Cependant, les années gagnées ne sont pas forcément des années de bonne santé. En effet alors que l'espérance de vie à 65 ans s'élevait à 19,4 ans pour les hommes et 23,2 ans pour les femmes en 2018, l'espérance de vie sans incapacité à 65 ans n'était que de 10,1 ans pour les hommes et 11,2 ans pour les femmes.

Face à cette situation préoccupante et une prise en charge des pouvoirs publics insuffisante, des produits d'assurance dépendance sont commercialisés depuis plus d'une trentaine d'années par les différents organismes (compagnies, mutuelles et institutions de prévoyance). D'où la multitude de définitions de la dépendance basées sur les Actes de la Vie Quotidienne (AVQ), la grille Autonomie Gérontologie Groupe Iso-Ressource (AGGIR) ou la grille Colvez entres autres. La modélisation du risque devient alors une nécessité pour chaque acteur en fonction de la nature de son portefeuille et du contrat proposé.

Modélisation

Nos travaux portent sur un contrat d'assurance de groupe à adhésion obligatoire couvrant la dépendance totale uniquement. La garantie dépendance est adossée à d'autres garanties prévoyance avec une définition basée sur des AVQ. Le portefeuille est en quasi-fermeture au 31/12/2008. Afin de nous assurer de la bonne qualité et de la pertinence des données utilisées un traitement a été effectué au préalable sur les données brutes en faisant des tests de cohérence et en nous basant sur notre connaissance du risque assuré et du contrat qui régit la garantie. Les statistiques descriptives ont permis de souligner les limites des données notamment l'historique relativement récent des sinistres. Nous avons néanmoins construits nos lois d'expérience sur la base des données retenues et avons utilisés différentes méthodes pour les prolonger.

Notre approche utilise un modèle multi-états : les cotisants sont susceptibles de rentrer en dépendance ou de mourir directement sans passer par cet état ; les dépendants quant à eux ne sont exposés qu'à la mortalité, la dépendance totale étant irréversible.



Dans un premier temps, une loi de maintien en dépendance a été construite avec l'estimateur de Kaplan Meier avant d'estimer une loi de maintien pour les hommes et une autre pour les femmes avec le même estimateur. Le modèle de Cox a permis de juger de la pertinence de construire des lois par tranche d'âges et par sexe confirmée par les tests de rapport de maxima de vraisemblance de Wald et du logrank. Le paramètre du modèle de Cox ne pouvant être supposé constant avec le temps, hypothèse confirmée par le test des résidus de Schoenfeld, le modèle de Cox n'a pas été utilisé pour déterminer les différentes lois qui ont été estimées par Kaplan Meier séparément. En raison du manque de données à partir de la cinquième année de dépendance, les lois de maintien ont été retenues sur les quatre premières années puis prolongées en appliquant des taux de majorations aux tables réglementaires. Les coefficients utilisés diminuent avec l'âge en se rapprochant de 1, ce qui revient à dire que la survie des dépendants se rapproche de celle de la population générale.

Ensuite, nous avons estimé des lois d'incidence brutes pour l'ensemble de la population puis par sexe et avons pu confirmer la pertinence de modéliser séparément les lois d'incidence hommes et femmes en utilisant le test de Mantel-Haenszel (test du log-rank). Nous avons ensuite appliqué et comparé les méthodes de lissage non paramétriques des Moyennes Mobiles et de Whittaker-Henderson, puis les lissages paramétriques de Gompertz-Makeham et Thatcher sur nos deux lois d'incidence brutes. Les critères de fidélité et de régularité ont permis de faire des choix qui ont été réconfortés par des backtestings. Le lissage obtenu avec la méthode de Whittaker-Henderson a été retenu pour la loi d'incidence des hommes et celui de la méthode des Moyennes Mobiles pour la loi d'incidence des femmes. Ces lois ont été construites pour des âges compris entre 60 et 85 ans pour les hommes, et pour les femmes entre 60 et 89 ans car au-delà le manque de données rendrait l'estimation non fiable. L'objectif étant d'avoir les lois sur les âges [16;110], les lois ont été extrapolées sur les âges avant 60 ans dans un premier temps et ensuite sur les âges au-delà de 85 et 89 ans respectivement pour les hommes et les femmes. Sur les âges inférieurs les deux méthodes testées donnent des résultats assez similaires et nous avons retenues le prolongement obtenu par la méthode de Gompertz-Makeham. À l'inverse sur les âges élevés, le prolongement obtenu par la méthode de Gompertz-Makeham traduit une croissance exponentielle qui ne nous a pas paru adapté à l'incidence en dépendance, notre choix a porté sur le prolongement de la méthode de Thatcher plus adapté à notre connaissance du risque de perte d'autonomie.

Puis, nous avons construit de façon similaire aux lois d'incidence, deux lois de mortalités des assurés autonomes. Les lois bruts ont été lissées avec les mêmes méthodes que les lois d'incidence. Les lissages de Whittaker-Henderson ont été retenues à la fois pour les hommes et les femmes. Ces lois construites sur une plage d'âges compris entre 60 et 90 ans ont été prolongées en se basant sur l'écart de mortalité observé par rapport aux tables réglementaires TH00-02 et TF00-02 entre 80 et 90 ans.

Impact sur les provisions

Enfin, nous avons mesurer l'impact des lois établies sur la Provision pour Sinistres en Cours (PSC, aussi appelé PM de rentes par abus de langage) et la Provision pour Risques en Cours (PRC). En effet les lois construites peuvent permettre de projeter le portefeuille mais aussi de calculer des provisions sur la base de l'expérience qu'on a acquiert, avec donc plus de précisions que les lois de tarification. Sur la PSC qui n'est impactée que par la loi de maintien en dépendance, la provision calculée présente un écart négligeable de 1% à la baisse par rapport à la provision estimée avec les derniers lois du Réassureur. Cependant, un écart de 10% à la hausse par rapport à la PSC enregistrée dans les comptes est notée. Quant à la PRC, elle est impactée par les trois lois construites avec une baisse de 6,86% par rapport à la PRC calculée par le Réassureur avec ses dernières lois. L'impact sur la PRC enregistrée dans les comptes est une surestimation de 15,37%. N'ayant pas les lois utilisés par le Réassureur nous ne pouvons pas quantifier l'impact isolé de chacune de nos lois.

Perspectives

Pour aller plus loin, il serait intéressant de mettre à jour ces lois avec un historique de données plus profond. Le lissage et surtout le prolongement des lois pourraient être améliorés en utilisant des techniques d'apprentissage automatique se basant sur des données agrégées sur un ensemble de portefeuilles. Ainsi les données du portefeuilles seront utilisées pour la constructions des lois brutes puis le lissage et le prolongement se fera en utilisant des données plus générales.

Résumé

Ces dernières années, la prise en charge de la perte d'autonomie des personnes âgées ou dépendance est devenue un enjeu sociétal. En effet, les progrès de la médecine ont permis des gains en espérance de vie non négligeable. Selon l'INSEE, entre 1950 et 2019, les femmes ont gagné 3,7 années d'espérance de vie et les hommes 6 années. Cependant, les années gagnées ne sont pas forcément des années de bonne santé.

Le rapport de la concertation Grand âge et autonomie de mars 2019 estime que les personnes en perte d'autonomie passeraient de 1,3 millions en 2017 à 2,2 millions en 2050. Face à cette situation préoccupante et une prise en charge des pouvoirs publics insuffisante, des produits d'assurance dépendance sont commercialisés depuis plus d'une trentaine d'années par les différents organismes (compagnies, mutuelles et institutions de prévoyance).

Il est donc nécessaire de modéliser la dépendance afin de mieux cerner le risque et les incertitudes y afférents. Étant donné que chaque portefeuille d'assurés a ses spécificités (Catégorie socio-professionnelle, proportion d'hommes/femmes, âge moyen...) qui font que la modélisation d'un risque similaire basée sur une autre population ne capte pas toutes les particularités, il est indispensable d'assurer un bon pilotage. La profondeur d'historique et le caractère exhaustif des données disponibles peuvent être problématiques mais des méthodes de lissage et de prolongement proposent de bons compromis.

Ce mémoire se donne pour objectif la construction et l'exploitation de lois d'expérience d'un portefeuille d'assurance dépendance de groupe adossé à un ensemble de garanties prévoyance. La première partie de nos travaux a été consacrée à la présentation du risque dépendance avec le contexte démographique actuel et prévisionnel. Puis, dans la deuxième partie, le contrat afférent au portefeuille étudié a été présenté, suivi des données disponibles, de leur traitement et une partie de statistiques descriptives. Ensuite, les lois d'expériences du portefeuille ont été construites en mettant l'accent sur les problématiques du lissage et de l'extrapolation. Enfin, dans la dernière partie, les lois d'incidence en dépendance, de maintien des dépendants et de mortalité des valides obtenues ont été exploitées pour quantifier leur impact au niveau des provisions (PSC et PRC).

Mots clés : Dépendance, Kaplan Meier, Gompertz-Makeham, Whittaker-Henderson, Thatcher, Cox

Abstract

Supporting the loss of autonomy of the elderly has become a societal issue. Indeed, advances in medicine have enabled significant gains in life expectancy. According to INSEE, between 1950 and 2019, women gained 3.7 years of life expectancy and men 6 years. However, years gained are not necessarily years of good health.

The public french report "Concertation Grand âge et autonomie de mars 2019" estimates that people with loss of autonomy will increase from 1.3 million in 2017 to 2.2 million in 2050. Faced with this worrying situation and support from the public authorities insufficient, long-term care insurance products have been marketed for more than thirty years.

It is therefore necessary to model the dependence in order to better understand the risk and the related uncertainties. Given that each portfolio of policyholders has its specificities (socio-professional category, proportion of men / women, average age, etc.) which means that the modeling of a similar risk based on another population does not capture all the particularities, it is essential to ensure a good piloting. The depth of history and completeness of the data available can be problematic, but smoothing and extension methods offer good compromises.

The objective of this paper is to construct and use experimental laws for a group dependency insurance portfolio backed by a set of provident guarantees. The first part of our work was devoted to the presentation of the dependency risk with the current demographic and forecast context. Then, in the second part, the contract relating to the portfolio studied was presented, followed by the available data, their processing and part of descriptive statistics. Then, the experimental laws of the portfolio were built with an emphasis on the problems of smoothing and extrapolation. Finally, in the last part, the laws of incidence in dependency, maintenance of dependents and mortality obtained were used to quantify their impact on provisions (PSC and PRC) .

Keywords : Long-term care, Kaplan Meier, Gompertz-Makeham, Whittaker-Henderson, Thatcher, Cox

Remerciements

J'exprime toute ma gratitude et ma reconnaissance envers Adeline GERARD et Oumnia BENTAHAR pour l'accueil et l'occasion qui m'a été offerte à travers cette année d'apprentissage, et par elles toute l'équipe qui m'a accueilli. Mention spéciale à Oumnia Bentahar qui a encadré mes travaux avec une attention particulière.

Je remercie également mon tuteur académique Guillaume BIESSY pour ses conseils avisés et le temps accordé, ainsi que l'ensemble du corps professoral de l'ISUP pour la formation délivrée ainsi que le soutien.

Mes remerciements vont aussi à l'endroit de toutes les personnes qui m'ont soutenu et motivé, en particulier mes parents et toute ma famille au sens où je l'entends.

Sommaire

Introduction	1
1 Présentation de la dépendance	3
1.1 Définitions	3
1.2 La dépendance en France	7
1.2.1 Contexte	7
1.2.2 Prise en charge	9
2 Contrat et traitement des données	11
2.1 Contexte et contrat	11
2.1.1 Contexte	11
2.1.2 Contrat	11
2.2 Données, traitement et statistiques descriptives	13
2.2.1 Données	13
2.2.2 Traitement des données	14
2.2.3 Statistiques descriptives	15
3 Construction des lois d'expérience	22
3.1 Outils mathématiques	22
3.2 Lois de maintien en dépendance	24
3.2.1 Statistiques descriptives sur les rentiers	24
3.2.2 Estimateur actuariel	27
3.2.3 Estimateur de Kaplan-Meier	27
3.2.4 Modèle de Cox	32
3.3 Lois d'incidence dépendance	40
3.3.1 Test du log-rank	44
3.3.2 Lissage des taux bruts	45
3.3.3 Comparaison des lissages	54
3.3.4 Extrapolation sur les âges inférieurs	59
3.3.5 Prolongement des lois sur les âges élevés	60
3.4 Lois de mortalité des valides	63
3.4.1 Taux bruts	63
3.4.2 Lissage	64
3.4.3 Extrapolation des lois de mortalité des valides	66

4 Impacts sur le provisionnement **69**
4.1 Impact sur la PSC 69
4.2 Impact sur la PRC 70

Conclusion **73**

Bibliographie **76**

Annexe **77**

Introduction

Ces dernières années, l'allongement de l'espérance de vie et l'arrivée de la génération "baby-boom" aux âges élevés font de la perte d'autonomie des personnes âgées ou dépendance un enjeu sociétal. En effet, les progrès de la médecine ont permis des gains en espérance de vie non négligeable. Selon l'INSEE, entre 1950 et 2019, les femmes ont gagné 3,7 années d'espérance de vie et les hommes 6 années. Cependant, les années gagnées ne sont pas forcément des années de bonne santé. C'est ce qui rend pertinent l'usage de l'Espérance de Vie Sans Incapacité (EVSI).

Estimée à 1,5 millions en 2017 par le rapport de la concertation Grand âge et autonomie de mars 2019, les personnes âgées de plus de 85 ans devraient atteindre 4,8 millions en 2050, soit une hausse de 220%. Parallèlement, les personnes en perte d'autonomie passeraient de 1,3 millions en 2017 à 2,2 millions en 2050. Face à cette situation préoccupante, le système public français a commencé la prise en charge de la dépendance en 1985 à travers des dispositifs qui se sont relayés et une politique de sensibilisation.

Cependant, cette prise en charge des pouvoirs publics ne saurait être suffisante. En effet, selon le rapport de la concertation Grand âge et autonomie de mars 2019, les pensionnaires des établissements en situation de dépendance ont un reste à charge moyen de 1 850 euros. Voilà pourquoi, des produits d'assurance dépendance sont commercialisés depuis plus d'une trentaine d'années par les différents organismes (compagnie, mutuelles et institution de prévoyance). Leur objectif est de financer les coûts liés à une éventuelle perte d'autonomie.

Il est donc justifié et nécessaire de modéliser la dépendance afin de mieux cerner le risque et les incertitudes y afférents. Cela pourrait prévenir les mauvaises surprises sur un risque en pleine évolution et des pertes pour les acteurs. Néanmoins, chaque portefeuille d'assurés a ses spécificités (Catégorie socio-professionnelle, proportion d'hommes/femmes, âge moyen...) qui font que la modélisation d'un risque similaire basée sur une autre population ne capte pas toutes les particularités. Il est alors intéressant de faire ce travail de suivi. La profondeur d'historique et le caractère exhaustif des données disponibles peuvent être problématiques mais des méthodes de lissage et de prolongement proposent de bons compromis.

C'est dans cette optique que ce mémoire se donne pour objectif la construction et l'exploitation de lois d'expérience d'un portefeuille d'assurance dépendance de groupe adossé à un ensemble de garanties prévoyance.

La première partie de nos travaux sera consacrée à la présentation du risque dépendance avec le contexte démographique actuel et prévisionnel ainsi que la prise en charge par les pou-

voirs publics et organismes d'assurance en France. Puis, dans la deuxième partie, le contrat afférent au portefeuille étudié sera présenté, suivi des données disponibles, de leur traitement et une partie de statistiques descriptives. Ensuite, les lois d'expériences du portefeuille seront construites en mettant l'accent sur les problématiques du lissage et du prolongement. Il s'agira dans un premier temps des lois de maintien des dépendants, ensuite des lois d'incidence en dépendance, puis des lois de mortalité des valides. Enfin, dans la dernière partie, les lois obtenues seront exploitées pour quantifier leur impact au niveau des provisions (PM et PRC) et aussi produire des indicateurs de suivi démographique et financier du portefeuille.

Chapitre 1

Présentation de la dépendance

Dans ce premier chapitre, la dépendance sera présentée sous différents angles et à travers différentes définitions. Le cas de la France sera ensuite étudiée plus en détails.

1.1 Définitions

Différentes définitions admises et différentes grilles (France et internationale)

Il n'existe pas une mais des définitions de la dépendance qui varient en fonction des acteurs et des zones géographiques.

Généralement définie comme le besoin permanent et irréversible d'une tierce personne pour effectuer les actes de la vie courante, la dépendance est souvent le résultat du vieillissement et des maladies chroniques voire plus rarement d'un accident.

Cette définition laisse apparaître une première grille d'évaluation basée sur un certain nombre d'Actes de la Vie Quotidienne (AVQ) dont la réalisation ne peut être effectuée même avec l'aide d'équipement adapté, et nécessite l'aide d'une tierce personne de manière permanente. Les AVQ sont utilisés dans de nombreux pays pour la définition de la dépendance. Plusieurs définitions basées sur 4, 5 ou 6 AVQ coexistent et utilisent les AVQ suivantes :

- Faire sa toilette
- S'alimenter
- S'habiller
- Se déplacer
- Être continent
- Effectuer ses transferts

La perte d'une AVQ doit être irréversible pour être considérée. La reconnaissance de la dépendance est assujettie à la perte d'un nombre minimum d'AVQ. A titre d'exemple, les personnes classées 3AVQ4 ont perdu 3 AVQ sur les 4 retenues pour la définition, c'est une définition usuelle de la dépendance totale.

Par ailleurs, des tests complémentaires neurologiques et psychiatriques peuvent être utilisés pour la prise en compte des dépendances d'origine neuropsychiatrique. L'un des tests les plus répandus est le «Mini Mental Score» de Folstein (MMS).

En plus des AVQ éventuellement combinées d'un test complémentaire, une autre définition est très utilisée en France, elle est basée sur la grille Autonomie Gérontologie Groupe Iso-Ressource (AGGIR). Cette classification est créée conjointement par les Médecins de la Sécurité Sociale et la «Société Française de Gérontologie» en 1997. Elle est utilisée par l'État français pour estimer le niveau de perte d'Autonomie des personnes âgées et calculer le montant de l'aide publique en cas de dépendance. Aujourd'hui utilisée également par certaines compagnies d'assurance, la grille AGGIR classe les individus en six niveaux, du plus dépendant (GIR 1) au plus autonome (GIR 6).

La grille AGGIR prend en compte des critères sociaux, environnementaux et psychologiques. Elle fait appel à dix-sept variables réparties comme suit :

- 10 variables dites discriminantes se rapportant à la perte d'autonomie physique et psychique
- 7 variables dites illustratives, se rapportant à la perte d'autonomie domestique et sociale.

Les variables discriminantes sont :

1. cohérence : converser et/ou se comporter de façon logique et sensée ;
2. orientation : se repérer dans le temps, les moments de la journée et dans les lieux ;
3. toilette : concerner son hygiène corporelle ;
4. habillage : s'habiller, se déshabiller, choisir ses vêtements ;
5. alimentation : se servir et manger des aliments préparés ;
6. hygiène de l'élimination : assurer l'hygiène de l'élimination urinaire et fécale ;
7. transferts : se lever, se coucher, s'asseoir ;
8. déplacement à l'intérieur : du logement ou de l'institution (à l'aide éventuellement d'une canne, d'un fauteuil roulant ou d'un déambulateur)
9. déplacement à l'extérieur : à partir de la porte d'entrée, à pied ou en fauteuil ;
10. communication à distance : utiliser les moyens de communication (téléphone alarme, sonnette).

Seules les huit premières variables interviennent dans le calcul du GIR, les 2 dernières sont des variables instrumentales au même titre que les variables illustratives.

Les variables illustratives visent à évaluer le contexte environnemental et social de la personnes et sont :

1. gestion ;
2. cuisine ;
3. ménage ;
4. transports ;

5. achats ;
6. suivi des traitements ;
7. activités de temps libre.

De plus, chaque variable est cotée selon trois modalités :

- A : fait seul, totalement, habituellement et correctement ;
- B : fait partiellement, ou non habituellement, ou non correctement ;
- C : ne fait pas

La figure 1.1 reprend la fiche parue au Journal Officiel en 1997 et qui est utilisée pour la cotation informatique dans les logiciels actuels.

In fine, de la classification ressortent les 6 groupes suivants, par ordre décroissante de besoin d'une assistance tierce personne :

GIR 1 : personnes confinées au lit ou au fauteuil ayant perdu leur autonomie mentale, corporelle, locomotrice et sociale, qui nécessitent une présence indispensable et continue d'intervenants ;

GIR 2 : d'une part, les personnes confinées au lit ou au fauteuil dont les fonctions mentales ne sont pas totalement altérées et qui nécessitent une prise en charge pour la plupart des activités de la vie courante, d'autre part, celles dont les fonctions mentales sont altérées mais qui ont conservé leurs capacités motrices. Le déplacement à l'intérieur est possible mais la toilette et l'habillage ne sont pas faits ou partiellement ;

GIR 3 : personnes ayant conservé leur autonomie mentale, partiellement leur autonomie locomotrice, mais qui nécessitent quotidiennement et plusieurs fois par jour des aides pour leur autonomie corporelle. Ainsi, la toilette et l'habillage ne sont pas faits ou faits partiellement. De plus, l'hygiène de l'élimination nécessite l'aide d'une tierce personne ;

GIR 4 : personnes qui n'assument pas seules leur transfert mais qui une fois levées, peuvent se déplacer à l'intérieur du logement. Elles doivent être aidées pour la toilette et l'habillage. La plupart s'alimentent seules ; ce groupe comprend aussi des personnes sans problème de locomotion mais qu'il faut aider pour les activités corporelles et les repas ;

GIR 5 : personnes autonomes dans leurs déplacements chez elles qui s'alimentent et s'habillent seules. Elles peuvent nécessiter une aide ponctuelle pour la toilette, la préparation des repas et le ménage ;

GIR 6 : personnes qui n'ont pas perdu leur autonomie pour les actes discriminants de la vie quotidienne.

Peuvent avoir droit à l'Allocation Personnalisée d'Autonomie (APA) les personnes classées dans les quatre premiers GIR.

Il existe d'autres méthodes d'évaluation de la dépendance telles que la grille de Colvez qui mesure la perte de mobilité uniquement. Cette diversité des définitions et des modes d'évaluation de la dépendance rend difficile la comparaison des contrats et des offres entre elles. C'est dans ce cadre que le label Garantie Assurance Dépendance (GAD) a été développé par la FFSA. Il s'agit d'une harmonisation à minima sur la définition et le vocabulaire utilisés ainsi qu'une garantie minimale. Ceci dans le but de faciliter la comparaison des contrats

Nom et prénom
 MI Sec.Soc.
 Adresse
 Né(e) le
 Âge

Fiche récapitulative AGGIR

Date de l'évaluation

Activités réalisées par la personne seule	Pour chaque item, cocher les cases quand les conditions ne sont pas remplies (Réponse NON)				Code	Code final	Activités corporelles, mentales, domestiques et sociales.	
	S	T	C	H				
1. Cohérence	communication	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	[]	<p>Codage intermédiaire</p> <p>Pour chaque item cochez les cases quand les conditions ne sont pas remplies (réponse NON).</p> <p>Puis codez secondairement par A, B ou C selon le nombre d'adverbes cochés dans les quatre cases S à H.</p> <ul style="list-style-type: none"> • Si aucun adverbe n'est coché codez A. (fait spontanément, totalement, correctement et habituellement) • Si tous les adverbes sont cochés codez C (ne fait pas) • Si une partie des adverbes seulement est cochée codez B. <p>Code final si sous-variables</p> <ul style="list-style-type: none"> • Cohérence : <ul style="list-style-type: none"> - AA = A ; - CC, CB, BC, CA, AC = C ; - AB, BA, BB = B • Orientation : <ul style="list-style-type: none"> - AA = A ; - CC, CB, BC, CA, AC = C ; - AB, BA, BB = B • Toilette : <ul style="list-style-type: none"> - AA = A ; - CC = C ; - Autres = B • Habillage : <ul style="list-style-type: none"> - AAA = A ; - CCC = C ; - Autres = B. • Alimentation : <ul style="list-style-type: none"> - AA = A ; - CC, BC, CB = C ; - Autres = B • Élimination : <ul style="list-style-type: none"> - AA = A ; - CC, BC, CB, AC, CA = C ; - Autres = B 	
	comportement	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			[]
2. Orientation	dans le temps	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	[]		
	dans l'espace	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			[]
3. Toilette	haut	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	[]		
	bas	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			[]
4. Habillage	haut	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	[]		
	moyen	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			[]
	bas	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			[]
5. Alimentation	se servir	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	[]		
	manger	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			[]
6. Élimination	urinaire	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	[]		
	fécale	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>			[]
7. Transferts		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	[]		
8. Déplacements à l'intérieur		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	[]		
9. Déplacements à l'extérieur		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	[]		
10. Alerter		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	[]		
11. Gestion		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	[]		
12. Cuisine		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	[]		
13. Ménage		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	[]		
14. Transports		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	[]		
15. Achats		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	[]		
16. Suivi du traitement		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	[]		
17. Activités du temps libre		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	[]		

Groupe iso-ressources
 Défini par le système informatique []

FIGURE 1.1 – Fiche récapitulative évaluation AGGIR
 Source :assurance-obseques.fr,13/08/2020

d'assurance dépendance entre-eux. Les contrats d'assurance labellisés GAD - Assurance Dépendance doivent présenter une définition commune, un montant minimal de rente pour la dépendance lourde et un vocabulaire commun. Ce qui est loin d'être une vérité générale sur le marché de la dépendance en France.

1.2 La dépendance en France

1.2.1 Contexte

En France, en 2019, l'espérance de vie à la naissance est de 85,6 ans pour les femmes et 79,7 ans pour les hommes. Ce qui représente une augmentation respective de 3,7 ans et 6 ans en vingt-cinq ans. Ci-dessous une illustration de l'évolution de l'espérance de vie de la population française entre 1950 et 2019.

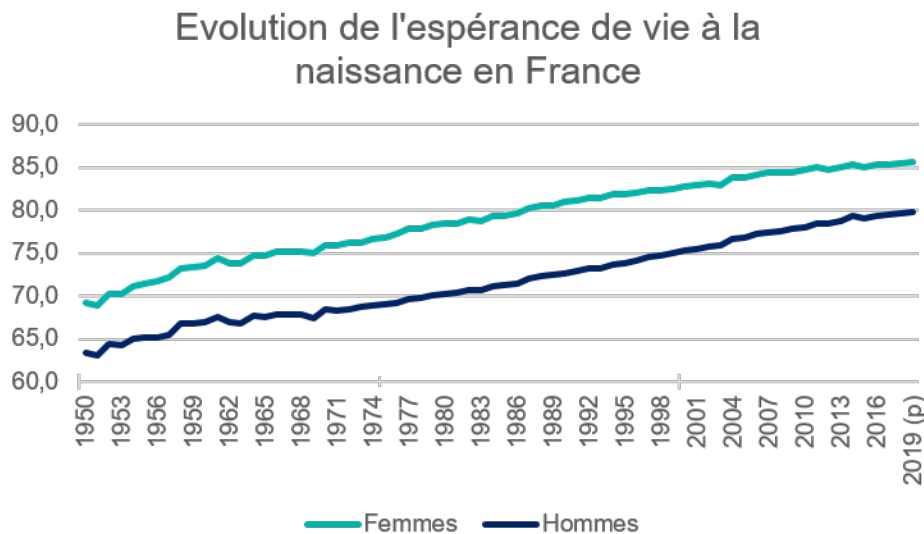


FIGURE 1.2 – Évolution de l'espérance de vie en France

Source : Insee, estimations de population et statistiques de l'état civil, 27/02/2020

En particulier, à 60 ans, en moyenne, une femme peut espérer vivre encore 27,8 ans et un homme 23,4 ans, soit une hausse respective de 2,9 ans et 3,7 ans en vingt-quatre ans.

Cependant, les années gagnées en espérance de vie ne seront pas forcément des années en bonne santé, ce qui rend pertinent l'usage de l'Espérance de Vie Sans Incapacité (EVSI). Selon une étude de la DREES basée sur les données INSEE, alors que l'espérance de vie à 65 ans s'élevait à 19,4 ans pour les hommes et 23,2 ans pour les femmes en 2018, l'EVSI à 65 ans n'était que de 10,1 ans pour les hommes et 11,2 ans pour les femmes. Il est également à noter que l'EVSI est restée quasi-stable sur les dix dernières années telle qu'il apparaît sur la figure 1.2.

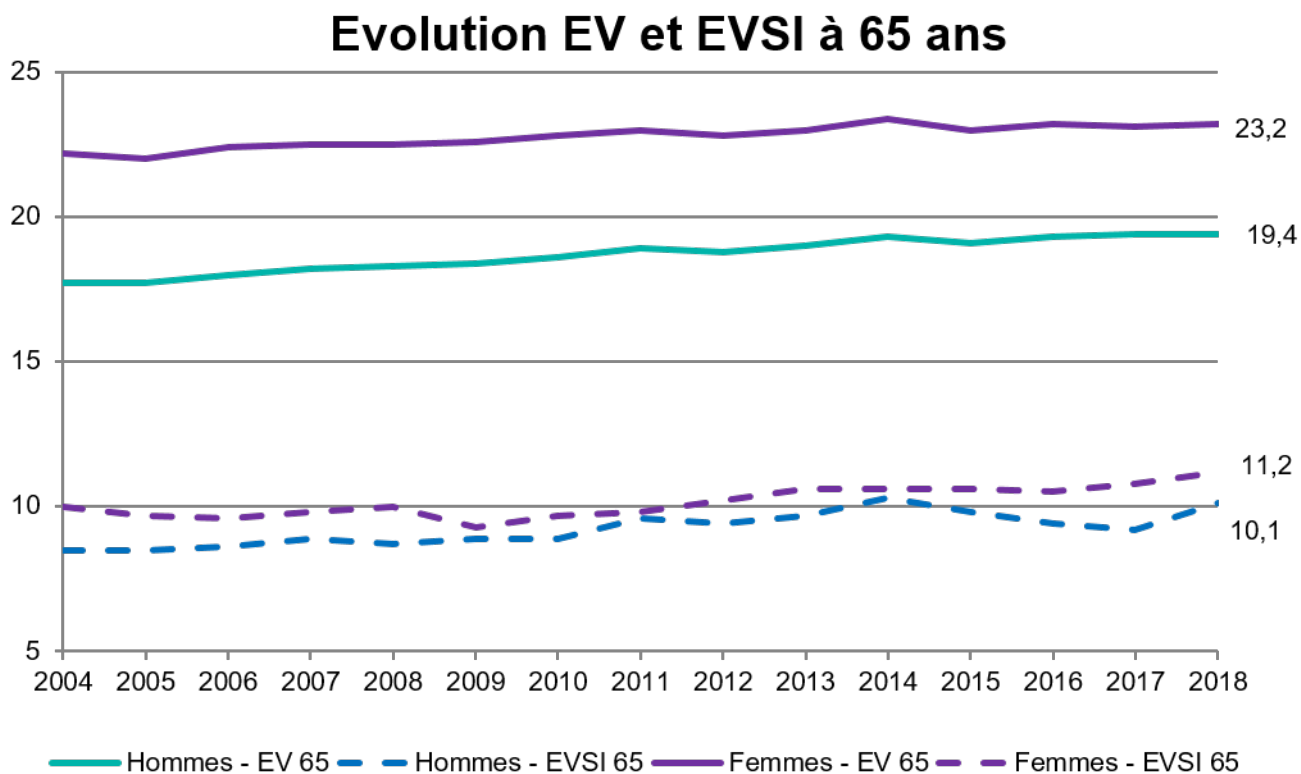


FIGURE 1.3 – Évolution comparée EV et EVSI

Source : Insee, Statistiques d'Etat civil et données issues de l'enquête SILC (Statistiques sur les revenus et les conditions de vie) ; EVSI : Calculs de la DREES avec la méthode d'Ehemu

Selon le rapport de la concertation Grand âge et autonomie de mars 2019, la France comptait 1,5 millions de personnes âgées de plus de 85 ans en 2017. Une augmentation de 220% est attendue à l'horizon 2050 portant les plus de 85 ans à 4,8 millions. Quant aux personnes en perte d'autonomie, elles représentaient 1,3 millions en 2017. Les prévisions stipulaient une hausse de 70% en 2050, soit 2,2 millions.

Les dépenses publiques concernant la prise en charge de la dépendance des personnes âgées étaient estimées à 24 milliards d'euros en 2014.

1.2.2 Prise en charge

La prise en charge de la perte d'autonomie par le système public français a commencé en 1985 avec l'Allocation Compensatrice Tierce Personne (ATP). Elle couvrait, sans condition d'âge, la dépendance totale nécessitant des soins. Une aide sociale supplémentaire était accordée aux personnes vivant avec un handicap. En 1996, l'ATP comptait 200 000 bénéficiaires dont 50 000 de plus de 60 ans. Elle est remplacée par la Prestation Spécifique Dépendance (PSD) en 1997.

La PSD était réservée aux personnes de plus de 60 ans classées GIR 1, 2 ou 3. Son attribution était soumise à des conditions de ressources avec un possible recouvrement sur la succession. En 2001, 150 000 personnes bénéficiaient de la PSD. Elle devient l'Allocation Personnalisée Autonomie (APA) le 1^{er} janvier 2002.

L'APA élimine les conditions de ressources et est ouverte aux GIR 4. Les conditions de ressources sont supprimées dans l'attribution de la prestation qui reste dépendante de celles-ci. De plus, la possibilité de recouvrement sur la succession est révoquée. En fin 2017, l'APA comptait 1 265 000 bénéficiaires de plus de 60 ans pour des prestations estimées à 8 milliards d'euros.

Malgré la prise en charge de la perte d'autonomie des personnes âgées par les pouvoirs publics, les charges peuvent être très lourdes. En effet, en établissement, le reste à charge mensuel moyen est estimé à 1 850 euros, montant supérieur aux revenus de 75% des personnes concernées. D'où l'émergence de l'assurance dépendance.

Si dans les années 85 les quelques produits individuels qui se lançaient ne prenaient en charge que la dépendance totale avec des rentes, le marché de l'assurance dépendance s'est démocratisé progressivement. Dès 1995, la dépendance partielle et l'assistance intègrent progressivement les contrats qui s'élargissent également à des produits groupe. Les définitions de la dépendance commencent à se baser sur la grille AGGIR et les AVQ et les prestations se diversifient intégrant des possibilités de capital. A partir des années 2000, les services d'assistance se développent et les dépendances totale et partielle cohabitent sur les contrats. Depuis 2009, l'avènement du label GAD Assurance Dépendance crée une nouvelle génération de contrat. Les couvertures de groupes se développent, la sélection médicale s'allège et les proches-aidants sont pris en compte dans les offres. L'heure est aussi à l'innovation avec la domotique, des prestations en cas de dépendance légère, des options capital décès ou la baisse des cotisations au-delà d'un certain âge.

Deux types de contrats d'assurance dépendance existent : les contrats à adhésion individuelle et les contrats collectifs. Les contrats individuels sont généralement à adhésion facultative et le montant des primes varie d'un assuré à l'autre, tandis que l'adhésion aux contrats collectifs est souvent obligatoire, et dans ce cas, la prime est assez intéressante pour l'ensemble du groupe. Quant à la garantie, elle peut être annuelle lorsque l'assureur ne couvre que le risque de l'année en cours ou viagère si l'assuré est couvert jusqu'à son décès. Dans ce cas, une provision pour risque croissant est en général constituée. Par ailleurs, certains contrats prévoient une réduction de rente en fonction du nombre d'années de cotisation en cas d'arrêt de paiement des primes.

Selon la FFA, 7,1 millions de personnes étaient couvertes contre le risque dépendance en fin 2017. Ce qui représentent 783 millions d'euros de cotisation dont 650 millions pour les sociétés d'assurance.

Parmi les 2,7 millions de personnes couvertes auprès des sociétés d'assurance en fin 2017, 1,6 million ont souscrit un contrat pour lequel la dépendance est la principale et unique garantie. 79% de ces contrats dépendance à garantie unique sont des adhésions individuelles avec une cotisation annuelle de 457 euros en moyenne et une rente mensuelle en cas de dépendance de 598 euros en moyenne en 2017. Ce sont 258,6 millions d'euros de prestations qui ont été versées en 2017 aux bénéficiaires des contrats individuels et collectifs dont la garantie principale et unique est la dépendance. Au même moment, les provisions constituées s'élèvent à 5,4 milliards d'euros au 31 décembre 2017.

Chapitre 2

Contrat et traitement des données

Dans ce chapitre seront exposées les caractéristiques du portefeuille étudié à travers le contrat en question et les données utilisées.

2.1 Contexte et contrat

2.1.1 Contexte

Assureur, coassureur et réassureur, CNP Assurances conçoit des solutions de protection en assurance vie (épargne et retraite), prévoyance, santé et couverture de prêts pour ses partenaires et leurs clients. Société anonyme au capital de 686 618 477 euros, le groupe compte 38 millions d'assurés en prévoyance, protection et santé dans le monde, et plus de 14 millions en épargne/retraite. Partenaire du monde paritaire et de nombreuses mutuelles, CNP Assurances propose des solutions sur mesures pour protéger leurs adhérents et les salariés des entreprises.

Pionnier de l'assurance dépendance, CNP Assurances a été l'un des premiers assureurs à se préoccuper des enjeux de dépendance. Le Groupe est l'un des leaders en assurance collective et détient 9% du marché en assurance individuelle en garantie principale et unique. CNP Assurances propose une gamme de produits à adhésion obligatoire ou facultative qui permet d'anticiper les besoins financiers et l'accompagnement en cas de perte d'autonomie. Notre étude porte sur l'un des contrats historique en adhésion collective obligatoire .

En effet, les données utilisées sont issues d'un portefeuille dépendance que réassure CNP Assurances depuis décembre 2006 auprès d'une mutuelle qui sera désignée La Mutuelle pour la suite. Par souci de simplicité, CNP Assurances pourra être appelé Le Réassureur.

2.1.2 Contrat

Nos travaux portent sur un contrat d'assurance de groupe à adhésion obligatoire couvrant la dépendance totale uniquement. La garantie dépendance est adossée à d'autres garanties prévoyance et ses principales caractéristiques sont les suivantes :

Généralités Le produit est lancé en 2002 par La Mutuelle et prévoit le versement d'une rente mensuelle aux adhérents en situation de dépendance totale avec une définition basée sur

cinq AVQ combinés à un système de points. Le groupe est en quasi-fermeture au 31/12/2008, seuls les enfants des membres participants peuvent devenir adhérents.

Bénéficiaires Peuvent souscrire au produit les membres participants (actifs et retraités) de La Mutuelle et les membres adhérents affiliés aux contrats collectifs sous réserve des conditions d'admission à leur bénéfice uniquement. En effet, le contrat ne couvre pas les conjoints et ayant droits mais le maintien des garanties est possible pour les anciens membres.

Conditions d'adhésion Pour être admis, le futur assuré devra avoir un âge compris entre 16 et 60 ans inclus au 31/12 de l'année en cours et remplir les conditions de santé suivantes :

- Ne pas être dépendant avec le besoin d'une ATP pour la réalisation d'un AVQ,
- Ne pas être en situation d'invalidité 2ème ou 3ème catégorie selon la Sécurité Sociale,
- Ne pas avoir un taux d'invalidité supérieur ou égal à 60% en tant que fonctionnaire relevant des articles L.27 ou L.29 et L.30 du Code des pensions civiles et militaires,
- Ne pas être bénéficiaire d'une rente accident de travail pour une incapacité supérieure ou égale à 40%.

Les personnes exonérées du ticket modérateur au titre d'une affection liste, hors liste ou d'une polyopathie définie par la réglementation de la Sécurité Sociale sont admises sous réserve de l'application d'un délai de carence. Quant aux personnes en arrêt de travail, elles ne peuvent être admises qu'après reprise de leur activité, sous réserve des dispositions ci-dessus concernant leur état de santé.

Définition de la dépendance . La dépendance est l'impossibilité, médicalement constatée, d'accomplir sans aide les actes élémentaires de la vie courante (s'alimenter, s'habiller, faire sa toilette, faire ses transferts, faire ses besoins naturels). Est considéré comme dépendant total pour l'attribution de la prestation, l'assuré totalisant au moins 33 points sur la base du barème résumé dans le tableau qui suit.

Capacité physique ou mentale à accomplir un acte	Seul	Avec une aide partielle	Avec une aide totale
S'alimenter (couper et porter les aliments à sa bouche, se servir à boire. . .)	0	3	10
S'habiller (se vêtir, se dévêtir)	0	3	10
Faire sa toilette (assurer son hygiène personnelle par sa toilette)	0	3	10
Faire ses transferts (se lever, se coucher, s'asseoir, se déplacer)	0	3	10
Faire ses besoins naturels	0	3	10

TABLE 2.1 – Définition de la dépendance selon le contrat

Date d'effet des garanties Ce contrat ne comporte pas de délai de carence à l'exception des personnes exonérées du ticket modérateur pour lesquelles le délai est de 12 mois. Les risques survenant pendant cette période ne sont pas garantis et ne donnent lieu à aucune indemnisation. Dans ce cas, les cotisations versées par le membre participant lui sont

intégralement remboursées. Le délai de carence ne s'applique pas en cas d'accident. Il existe cependant, pour tous les adhérents, une franchise de 90 jours à partir de la date de reconnaissance de l'état de dépendance totale par le médecin conseil.

Cotisations Les cotisations sont dues à compter de la prise d'effet de la garantie. Elles sont annuelles et fixées en fonction de l'âge atteint par l'assuré au 1er janvier de l'année selon un barème défini. L'âge est calculé par différence de millésime. Les cotisations cessent d'être dues dès la reconnaissance de la dépendance ou au décès de l'assuré. En cas d'arrêt de paiement des primes, il n'existe pas de valeur de réduction.

Prestations Les prestations sont des rentes viagères mensuelles versée en cas de dépendance totale selon la définition du contrat. Elles sont dues à partir de l'expiration du délai de franchise jusqu'au décès de l'assuré. La rente mensuelle d'un assuré décédé en cours de mois est entièrement dues aux héritiers légaux de ce dernier. La Mutuelle peut suspendre le service des prestations voire les supprimer si l'état du bénéficiaire n'en justifie plus le maintien. La rente mensuelle est de 340 au 1er janvier 2019 et sera indexée conformément au mécanisme de revalorisation.

Revalorisation et révision

La politique de revalorisation dépend uniquement de la stratégie de La Mutuelle. Les revalorisations des rentes dépendance en cours de service et des garanties sont décidées dans la limite disponible du fonds de revalorisation. Les rentes en cours de service sont revalorisées à due concurrence des provisions pour rentes dépendance en cours de service. Quant aux garanties en cours non sinistrées, elles sont revalorisées d'une part, par la mise à niveau des provisions pour risque croissant et d'autre part, par l'augmentation des cotisations proportionnellement à la revalorisation des garanties.

Les révisions tarifaires sont possibles sur la base des études périodiques réalisées par Le Réassureur avec les données transmises annuellement par La Mutuelle. Elles peuvent concerner aussi bien les adhésions nouvelles que celles en cours. Des éléments contractuels pourraient aussi être modifié en guise de compensation.

2.2 Données, traitement et statistiques descriptives

2.2.1 Données

Ce contrat étant historique avec une première souscription au 01/07/2002, nous disposons d'une base de données conséquente et robuste aussi bien pour les assurés valides que les sinistres.

Avec un recul de 16 années, nous disposons des données suivantes arrêtées au 03/12/2018 pour 424 181 assurés :

- Clé pour définir un assuré
- Date de naissance
- Date de début de garantie
- Genre (1 - Homme, 2 - Femme)
- Cotisation émise pour octobre 2018

- Motif de sortie :
- Date de sortie (31/12/9999 pour les assurés présents)
- Date de passage en dépendance (vide pour les valides)
- Minoration de cotisation
- Date de décès (vide pour les valides)

Les motifs de sorties sont au nombre de 4 :

1. Chute suite à décès avéré enregistré sur les garanties Décès statutaires
2. Chute sans décès avéré sur les garanties Décès statutaires
3. Chute sans adhésion retrouvée à une garantie Décès statutaire
4. Non sorti

Pour les sinistres, nous avons les mêmes informations pour 5 686 dépendants :

- Clé pour définir un assuré
- Date de naissance
- Date de début de garantie
- Genre (1 - Homme, 2 - Femme)
- Cotisation émise pour octobre 2018
- Motif de sortie :
- Date de sortie (31/12/9999 pour les sinistres en cours)
- Date de passage en dépendance
- Date de décès (31/12/9999 pour les décès non renseignés).

Les motifs de sortie restent les mêmes mais sont des décès en réalité.

Nous observons les premiers sinistres en décembre 2005 et avons des données robustes à partir de 2012, soit 6 ans de recul.

2.2.2 Traitement des données

Tout travail statistique nécessite des données de bonnes qualités. En actuariat, cette exigence est renforcée par la directive Solvabilité 2 qui veut que les données soient exactes, exhaustives et appropriées. Dans cette optique, la base de données a été analysée en tenant compte des caractéristiques du contrat et des informations fournies par La Mutuelle. Ce qui a valu les retraitements suivants.

Début de garantie Sur les 424 181 assurés de la base, 221 662 ont une date de début de garantie antérieure au lancement du produit. Leur date de début de garantie a été ramenée au 01/07/2002 correspondant à la création du produit. 107 assurés ont une date de sortie égale ou antérieure à la date de début de leur contrat et 2 assurés ont une date d'entrée en dépendance antérieure au début de la garantie. Ces lignes ont été supprimés, soit 109 assurés.

Passage en dépendance Parmi les 424 072 assurés restants, 16 ont une date d'entrée en dépendance égale ou postérieure à leur date de décès. On considère qu'il n'y a pas eu

de dépendance, les dates d'entrée en dépendance sont supprimées. 10 autres assurés ont une date d'entrée en dépendance égale ou postérieure à leur date de chute. On considère qu'il n'y a pas eu de dépendance, les dates d'entrée en dépendance sont supprimées.

A partir de ce fichier, une base des rentiers a été extraite.

Décès des dépendants 738 assurés sont sortis de dépendance sans date de décès renseignée. Parmi eux, 562 sont confirmés décédés à travers la garantie statutaire décès. La Mutuelle confirme qu'il s'agit bien d'un décès pour tous les 738 assurés. On crée une nouvelle colonne date de décès qui reprend la date de décès si elle a été notifié sinon la date de sortie.

Date de sortie Bien que l'extraction ait été faite en fin 2018, 2215 assurés ont une date de sortie en 2019 dont 2123 en janvier 2019. Ces sorties sont considérées comme des résiliations qui ont été renseignées dès le début de leur préavis, elles sont donc justifiées. Au final, 92 sorties en 2019 sont inexplicables, ce nombre étant négligeable pour pouvoir remettre en cause les études envisagées, les lignes en question sont conservées.

Les deux tableaux suivants résument les données in fine :

Nombre d'assurés	424 072
Âge moyen à l'adhésion	47
Part des hommes	55%
Autonomes Présents	308 637
Passages en dépendance	5 658
Dépendants Présents	2 458

TABLE 2.2 – Récapitulatif du portefeuille

Nombre de décès	24 048
Décès en autonomie	20 848
Décès en dépendance	3 200

TABLE 2.3 – Mortalité du portefeuille

2.2.3 Statistiques descriptives

Population assurée

La base compte 424 072 assurés comptabilisés depuis la date de lancement du produit. Ils sont répartis sur la figure qui suit en fonction des années d'adhésion. Les adhésions en 2002 constituent l'essentiel du portefeuille. Cela s'explique par le fait que le produit ait été mis en place en 2002 avec une adhésion collective obligatoire. La répartition des adhésions hors 2002 est aussi représentée pour plus de visibilité. Il est à noter les faibles adhésions à partir de 2009 qui correspondent à des souscriptions des enfants des adhérents de La Mutuelle, le

Répartition par année d'adhésion

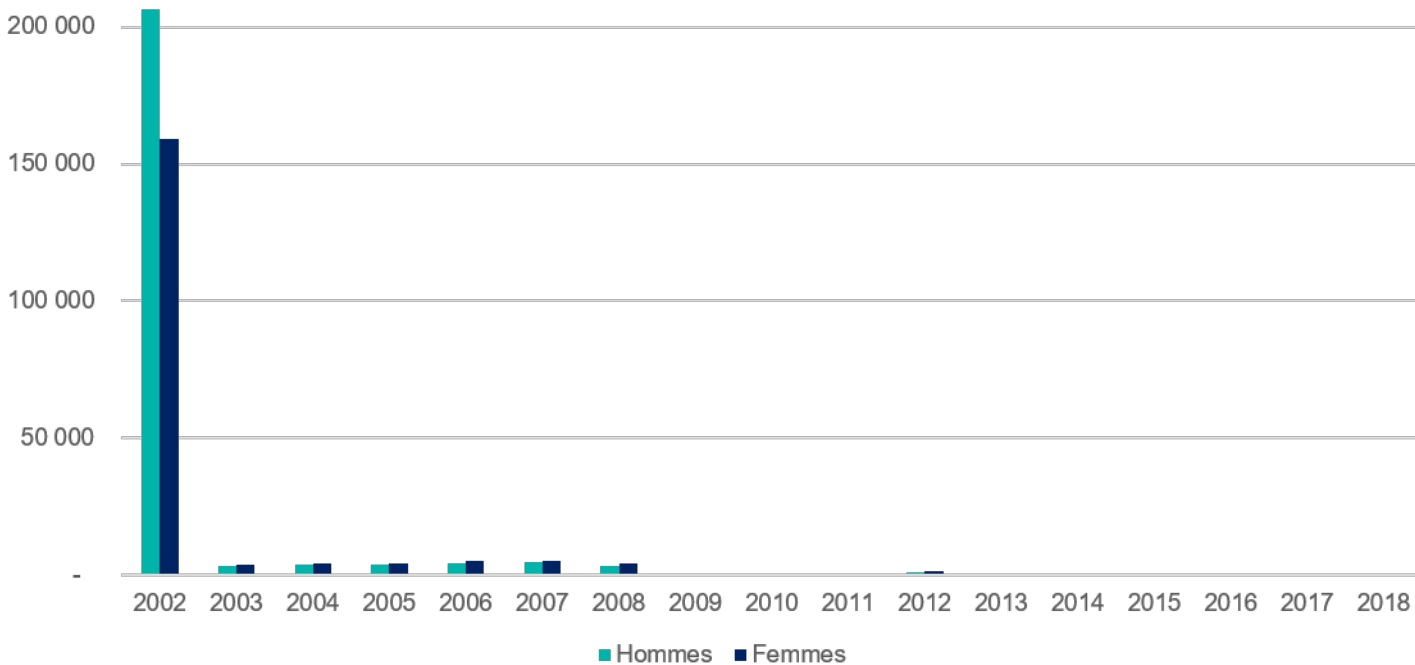


FIGURE 2.1 – Répartition des assurés par année d'adhésion

groupe étant en quasi-fermeture.

Les adhérents de 2002 sont majoritairement constitués d'hommes et les adhésions à partir de 2003 comptent un peu plus de femmes. Au final la population assurée est composée de près de 55% d'hommes.

Concernant l'âge, les souscriptions sont majoritairement faites à 21 ans. Cela peut également s'expliquer par le caractère obligatoire de l'adhésion pour toute personne rejoignant le groupe à supposer que les recrutements sont concentrés autour de cet âge. Un deuxième pic est localisé autour de 45 ans près de l'âge moyen d'adhésion qui est de 46,97 ans. Plus de 16% des assurés adhèrent au contrat après 60 ans, ce qui est une part non négligeable. La structure est détaillée sur la pyramide des âges.

Dépendance

Bien que la population assurée soit majoritaire masculine, les femmes sont les plus représentées parmi les assurés en situation de dépendance. Les entrées en dépendance se concentrent entre 75 et 95 ans avec un pic autour de 88 ans. L'âge moyen d'entrée en dépendance est 83,15 ans. Avant 80 ans les hommes sont plus nombreux parmi les dépendants mais la tendance s'inverse au-delà notamment parce que dans la population assurée, les femmes sont plus représentées à ces âges. Le graphique qui suit représente la pyramide des

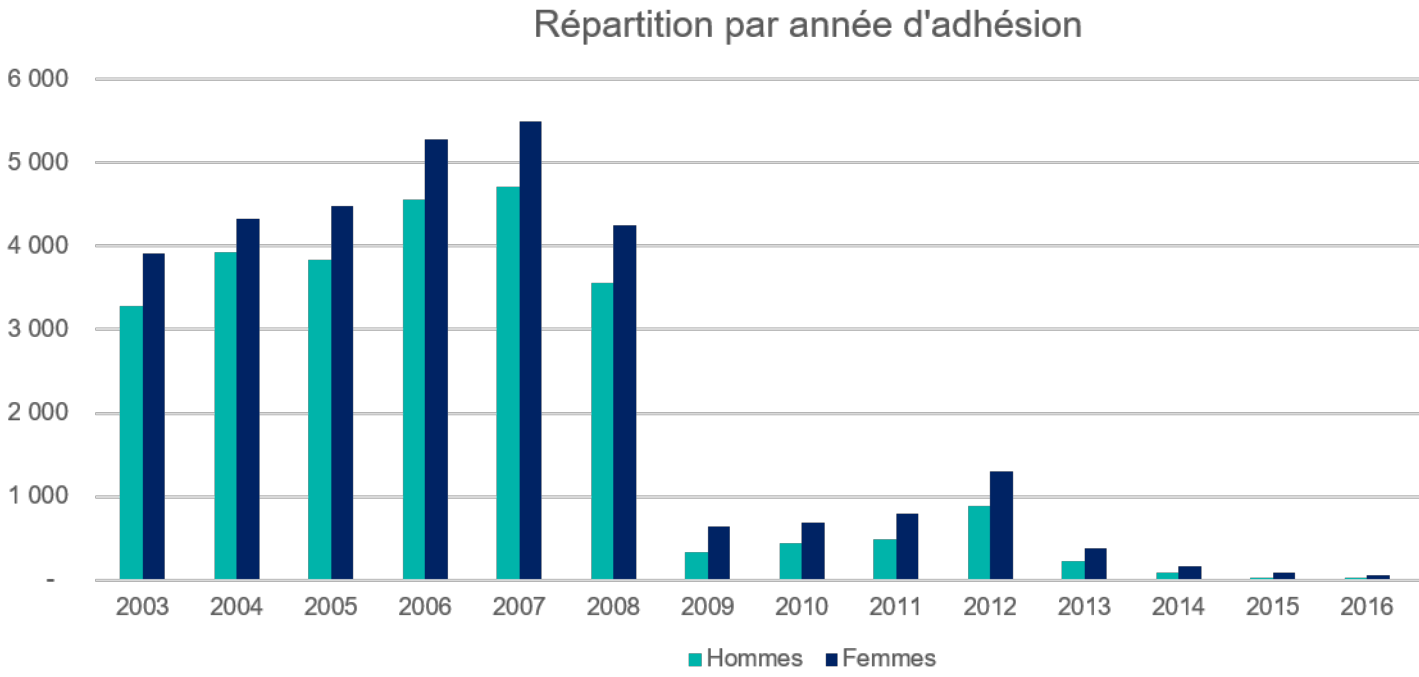


FIGURE 2.2 – Répartition des adhésions hors 2002

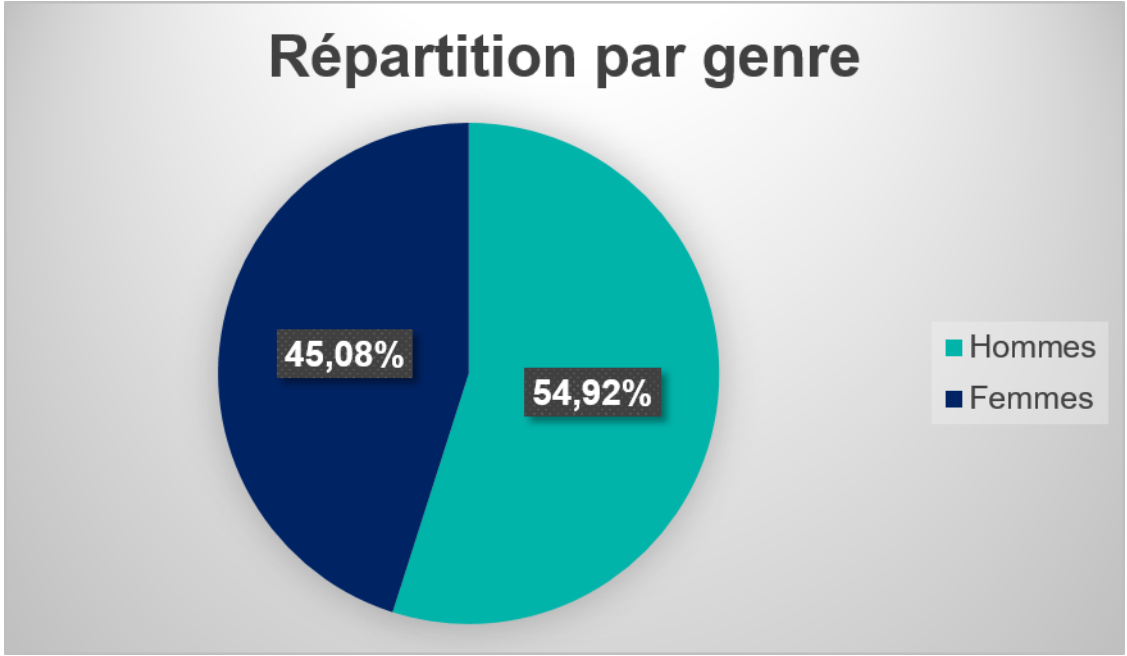


FIGURE 2.3 – Répartition genre

Pyramide des âges à l'adhésion

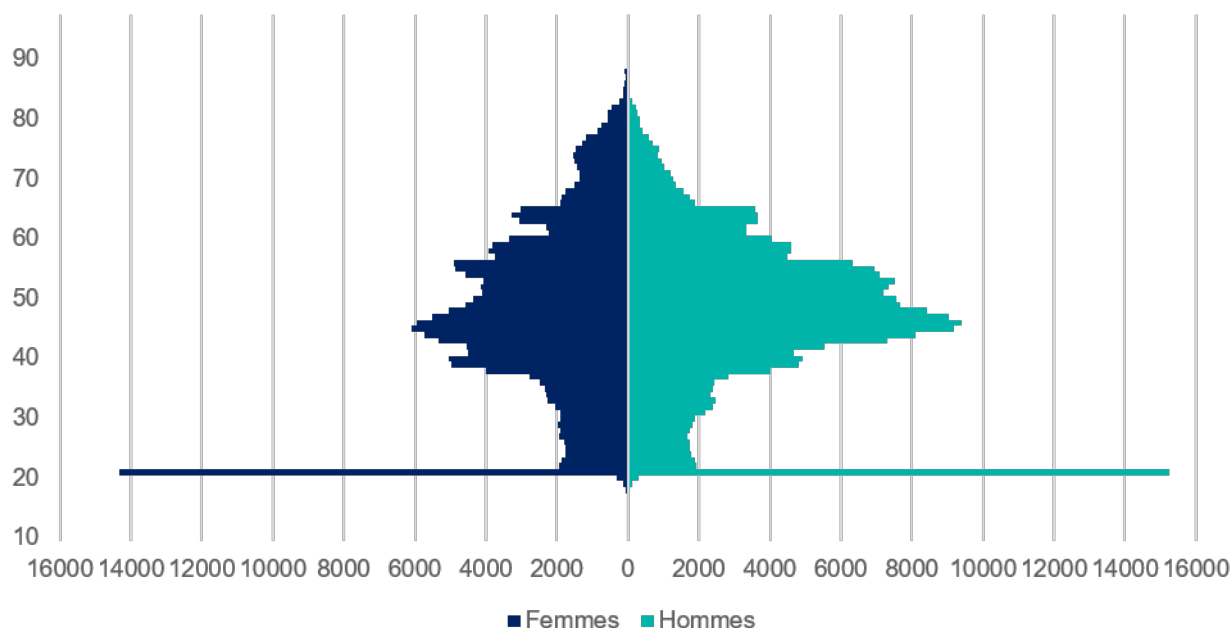


FIGURE 2.4 – Pyramide des âges

âges à la survenance de la dépendance depuis le lancement du produit.

Regardons à présent l'évolution de la survenance du risque dépendance au fil des années. L'entrée en dépendance est observée à partir de 2012, nous avons très peu de données avant. Le nombre de passages en dépendance augmente jusqu'en 2016 où la première baisse est constatée. La tendance haussière reprend en 2017 avec un pic atteint. Par rapport à 2017, moins d'entrées en dépendance sont notifiées en 2018 qui est une année incomplète.

Quant au nombre de décès en dépendance, il est en hausse depuis 2012. En effet, les dépendants qui décèdent sur une année peuvent provenir de toutes les survenances de dépendance antérieure, ce qui explique la constante augmentation. Les données étant extraites le 03/12/2018, l'année 2018 est incomplète, ce qui explique sa particularité. Aussi, l'écart entre les femmes et les hommes s'explique par la population dépendante très féminine. Les évolutions des entrées et des décès en dépendance sont représentées ci-après.

Portefeuille au 03/12/2018

Au 03/12/2018, les assurés sont à 72,78% valides et cotisants, ce qui est désigné sous "Autonomes". 20,97% des assurés sont sortis du portefeuille pour une raison autre que le décès, donc des résiliations. Les décès en état de dépendance et en autonomie notifiés représentent 5,67% alors que 0,58% de la population assurée sont en état de dépendance à cette date. Concernant les décès connus, une certaine stabilité est notée de 2015 à 2017.

Pyramide des âges des dépendants à la survenance

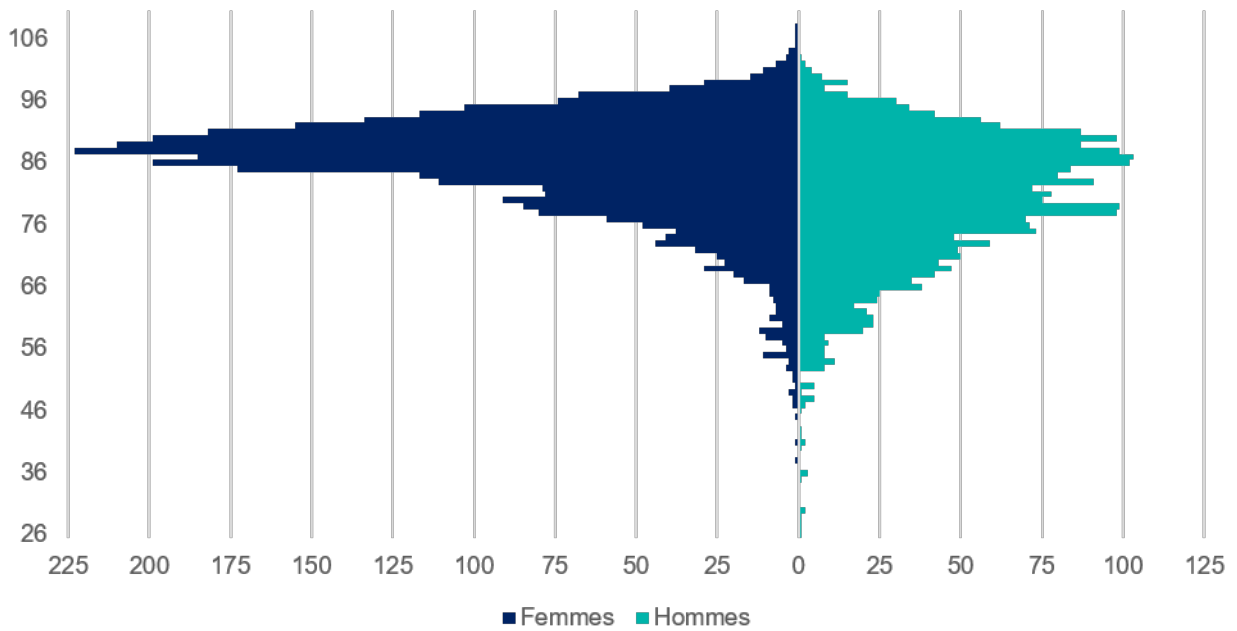


FIGURE 2.5 – Pyramide des âges à l’entrée en dépendance

PASSAGES EN DEPENDANCE PAR ANNEE DE SURVENANCE

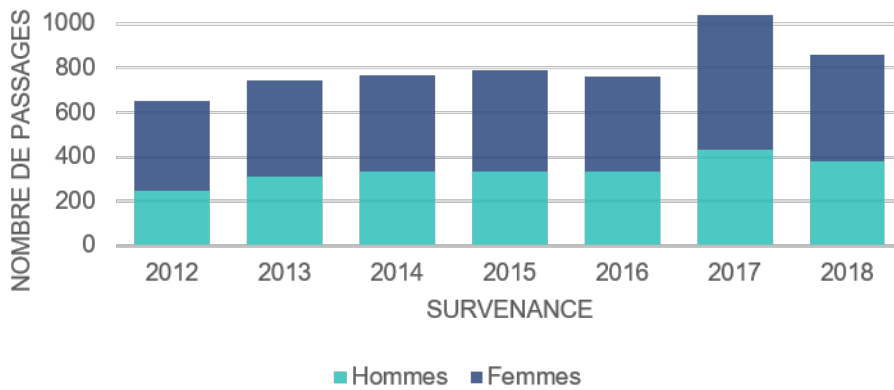


FIGURE 2.6 – Passages en dépendance

DÉCÈS DES DÉPENDANTS

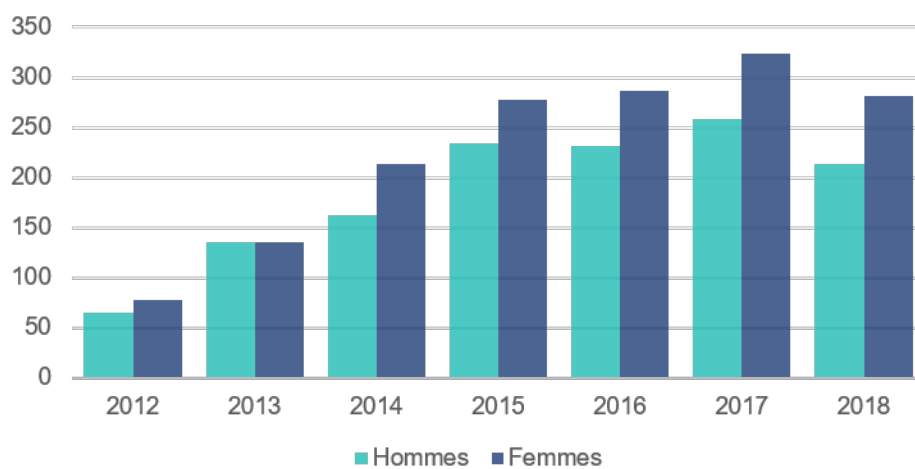


FIGURE 2.7 – Décès en dépendance

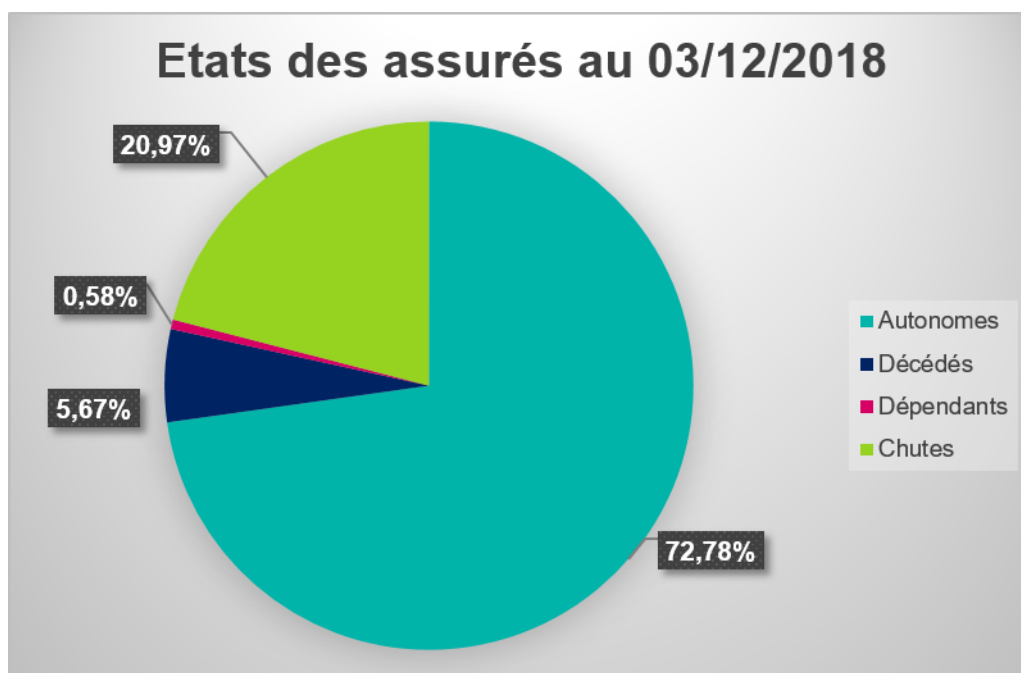


FIGURE 2.8 – État des assurés au 03/12/2018

DÉCÈS PAR SURVENANCE

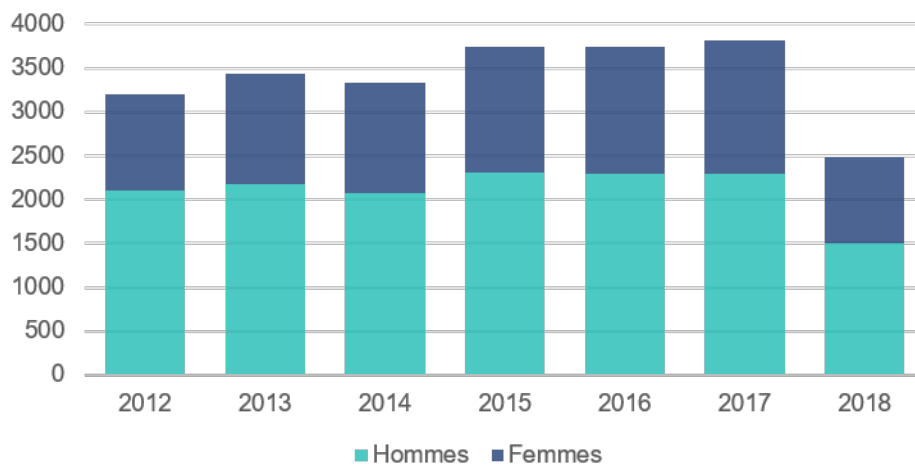


FIGURE 2.9 – Décès connus par survenance

Chapitre 3

Construction des lois d'expérience

Dans ce chapitre, seront construites des lois d'expérience concernant le portefeuille étudié : lois d'incidence dépendance, de mortalité des valides et de maintien des dépendants. Une attention particulière sera accordée au lissage et au prolongement.

Ces lois sont utiles au pilotage du portefeuille. En effet notre approche utilise un modèle multi-états : les cotisants sont susceptibles de rentrer en dépendance ou de mourir directement sans passer par cet état ; les dépendants quant à eux ne sont exposés qu'à la mortalité, la dépendance totale étant irréversible. C'est ce qui est résumé sur le graphique suivant.

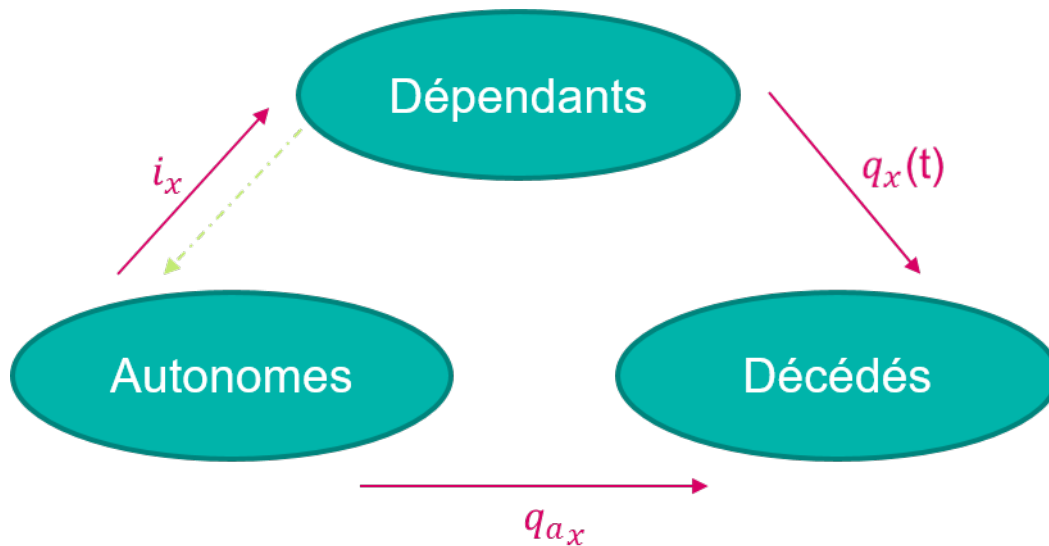


FIGURE 3.1 – Modélisation de la dépendance

3.1 Outils mathématiques

Soit T une variable aléatoire positive ou nulle, sa fonction de répartition $F(t) = \mathbb{P}[T \leq t]$ représente, pour t fixé, la probabilité de mourir avant l'instant t .

Lorsqu'elle existe, la densité de T qui représente la probabilité de mourir dans un petit

intervalle de temps après l'instant t s'écrit :

$$f(t) = \frac{d}{dt}F(t) = \lim_{h \rightarrow 0} \frac{\mathbb{P}[t \leq T < t + h]}{h}$$

En s'y prenant autrement, la fonction de survie représente, pour t fixé, la probabilité de survivre jusqu'à l'instant. Elle s'écrit :

$$S(t) = 1 - F(t) = \mathbb{P}[T > t]$$

Elle est décroissante par nature : $S(0) = 1$ et $\lim_{t \rightarrow \infty} S(t) = 0$. Après un instant $t + u$, la survie d'un individu sachant qu'il a déjà survécu jusqu'en t s'écrit :

$$S_u(t) = \mathbb{P}[T > t + u | T > t] = \frac{\mathbb{P}[T > t + u]}{\mathbb{P}[T > t]} = \frac{S(t + u)}{S(t)}$$

Pour spécifier un modèle de durée, la fonction de risque instantané, maladroitement répandue sous fonction de hasard, est très utilisée. Pour t fixé, elle est la probabilité par unité de temps conditionnée à la survie jusqu'en t , de mourir dans un petit intervalle de temps après t . Elle s'écrit :

$$h(t) = \lim_{u \rightarrow 0} \frac{\mathbb{P}[t < T \leq t + u | T \geq t]}{u} = \frac{f(t)}{S(t)} = -\frac{d}{dt} \ln(S(t))$$

La fonction de risque instantané détermine alors entièrement la loi de T . Son intégrale notée H est la fonction de hasard cumulée :

$$H(t) = \int_0^t h(s) ds$$

Toutes ces fonctions sont liées par la relation qui suit :

$$S(t) = \exp\left(-\int_0^t h(s) ds\right) = \exp(-H(t))$$

Les probabilités de décès et de survie conditionnelle seront très souvent abordées. Retrouvons leur expression en fonction de la fonction de survie. La probabilité de survie conditionnelle entre x et $x + t$ qui est la probabilité de survivre après $x + t$ ayant déjà survécu jusqu'en x s'écrit :

$${}_t p_x = \mathbb{P}[T > x + t | T > x]$$

Ainsi, la probabilité de décéder entre x et $x + t$ s'écrit :

$${}_t q_x = 1 - {}_t p_x = \mathbb{P}[T \leq x + t | T > x]$$

Soit, en utilisant la fonction de survie :

$${}_t p_x = \frac{S(x + t)}{S(x)} \quad \text{et} \quad {}_t q_x = \frac{S(x) - S(x + t)}{S(x)}$$

3.2 Lois de maintien en dépendance

Après avoir revus les outils mathématiques indispensables en analyse de survie, intéressons nous aux lois de maintien en dépendance. Elles sont utiles à l'estimation de la durée de versement des rentes et donc du montant des prestations en cas de dépendance. Les lois de maintien sont fonction de l'âge. Toutefois, l'âge sera ici décliné en deux composantes. En effet, nos lois seront des fonctions à deux dimensions, à savoir l'âge d'entrée en dépendance et la durée de maintien dans cet état. C'est qu'en pratique, une personne entrée en dépendance à 70 ans n'aura pas la même mortalité qu'une personne entrée en dépendance à 95 ans. De plus, la prise en compte de la durée de maintien en dépendance, permet de capter les effets intrinsèques de la dépendance indépendants du vieillissement naturel. Nous chercherons donc à estimer la probabilité $q_x(t)$ qu'une personne entrée en dépendance à l'âge x et maintenue en dépendance depuis t années décède l'année suivante. L'objectif est de construire une table de maintien pour des âges d'entrée en dépendance compris entre 16 et 106 ans et une ancienneté allant jusqu'à 90 ans.

Nous allons d'abord estimer les taux bruts de décès sur les cinq premières années de maintien avant de procéder au prolongement à l'aide des tables TH00-02 et TF00-02. Les différentes méthodes utilisées seront également comparées. Afin de prendre plus de recul, explorons les statistiques de notre population de rentiers.

3.2.1 Statistiques descriptives sur les rentiers

Les premiers sinistres sont survenus en 2006, mais jusqu'au 31/12/2011, seuls 40 sinistres sont enregistrés dont 37 en cours. C'est à partir de 2012, soit 10 ans après le lancement du produit, que les sinistres sont régulièrement observés.

Survenance	Présents au 01/01/N	Nouveaux Sinistres
2009	9	1
2010	9	7
2011	15	22
2012	37	826
2013	719	768
2014	1215	773
2015	1606	789
2016	1881	819
2017	2143	1016
2018	2455	628

TABLE 3.1 – Évolution des dépendants

Il est à noter que les nouveaux sinistres ont connu une hausse subite en 2012 suivie d'une légère baisse l'année suivante. Puis, jusqu'en 2016 les entées en dépendance ont sensiblement augmenté avant de noter une importante hausse en 2017. Cependant, la survénance 2018 n'est pas complète, d'où la baisse des nouveaux sinistres enregistrés. Regardons à présent la répartition et l'état de ces sinistres.

	Dépendants	Décédés	Pourcentage de décès
Hommes	2085	1144	54,9%
Femmes	2905	1395	48,0%
Total	4990	2539	50,9%

TABLE 3.2 – Répartitions des sinistres et décès par sexe

Qu'en est-il de la durée de maintien ? La réponse peut être déclinée en s'intéressant aux âges d'entrée et de sortie de l'état de dépendance, ou directement à la durée passée en dépendance.

	Âge à la survenance	Âge à la sortie
Moyenne	79.8	82.7
Minimum	27	36
Maximum	103	104

TABLE 3.3 – Hommes : âges de dépendance

	Âge à la survenance	Âge à la sortie
Moyenne	85.5	88.2
Minimum	38	50
Maximum	108	108

TABLE 3.4 – Femmes : âges de dépendance

Nous remarquons que les femmes, qui représentent 58,22% des personnes dépendantes, rentrent en dépendance plus tardivement, en moyenne 6 ans après la survenance chez les hommes. L'âge à la sortie ne prend en compte que les sinistres clôturés au 31/12/2017. Sur cette base, la durée moyenne de maintien serait de 3 ans environ pour les deux genres. Les pyramides de maintien fournissent plus de détails à ce propos.

D'abord, une grande partie des assurés décèdent la première année de dépendance. Cela correspond à 43% des hommes et 37% des femmes pour l'ensemble des dépendants. Ces proportions atteignent respectivement 49% et 42% pour les sinistres clôturés. Lesquels présentent une durée de maintien maximum n'excédant pas 5 ans avec des quantités pas assez significatives. Cela veut dire que la construction ne sera pas robuste au-delà de 6 ans de maintien et reconforte l'idée d'estimation des taux bruts sur les cinq premières années pour ensuite faire un prolongement.

Les notions de censure et troncature ne seront pas revues en détails mais il n'est pas insensé de souligner que le surplus entre les deux pyramides de maintien est constitué de censures à droite. Elles sont très prononcées sur la 1ère année de maintien, ce qui suscite une attention particulière pour les survenances de 2017 étant donné que l'extraction est faite en 2018.

Durée de maintien des dépendants

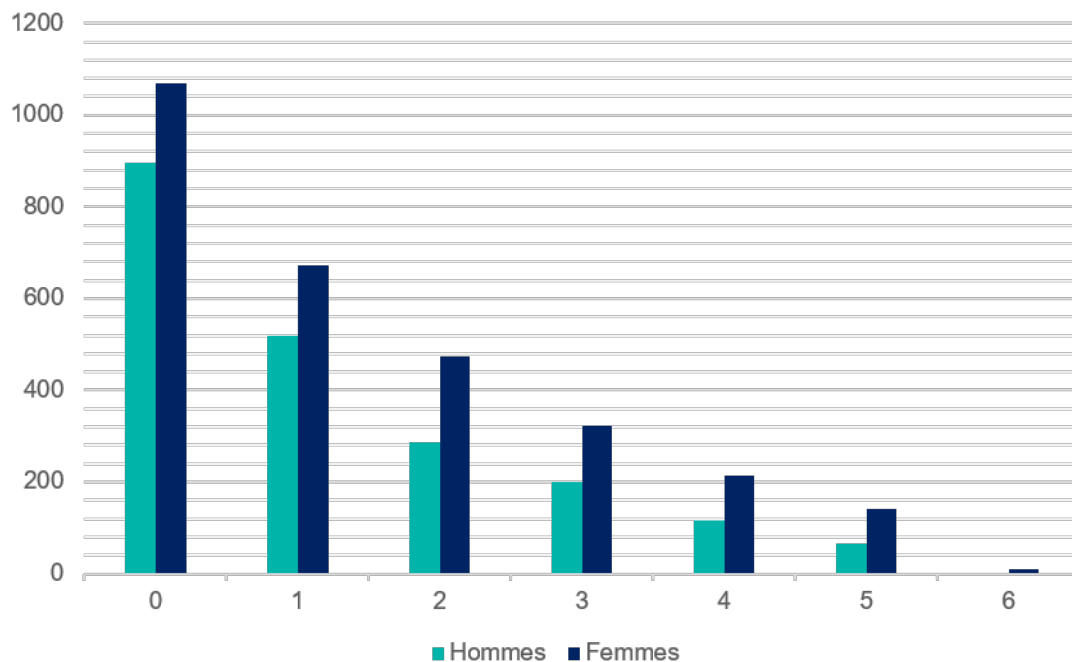


FIGURE 3.2 – Durée de maintien en dépendance

Durée de maintien des dépendants décédés

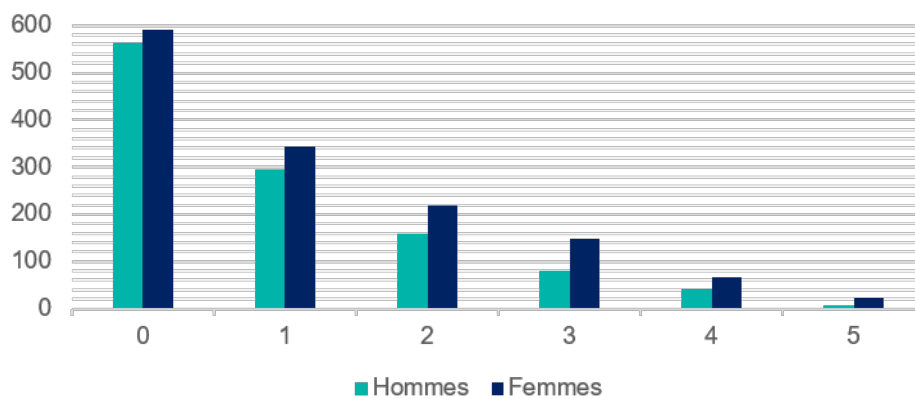


FIGURE 3.3 – Durée de maintien en dépendance des sinistres clos

La loi de maintien a été construite dans un premier temps selon la méthode de l'estimateur actuariel puis de l'estimateur de Kaplan-Meier. Ces deux méthodes vont être présentées ainsi que leurs résultats.

3.2.2 Estimateur actuariel

Cet estimateur est généralement utilisé quand les dates exactes de décès et de fin d'observation en cas de censures ne sont pas disponibles. Dans ce cas, les censures sont supposées survenir en moyenne en milieu d'année.

Soient :

- d_i le nombre de réalisations du risque au temps t_i ;
- c_i le nombre de censures dans l'intervalle $[t_i; t_{i+1}$;
- n_i le nombre d'assurés réellement exposés au risque en t_i

Ainsi nous avons :

$$n_i = n_{i-1} - \frac{c_i}{2}$$

et

$$\text{maintien}_i = \text{maintien}_{i-1} * \left(1 - \frac{d_i}{n_i}\right)$$

Cependant, le décès étant la seule cause de sortie de l'état de dépendance, nos observations ne comporte pas de censures à l'exception des censures à droites qui correspondent à la non réalisation du risque de décès à la date d'extraction.

3.2.3 Estimateur de Kaplan-Meier

Dans le cadre de l'estimation de la durée de maintien en dépendance, le principe de l'estimateur de Kaplan-Meier est le suivant : être maintenu en dépendance après un instant t , c'est l'être juste avant t et ne pas sortir de cet état à l'instant t .

Soient :

- d_i le nombre de réalisations du risque au temps t_i ;
- c_i le nombre de censures dans l'intervalle $[t_i; t_{i+1}$;
- n_i le nombre d'assurés exposés au risque en t_i

La fonction de survie S se définit comme suit :

$$S(t) = \mathbb{P}(T > t)$$

En utilisant les probabilités conditionnelles :

$$\begin{aligned} S(t_i) &= \mathbb{P}(T > t_i) \\ &= \mathbb{P}(T > t_i \cap T > t_{i-1}) \\ &= \mathbb{P}(T > t_i | T > t_{i-1}) \times \mathbb{P}(T > t_{i-1}) \\ &= S_{i-1}(t_i) \times S(t_i) \end{aligned}$$

Maintien en dépendance par estimateur actuariel

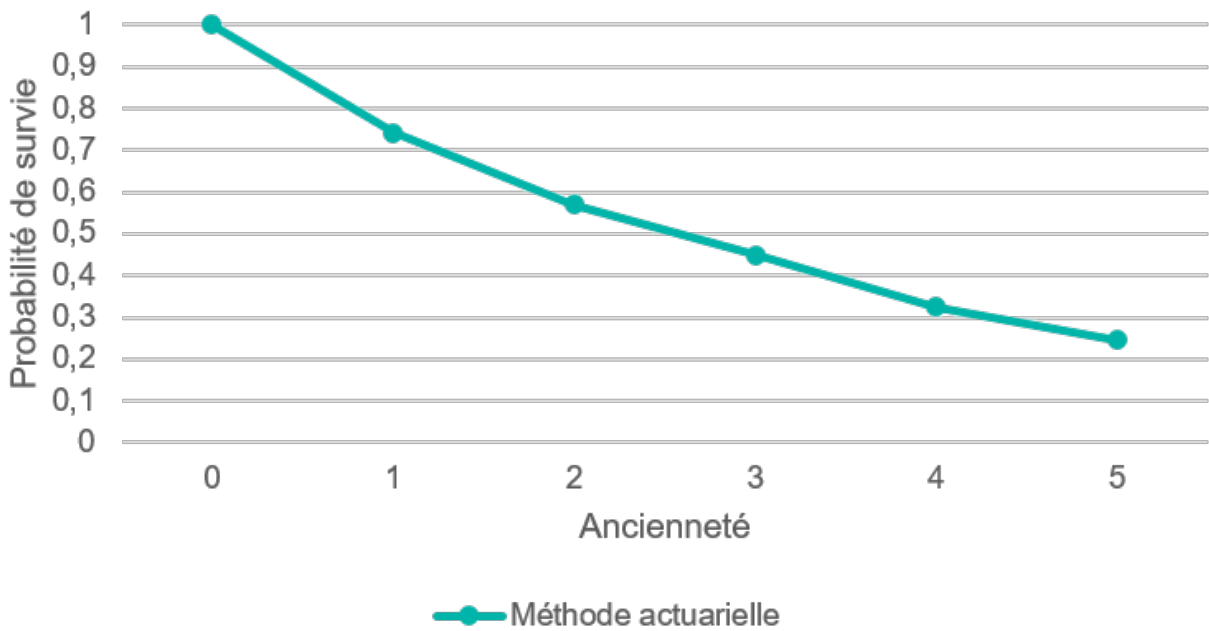


FIGURE 3.4 – Loi de maintien en dépendance par estimateur actuariel

Comme $S_{t_{i-1}}(t_i)$ peut être estimé par $1 - \frac{d_i}{n_i}$, $S(t_i)$ peut alors être estimé par $\hat{S}(t_i)$:

$$\hat{S}(t_i) = \left(1 - \frac{d_i}{n_i}\right) \times S(t_{i-1})$$

En faisant une récurrence de proche en proche et en notant que $\hat{S}(0) = 1$, l'estimateur de Kaplan-Meier de la fonction de survie s'écrit finalement :

$$\begin{aligned} \hat{S}_{KM}(t) &= \prod_{i|t_i \leq t} \hat{S}_{t_{i-1}} \\ &= \prod_{i|t_i \leq t} \left(1 - \frac{d_i}{n_i}\right) \end{aligned}$$

Un estimateur de la variance de l'estimateur de Kaplan-Meier nous est donné par la formule de Greenwood :

$$Var\left(\hat{S}_{KM}(t)\right) = \hat{S}_{KM}(t)^2 \cdot \sum_{i|t_i \leq t} \frac{d_i}{(n_i - d_i) \cdot n_i}$$

Par ailleurs, $\frac{\sqrt{n}(\hat{S}_{KM}(t) - S_{KM}(t))}{S_{KM}(t)}$ converge vers un processus gaussien centré. D'où $\hat{S}_{KM}(t)$

converge vers un processus gaussien centré sur $S_{KM}(t)$. Nous en déduisons un intervalle de confiance asymptotique de niveau $1-\alpha$:

$$IC_{\alpha}^A = \left[\hat{S}_{KM}(t) - q_{\alpha/2}^{N(0,1)} \sqrt{Var(\hat{S}_{KM}(t))}, \hat{S}_{KM}(t) + q_{\alpha/2}^{N(0,1)} \sqrt{Var(\hat{S}_{KM}(t))} \right]$$

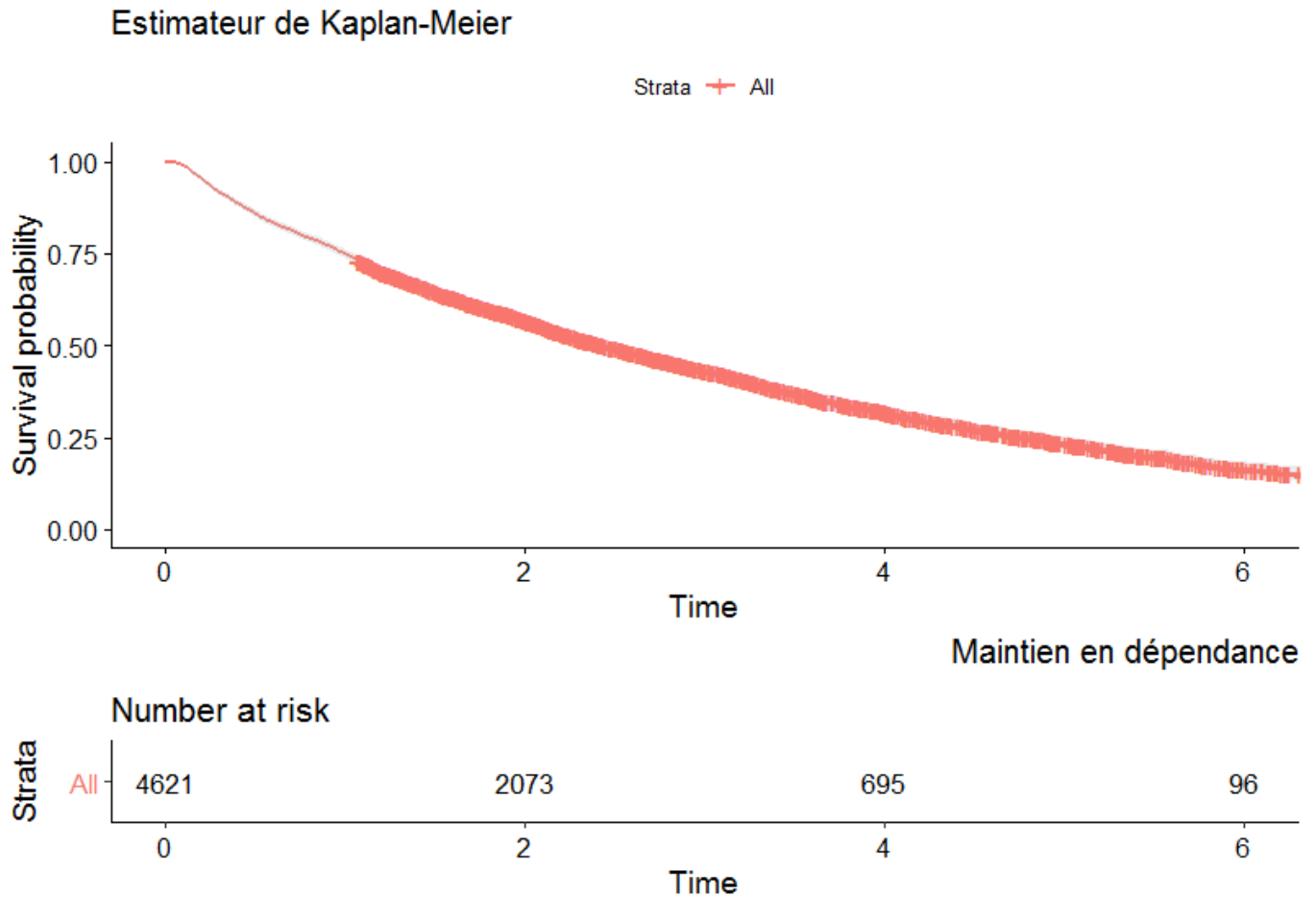


FIGURE 3.5 – Fonction de survie du maintien en dépendance par l’estimateur Kaplan-Meier

La comparaison avec l’estimateur actuariel ne révèle pas de différence notable, l’hypothèse d’entrée en dépendance en milieu d’année est globalement acceptable.

Une loi prenant en compte la franchise a été également construite, la comparaison ne révèle qu’une légère différence.

Nous avons également regarder l’influence du sexe sur le maintien en dépendance. Les femmes survivent plus longtemps en dépendance que les hommes.

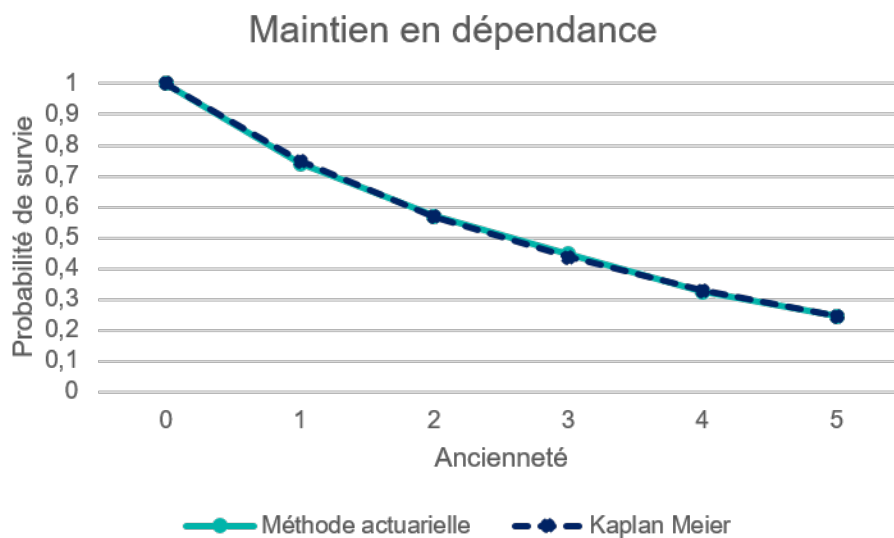


FIGURE 3.6 – Comparaison estimateurs actuariel et Kaplan-Meier

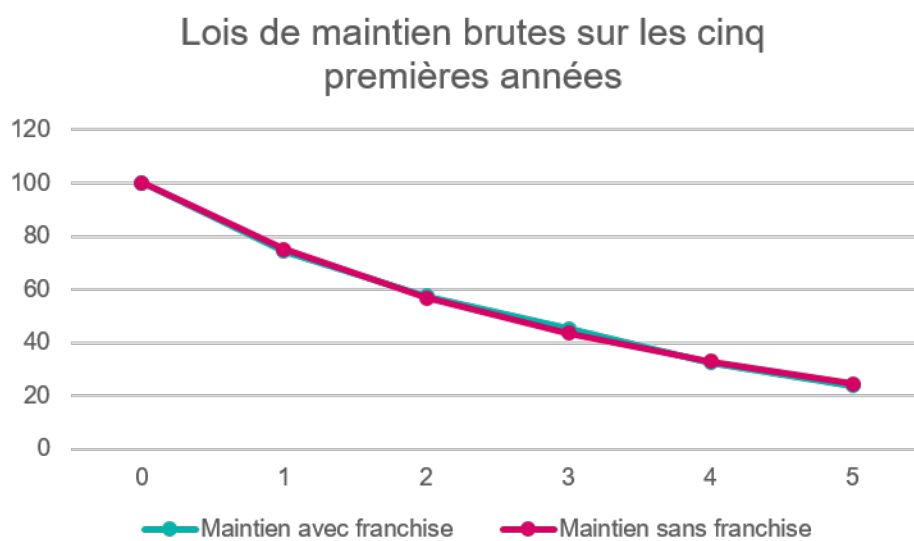


FIGURE 3.7 – Impact de la franchise sur le maintien en dépendance

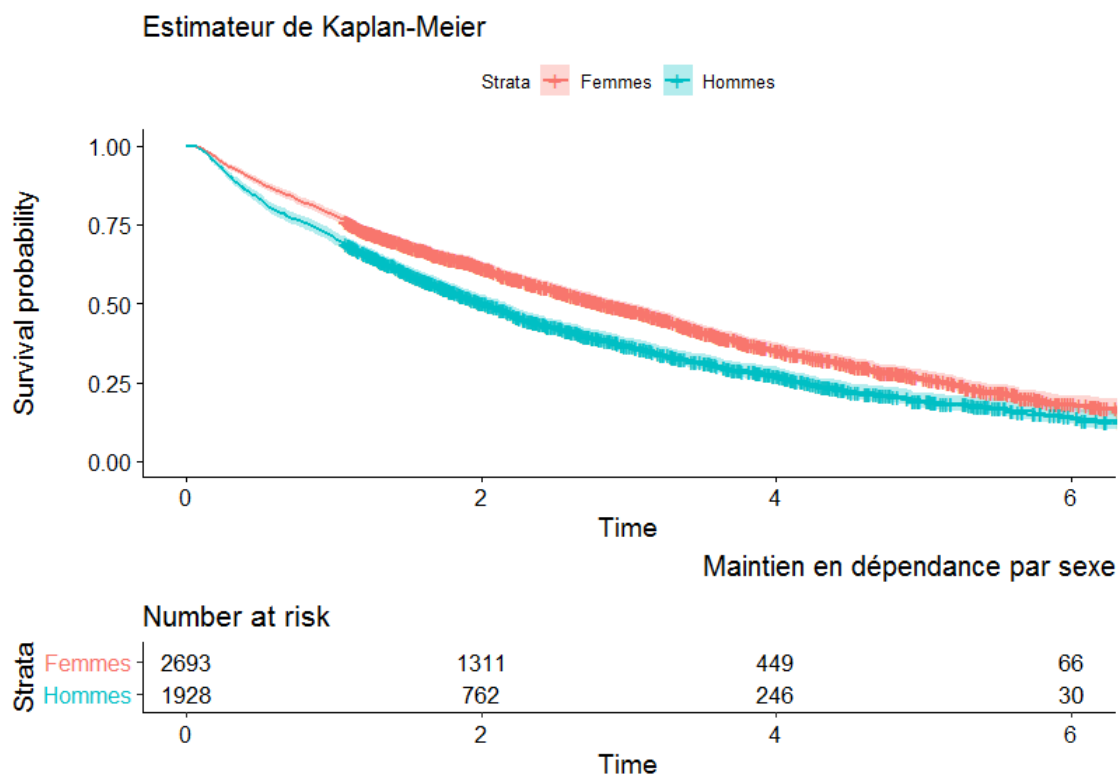


FIGURE 3.8 – Comparaison lois de maintien en dépendance par sexe

3.2.4 Modèle de Cox

Le modèle à risque proportionnel de Cox est un modèle semi-paramétrique qui suppose un effet multiplicatif des variables explicatives sur la fonction de risque instantanée de base. Son objectif est d'évaluer cet impact. L'intérêt majeur est la quantification de l'influence des variables explicatives sans avoir à estimer le risque absolu.

Principe

Soient :

- $X=(X_1,\dots,X_p)$ un vecteur de variables aléatoires binaires
- h_0 la fonction de risque instantané de base
- $\beta=(\beta_1,\dots,\beta_p)$ un vecteur de coefficients réels

Le risque instantané de survenance de l'évènement est exprimé en fonction de l'instant t et des variables explicatives X_j . Il traduit la probabilité de réalisation de l'évènement à l'instant t sachant qu'il n'est pas encore survenu jusqu'à cet instant et s'écrit :

$$h(t, X) = h_0(t) \times \exp(\beta' X)$$

Les paramètres β_j représente l'effet des variables explicatives X_j sur le risque instantané de base h_0 .

Hypothèses

Le modèle de Cox repose sur des hypothèses précises :

- Il existe une relation log-linéaire entre la fonction de risque instantané et les variables explicatives :

$$\log \left[\frac{h(t, X)}{h_0(t)} \right] = \beta' X$$

- Le risque instantané de survenance de l'évènement est multiplié par une constante quand on change la valeur d'une variable. Autrement dit, le rapport entre deux fonctions de risque instantané pour deux individus i_1 et i_2 n'ayant qu'une variable qui diffère entre eux, est proportionnel et ne dépend que de X_{i_1} et X_{i_2} qui sont indépendants du temps :

$$\frac{h(t, X_{i_1})}{h(t, X_{i_2})} = \frac{h_0(t)\exp(\beta' X_{i_1})}{h_0(t)\exp(\beta' X_{i_2})} = \exp(\beta'(X_{i_1} - X_{i_2}))$$

Estimation

Soient

- D le nombre de décès observés parmi les n individus à l'étude
- $T_1 < T_2 < \dots < T_D$, les instants de décès distincts,
- $(1), (2), (3), \dots, (D)$, les indices des individus décédés respectivement en T_1, T_2, \dots, T_D ,
- X_i , la valeur des covariables de l'individu i ,
- R_i , l'ensemble des individus vivants et toujours exposés au risque juste avant T_i

L'idée de Cox est qu'aucune information ne peut être donnée sur β par les intervalles pendant lesquels aucun évènement n'a eu lieu, car on peut concevoir que h_0 soit nulle dans ces intervalles. De même, les moments où se produisent les censures sont supposés n'apporter

que peu ou pas d'information sur β . On travaille alors conditionnellement à l'ensemble des instants où un décès a lieu.

La probabilité que l'individu i décède en T_i sachant qu'il y a un décès à cet instant s'écrit :

$$\frac{h_0(T_i) \exp(\beta' X_{(i)})}{\sum_{j \in R_i} h_0(T_j) \exp(\beta' X_j)} = \frac{\exp(\beta' X_{(i)})}{\sum_{j \in R_i} \exp(\beta' X_j)}$$

Cette probabilité dépend uniquement du paramètre β , elle est indépendante de la fonction de risque instantanée. Avec des contributions à la vraisemblance à chacun des temps de décès distincts, la vraisemblance de Cox est donc le produit sur les temps de décès :

$$L_{Cox}(\beta) = \prod_{i=1}^D \frac{\exp(\beta' X_i)}{\sum_{j \in R_i} \exp(\beta' X_j)}$$

La vraisemblance partielle ne dépend pas de la fonction de risque instantanée h_0 , β peut être estimé par maximisation de la vraisemblance partielle, sans connaître la fonction de risque instantané de base. Une estimation du vecteur de paramètres β est alors obtenue en annulant les dérivées premières de la log-vraisemblance. La fonction *coxph* du logiciel R sera utilisée pour ces calculs.

Une fois les paramètres β estimés, la relation suivante permet de retrouver la fonction de survie pour un vecteur de covariable X :

$$S(t, X) = \exp\left(-\int_0^t h(u, X) du\right)$$

Ou alors, en utilisant la fonction de risque cumulée :

$$S(t, X) = \exp\left(-H_0(t) \exp(\beta' X)\right)$$

Une approximation de la fonction de risque cumulée est donné par l'estimateur de Breslow :

$$\hat{H}_0(t) = \sum_{i: T_i \leq t} \frac{d_i}{\sum_{j \in R_i} \exp(\beta' X)}$$

Ce qui conduit à l'estimateur suivant de la fonction de survie pour un vecteur de covariable X :

$$\hat{S}(t, X) = \exp\left(-\hat{H}_0(t) \exp(\beta' X)\right)$$

Ce raisonnement suppose des temps d'évènements distincts, ce qui est rarement le cas avec des données réelles où le pas de temps est en général mensuel ou annuel.

Effet d'une covariable

Le modèle de Cox permet d'isoler l'influence des variables sur la loi de maintien et leur corrélation.

Soient deux individus i et j aux caractéristiques identiques sauf pour la k^{e} variable ($\forall u \neq k, X_{iu} = X_{ju} = X_u, X_{ik} = 1$ et $X_{jk} = 0$) et dont les fonctions de risque instantané sont respectivement $h_i(t)$ et $h_j(t)$, l'effet de la k^{e} variable s'évalue en faisant le rapport des fonctions :

$$\frac{h_i(t)}{h_j(t)} = \frac{h_0(t) \exp(\beta_1 * X_1 + \beta_2 * X_2 + \dots + \beta_k * 1 + \dots + \beta_p * X_p)}{h_0(t) \exp(\beta_1 * X_1 + \beta_2 * X_2 + \dots + \beta_k * 0 + \dots + \beta_p * X_p)} = \exp(\beta_k)$$

D'où,

$$S_i(t) = S_j(t)^{-\exp(\beta_k)}$$

Quant à la corrélation entre les individus i et j précédemment définis, il suffit de regarder le signe de β_k :

- Si $\beta_k > 0$, alors $\exp(\beta_k) > 1$: le risque est plus élevé pour l'individu i que pour l'individu j . La covariable $X_k = 1$ augmente le risque.
- Si $\beta_k = 0$, alors $\exp(\beta_k) = 1$: le risque est identique pour les individus i et j . La covariable X_k n'a pas d'effet sur le risque.
- Si $\beta_k < 0$, alors $\exp(\beta_k) < 1$: le risque est plus faible pour l'individu i que pour l'individu j . La covariable $X_k = 1$ atténue le risque.

Application

La significativité de la variable "sexe" sur la loi de maintien a été testée à l'aide du modèle de Cox. Voici les résultats :

```
call:
coxph(formula = surv(time = (AgeFinJours_KM - AgeDebJours_KM)/365.25,
  Maintien_dep$Censure) ~ Homme, data = Maintien_dep)

n= 4621, number of events= 2415

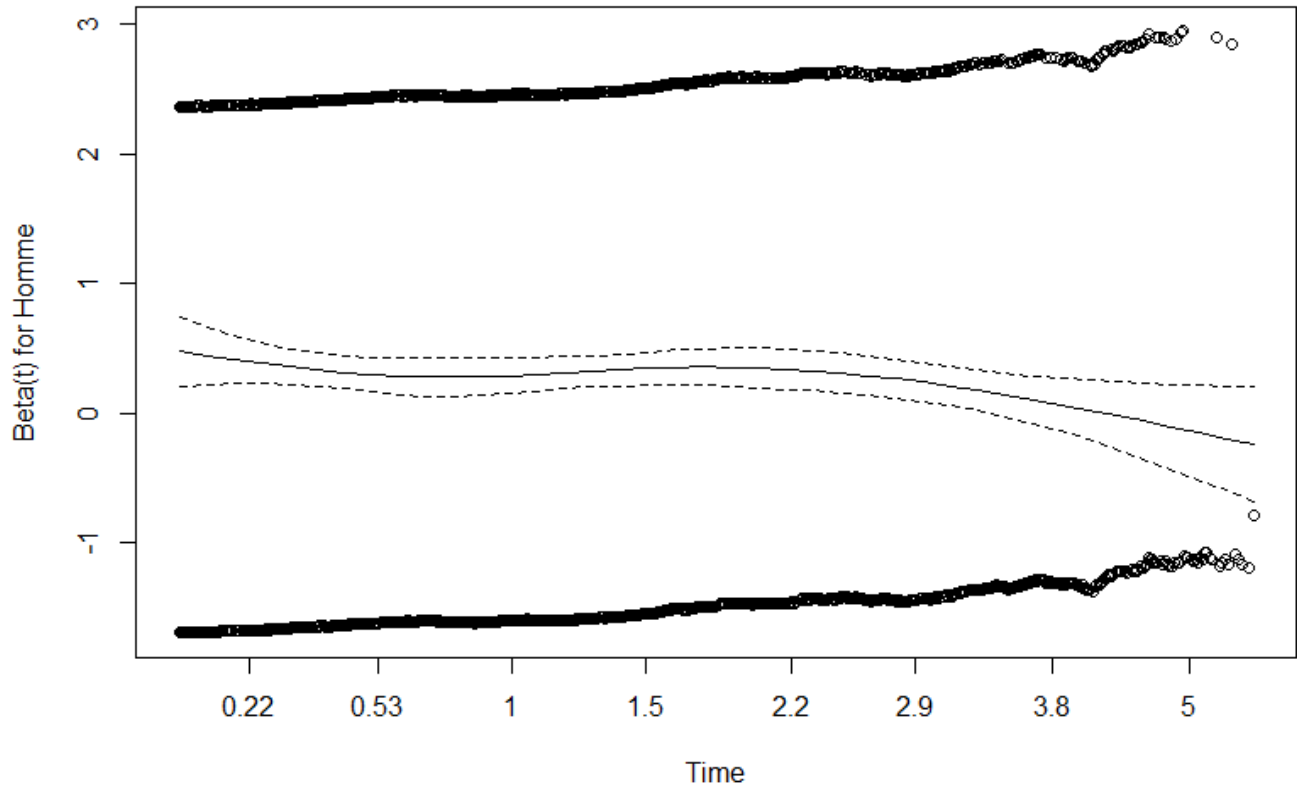
      coef exp(coef) se(coef)      z Pr(>|z|)
Homme 0.29712    1.34597  0.04096  7.254 4.05e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

      exp(coef) exp(-coef) lower .95 upper .95
Homme      1.346      0.743    1.242    1.458

Concordance= 0.541 (se = 0.006 )
Likelihood ratio test= 51.95 on 1 df,  p=6e-13
Wald test               = 52.62 on 1 df,  p=4e-13
score (logrank) test = 53 on 1 df,  p=3e-13
```

Le coefficient β de la variable "sexe" (être un homme ou pas), est de 0.297 avec une p -value de $4.05e^{-13}$ significativement faible permettant de rejeter l'hypothèse de nullité du coefficient. Ce résultat est aussi confirmé par les tests de rapport des maxima de vraisemblance, de Wald et du logrank. Le sexe est donc une variable discriminante pour la loi de maintien en dépendance.

L'évolution de la valeur du coefficient β est représenté ci-dessous. Il est positif au début puis négatif à partir de la fin de la 3^e année de maintien. En d'autres termes, la mortalité des hommes est supérieure à celle des femmes les premières années de maintien en dépendance et la tendance s'inverse à partir de la fin de la 3ⁱème année.



Étant donné que β ne peut être supposé constant dans le temps, le modèle de Cox ne permet pas d'estimer les lois de maintien par sexe. Ce qui est confirmé par le test des résidus de Schoenfeld. En effet, la $p - value$ obtenue égale à 0.043 permet de rejeter l'hypothèse d'indépendance temporelle au seuil de 5%.

L'objectif de base étant de construire des lois par âge d'entrée en dépendance et par durée de maintien, l'impact de l'âge à la survenance sera étudié. Par souci de robustesse et pour des raisons d'ordre pratique, les dépendants sont regroupés par classes d'âges d'entrée à la dépendance. Les classes d'âges ont été choisies en veillant à avoir des effectifs comparables. Ci-après les résultats de l'estimation de la fonction de survie par Kaplan-Meier sur chaque classe d'âges. La fonction de survie des moins de 80 ans est supérieure aux deux autres fonctions de survie, l'écart se creuse avec la durée de maintien. Cependant, bien qu'étant supérieure à celle des plus de 89 ans, la fonction de survie des assurés entrés en dépendance entre 80 et 89 ans lui est très proche.

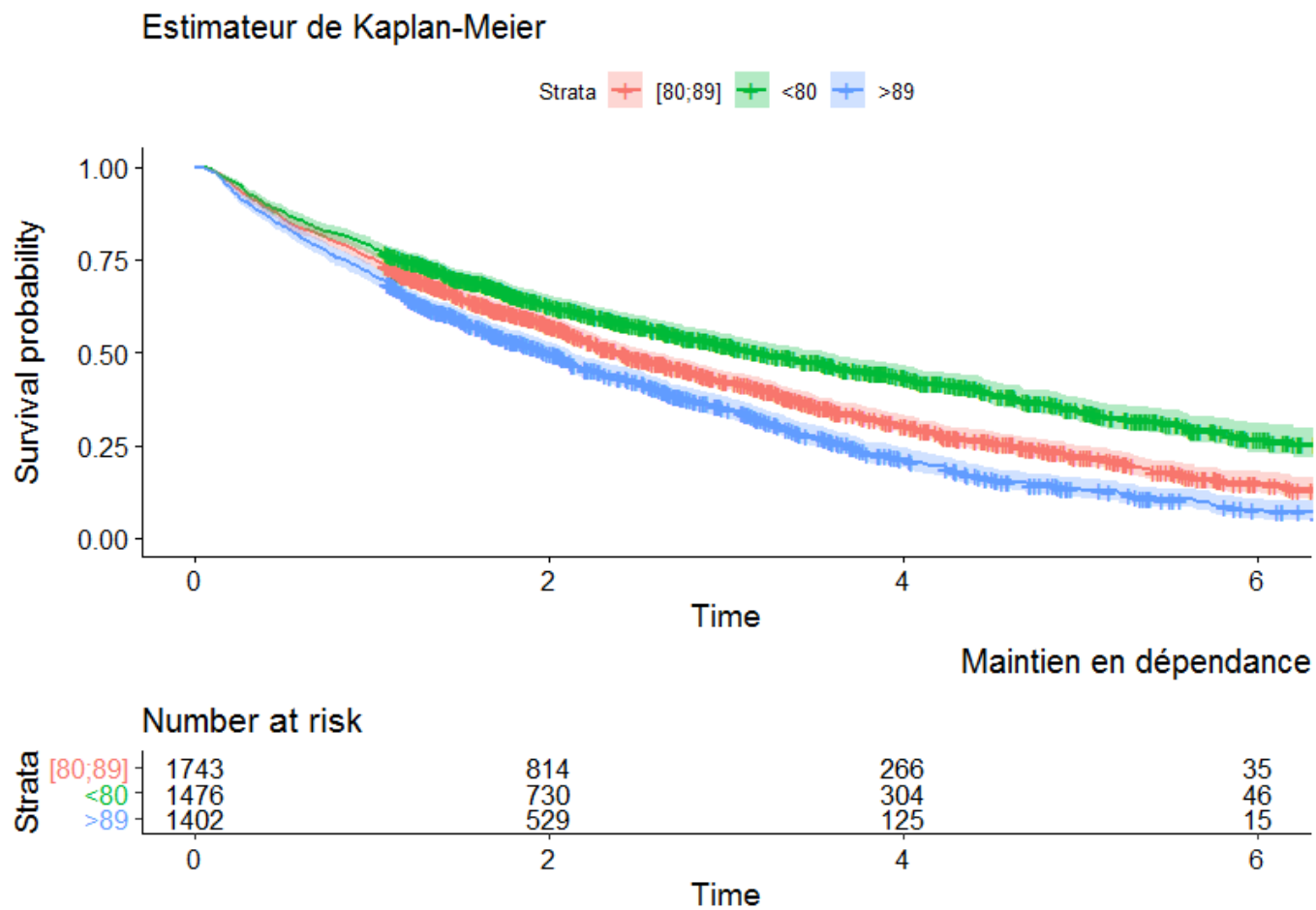


FIGURE 3.9 – Lois de maintien en dépendance par classe d'âge

Le modèle de Cox a été utilisé pour tester l'influence de la classe d'âge sur le maintien en dépendance. La classe d'âges d'entrée en dépendance [80;89] est prise pour référence. Pour rappel, l'âge moyen d'entrée en dépendance est 83 ans. Les résultats montrent un impact significatif de la classe d'âges d'entrée en dépendance. En effet, les personnes entrées en dépendance avant 80 ans ont un risque de mortalité global moins élevé ($\beta_1 = -0.31 < 0$) et celles entrées en dépendance après 89 ans ont un risque de décès en général plus élevé ($\beta_2 = 0.17 > 0$). En résumé, le maintien en dépendance diminue globalement avec la classe d'âges d'entrée en dépendance.

```
Call:
coxph(formula = surv(time = (AgeFinJours_KM - AgeDebJours_KM)/365.25,
  Maintien_dep$Censure) ~ Age_entree_dep, data = Maintien_dep)

n= 4621, number of events= 2415

              coef exp(coef) se(coef)      z Pr(>|z|)
Age_entree_depMoins de 80 -0.31280   0.73140  0.05089 -6.147 7.89e-10 ***
Age_entree_depPlus de 89  0.17327   1.18919  0.04780  3.625 0.000289 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

              exp(coef) exp(-coef) lower .95 upper .95
Age_entree_depMoins de 80   0.7314   1.3672   0.662   0.8081
Age_entree_depPlus de 89   1.1892   0.8409   1.083   1.3060

Concordance= 0.543 (se = 0.006 )
Likelihood ratio test= 87.17 on 2 df,  p=<2e-16
Wald test               = 84.55 on 2 df,  p=<2e-16
Score (logrank) test = 85.74 on 2 df,  p=<2e-16
```

De ces différentes analyses, il ressort que le sexe et la classe d'âges d'entrée en dépendance sont des variables discriminantes, la loi de maintien en dépendance devra être déclinée en deux lois :

- une loi de maintien pour les femmes en fonction de la classe d'âges d'entrée et de l'ancienneté en dépendance
- une loi de maintien pour les hommes en fonction de la classe d'âges d'entrée et de l'ancienneté en dépendance

Étant donné que l'effet des covariables discriminantes dépend du temps, chacune de ces lois sera estimée par la méthode Kaplan-Meier. Il est aussi possible de poursuivre avec Cox en modélisant les données avec un effet dépendant du temps pour ces variables. voir «Extensions du modèle de Cox : Variables dépendantes du temps, Effets non linéaires» du Professeur Roch GIORGI de la Faculté de Médecine d'Aix-Marseille Université.

Les différentes fonctions de survie ont été représentées ci-dessous

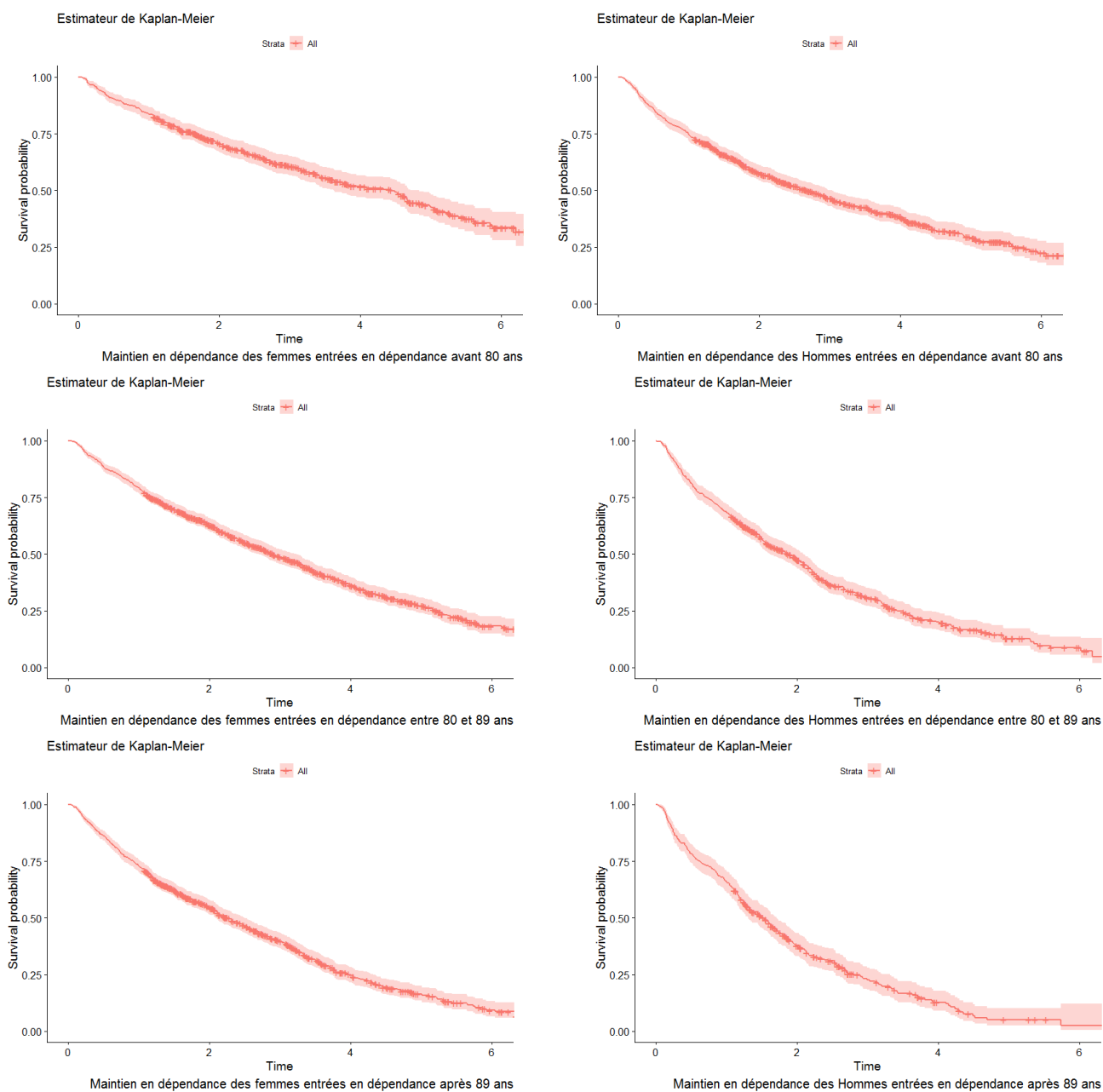


FIGURE 3.10 – Lois de maintien en dépendance par sexe et classe d'âges

En raison du manque de données à partir de la cinquième année de dépendance, ces lois ont été retenues pour les 4 premières années de dépendances puis ont été prolongées en appliquant un taux de majoration aux tables réglementaires TH00-02 et TF00-02. Le choix des coefficients est basé sur la relation de proportionnalité moyenne entre les taux de décès des lois d'expérience et ceux des tables réglementaires en regardant les 3ème et 4ème année de maintien en dépendance.

Cela revient à déterminer, pour chacune des lois d'expériences construites, un α_C tel que :

$$q_C = \alpha_C q_C^R$$

Avec :

- C un individu fictif d'âge l'âge moyen x_C de la classe de la loi à prolonger et de sexe le sexe des individus de la classe en question
- R la table réglementaire correspondante à la classe en question : TH00-02 pour les hommes et TF00-02 pour les femmes.
- q_C la probabilité pour un individu C dépendant d'âge x_C de décéder dans l'année.
- q_C^R la probabilité, issue de la table réglementaire R pour un individu valide C d'âge x_C de décéder dans l'année.

En utilisant les probabilités obtenues par l'estimation de Kaplan-Meier sur chaque classe, nous obtenons les valeurs de α_C suivants :

	Moins de 80 ans	80 - 89 ans	Plus de 89 ans
Hommes	3,15	2,12	1,53
Femmes	4,33	2,37	1,58

Plus l'âge d'entrée en dépendance est élevé, plus la survie de l'assuré dépendant se rapproche de la survie de la population générale. Cela se traduit par un coefficient qui diminue avec l'âge et se rapproche de 1.

3.3 Lois d'incidence dépendance

Il existe deux approches pour modéliser la dépendance. La première propose une vision en flux (taux d'incidence) et la seconde une vision en stock (taux de prévalence). Le taux d'incidence mesure pour chaque âge donné, le ratio de nouvelles entrées en dépendances par rapport à la population totale exposée ayant cet âge. Le taux de prévalence quant à lui, mesure le ratio de personnes dépendantes (peu importe leur ancienneté dans cet état) ayant un âge donné par rapport à la population totale exposée ayant cet âge.

La première approche est celle qui sera utilisée dans le cadre de ce mémoire.

Notons :

- e_x^c l'exposition centrale observée entre les âges x et $x + 1$
- n_x^2 le nombre d'entrées en dépendance observées entre les âges x et $x + 1$

Il convient de noter que l'exposition centrale correspond ici au temps total passé par l'ensemble des individus dans l'état d'autonomie. Les individus dépendants ne sont en effet plus soumis au risque.

La base à notre disposition contient :

- La date de naissance notée DN ,
- La date d'adhésion notée DA ,
- La date de sortie notée DS ,
- La date d'entrée en dépendance notée DE ,
- La date de décès notée DC , si elle est connue,
- Le motif de sortie notée MS

En restreignant notre intervalle d'étude entre les dates Deb et Fin , nous avons pour les n_c cotisants :

- L'âge de début d'observation x_i :

$$x_i = \max(DA_i, Deb) - DN_i$$

- L'âge de fin d'observation y_i :

$$y_i = \min(DS_i, DE_i, Fin) - DN_i$$

À partir de là, nous calculons :

•

$$e_x^c = \sum_{i=1}^{n_c} \max[0, \min(y_i, x + 1) - \max(x_i, x)]$$

•

$$n_x^2 = \sum_{i=1}^{n_c} \mathbf{1}_{\{DE < Fin\}} \mathbf{1}_{[x, x+1]}(y_i)$$

Le taux brut d'incidence en dépendance à l'âge x est alors estimé par :

$$\hat{i}_x = \frac{n_x^2}{e_x^c}$$

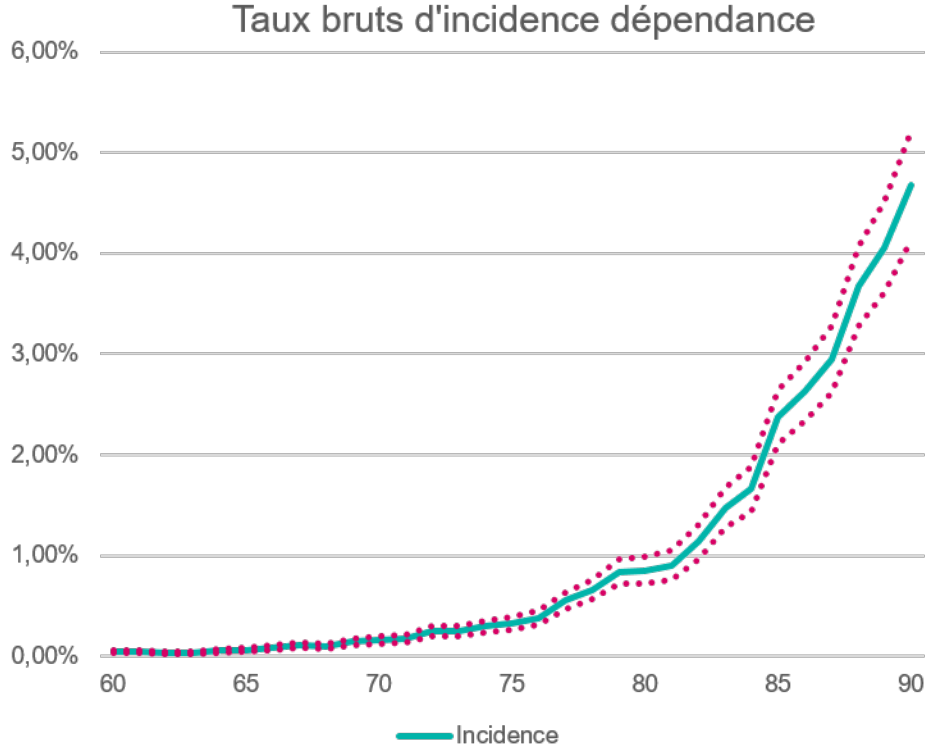


FIGURE 3.11 – Taux d'incidence bruts

Ceci est une version de l'estimateur de Hoem qui est un estimateur sans biais du taux d'incidence i .

Soit n_x , le nombre de cotisants valides d'âge x , le nombre d'entrées en dépendance est estimé par $n_x * \hat{i}_x$. Il correspond à une somme de n_c variables aléatoires indépendantes suivant une loi de Bernoulli $B(i_x)$. Il s'agit alors d'une loi Binomiale $B(n_x, i_x)$. En utilisant le Théorème Central Limite et le Théorème de Slutsky, découle le résultat de convergence qui suit :

$$\hat{i}_x \xrightarrow{\mathcal{L}} \mathcal{N}\left(i_x, \sqrt{\frac{i_x(1-i_x)}{n_x}}\right)$$

En notant $q_{\frac{\alpha}{2}}$, le quantile d'ordre α de la loi normale centrée réduite, pour n_x suffisamment grand, l'intervalle de confiance asymptotique de niveau α s'écrit :

$$IC_{\alpha} = \left[\hat{i}_x - q_{\frac{\alpha}{2}} \sqrt{\frac{i_x(1-i_x)}{n_x}}, \hat{i}_x + q_{\frac{\alpha}{2}} \sqrt{\frac{i_x(1-i_x)}{n_x}} \right]$$

Ainsi sont obtenus les taux bruts d'incidence et les intervalles de confiance associés .

Le taux d'incidence en dépendance augmente avec l'âge, ce qui est raisonnable vu la nature du risque. Plus précisément, le taux d'incidence reste quasi-constant jusqu'à 70 ans

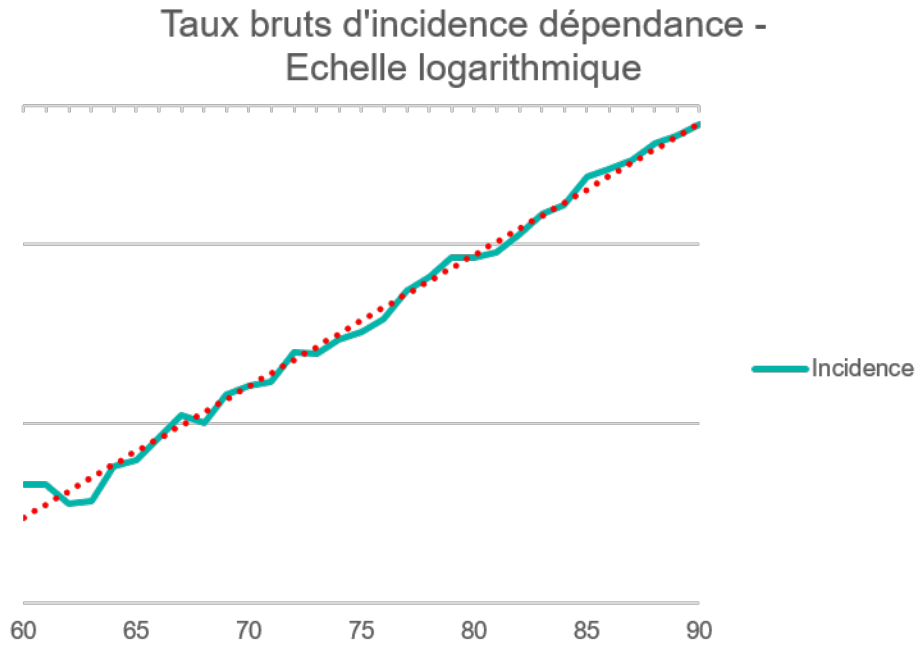


FIGURE 3.12 – Taux d'incidence bruts en échelle logarithmique

avant d'augmenter lentement entre 70 et 80 ans, il croit de façon exponentielle au-delà.

La figure 3.12 représente les taux bruts d'incidence avec une échelle logarithmique. Elle nous permet de confirmer la tendance de croissance exponentielle. La courbe des taux bruts d'incidence en échelle logarithmique s'ajuste bien à une droite à partir de 65 ans.

Ce qui confirme la tendance de croissance exponentielle à partir de 65 ans, donc sur les âges de forte exposition à la dépendance

Les taux bruts d'incidence ont aussi été calculés par sexe. La comparaison est reprise par la figure 3.13. Jusqu'à 80 ans, les taux d'incidence sont quasi-confondus avec les taux d'incidence des femmes légèrement en dessous par moment. A partir de 80 ans, les taux d'incidence des femmes sont sensiblement supérieurs à ceux des hommes sauf entre 81 et 84 où ils sont presque similaires. Cependant, à 90 ans, le taux d'incidence des hommes est largement supérieur à celui des femmes. Cela pourrait être dû à la faible exposition des hommes sur ces âges.

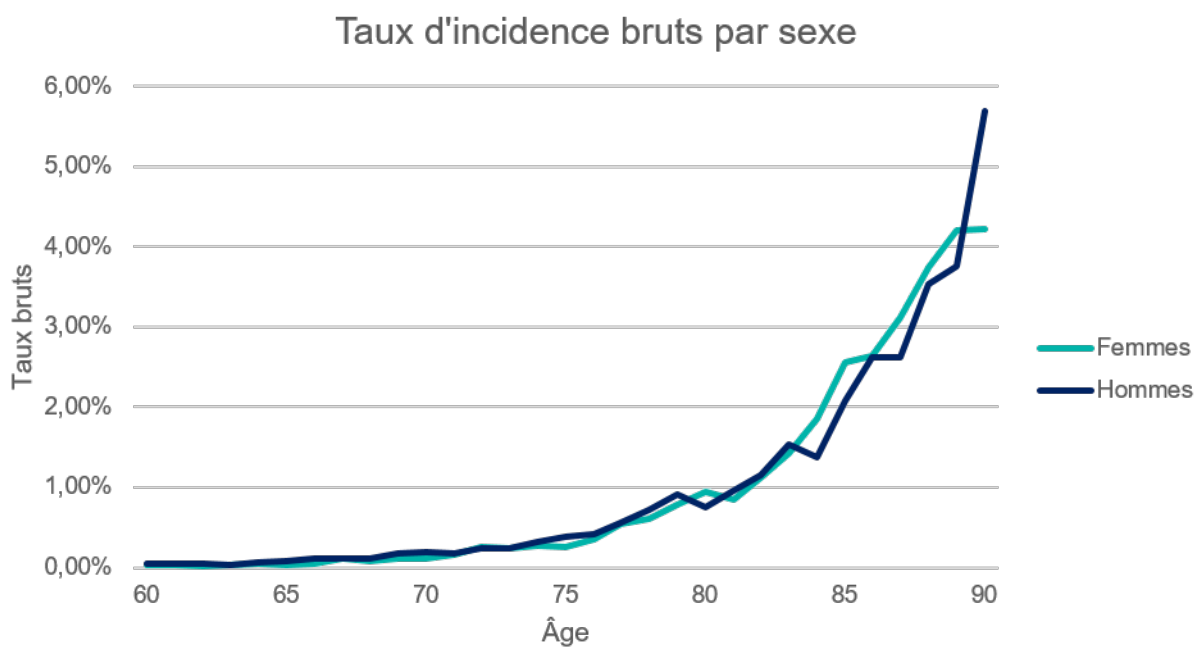


FIGURE 3.13 – Comparaison taux d'incidence bruts par sexe

Pour juger de la pertinence de poursuivre, ou pas, la construction en différenciant le sexe, nous allons utiliser le test du log-rank.

3.3.1 Test du log-rank

Le test de Mantel-Haenszel communément appelé test du log-rank est l'un des tests les plus utilisés en analyse de survie pour comparer les caractéristiques de deux échantillons. Il est en pratique très utilisé pour savoir si une variable a un effet significatif sur la quantité modélisée. Le test suppose d'emblée que les facteurs de risques sont les mêmes pour les deux groupes et compare les taux observés du deuxième groupe avec ceux du premier.

Soient deux échantillons A et B, indépendants et censurés. Le but du test est de vérifier l'hypothèse d'égalité des fonctions de survie des deux échantillons A et B.

- **Hypothèse \mathcal{H}_0** : $S_A = S_B$ les fonctions de survie des deux échantillons sont les mêmes
- **Hypothèse \mathcal{H}_1** : $S_A \neq S_B$ les fonctions de survie des deux échantillons ne sont pas les mêmes

Le test consiste à construire des statistiques sur les sommes des différences entre les taux théoriques et les taux observés ($i_{t,j}^{th} - i_{t,j}^{obs}$) qui sont asymptotiquement gaussiennes sous \mathcal{H}_0 . j est l'indice du groupe ($j \in \{1, 2\}$ respectivement pour A et B).

Soit

$$\Psi_j = \frac{[\sum_i (i_{t,j}^{th} - i_{t,j}^{obs})]^2}{\sum_i Var i_{t,j}^{obs}}$$

Les Ψ_j suivent une loi de Khi-deux de degré de liberté le nombre d'échantillon moins 1. La conclusion du test se fait avec la p-value de niveau $1 - \alpha$ (=95% en général) :

- Si $p \leq \alpha$: on rejette l'hypothèse \mathcal{H}_0
- sinon on ne rejette pas l'hypothèse \mathcal{H}_0

Nous avons effectué le test sur R et les résultats nous permettent de rejeter l'hypothèse \mathcal{H}_0 même avec un seuil de 99% (p-value très faible).

```
call:
survdif(formula = surv(incidence$`Age fin observation`, incidence$`Fin observation
pour dépendance`) ~
incidence$Genre)

n=424071, 1 observation deleted due to missingness.

      N Observed Expected (O-E)^2/E (O-E)^2/V
incidence$Genre=1 232897    2388    2172    21.4    36
incidence$Genre=2 191174    3270    3486    13.3    36

Chisq= 36 on 1 degrees of freedom, p= 2e-09
```

Le test du log-rank nous conduit à poursuivre la construction de notre loi d'incidence avec une loi pour les hommes et une autre pour les femmes. Ce qui semble être raisonnable au regard des données disponibles et de la différence d'exposition des deux sexes au risque. Rappelons le, les hommes sont exposés au risque de façon plus précoces par rapport aux femmes, ce qui fait que sur les âges élevés (vers 90 ans) peu d'hommes sont toujours vivants et valides.

3.3.2 Lissage des taux bruts

Après la détermination des taux bruts, des irrégularités sont constatées au niveau de certains âges. Cela serait dû au manque de données. Pour y remédier, les taux bruts seront lissés tout en maximisant deux critères : la fidélité des taux lissés aux taux bruts et leur régularité.

Le lissage sera effectué entre les âges 60 et 90 car en dehors de cette tranche la faible exposition pourrait conduire à des aberrations. Le lissage permet de réduire les irrégularités causées par le manque de données sur certains âges afin d'avoir une loi plus réaliste et sensée. La méthode choisie sera celle qui allie le mieux les critères de fidélité et de régularité.

Méthode des Moyennes mobiles

L'une des méthodes de lissage les plus intuitives est celle des moyennes mobiles. Elle est le plus souvent utilisée dans l'analyse des séries temporelles. Le principe est le suivant, pour tout âge x , le taux lissé est obtenu en faisant une moyenne arithmétique de $2h+1$ taux :

- les taux aux h âges précédents l'âge x ($x - 1, x - 2, \dots, x - h$)
- le taux à l'âge x
- les taux aux h points suivants l'âge x ($x + 1, x + 2, \dots, x + h$)

Soient \hat{i}_x et i_x respectivement taux bruts et lissés pour l'âge x , la formule des moyennes mobiles est la suivante :

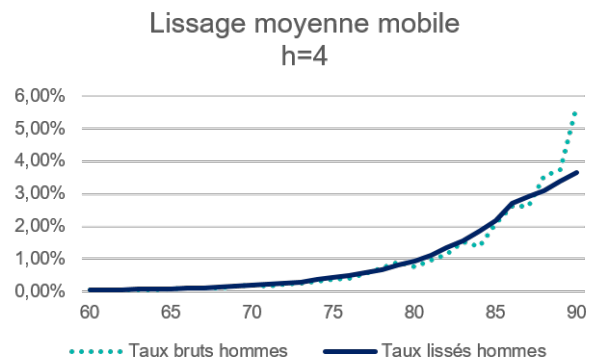
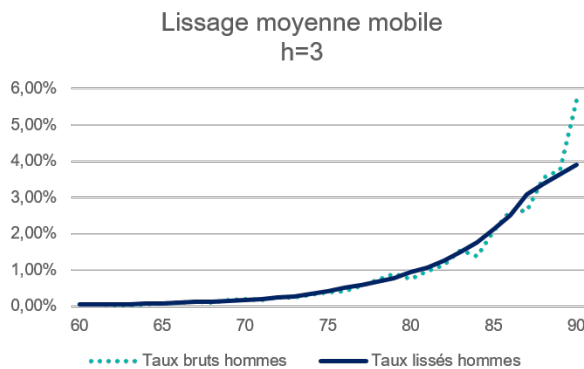
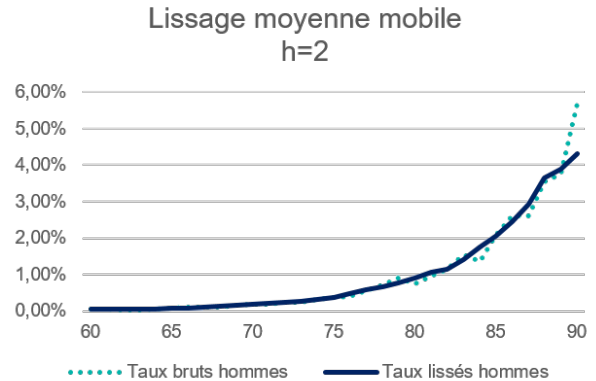
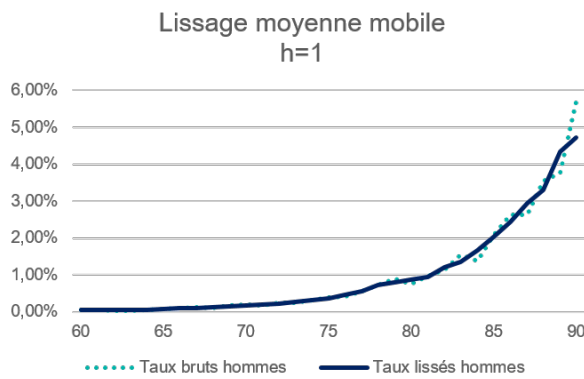
$$i_x = \frac{1}{2h + 1} \sum_{j=-h}^h \hat{i}_{x+j}$$

Il apparaît que pour les taux en début et en fin de la plage de lissage, un manque de données se posera. En effet, les h premiers taux bruts et h derniers taux bruts seront concernés par cette limite de la méthode des moyennes mobiles. Nous adapterons la formule aux données disponibles. Par exemple pour $h = 2$, soient v et u respectivement premier et dernier âge de notre intervalle de lissage, le taux lissé i_{v+1} sera la moyenne mobile en utilisant :

- le taux brut \hat{i}_v
- le taux brut \hat{i}_{v+1}
- les taux bruts \hat{i}_{v+2} et \hat{i}_{v+3}

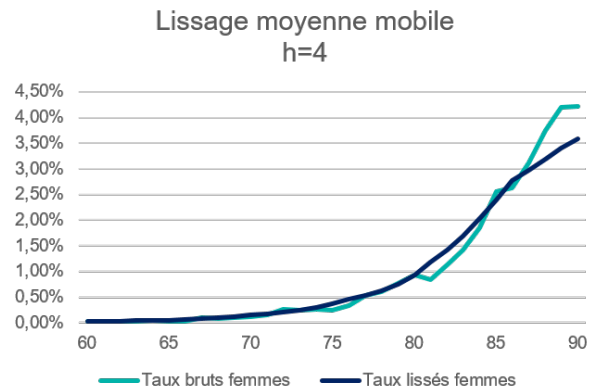
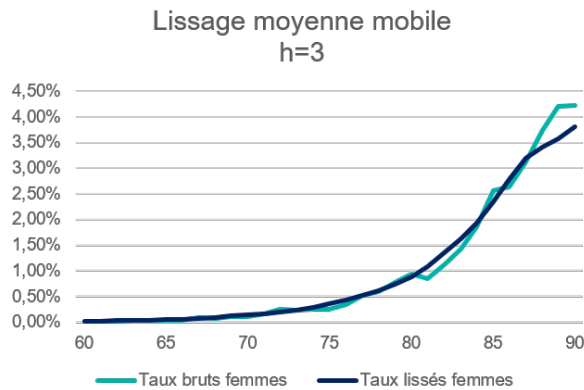
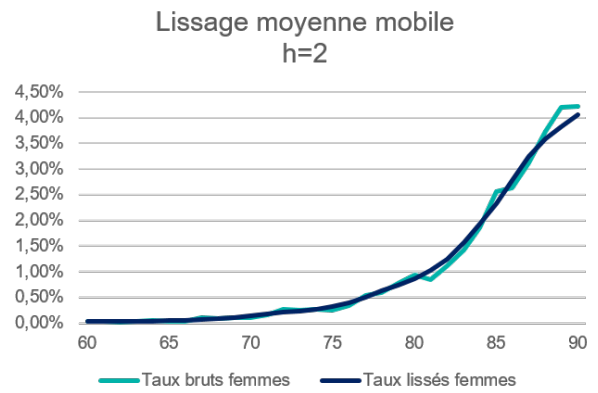
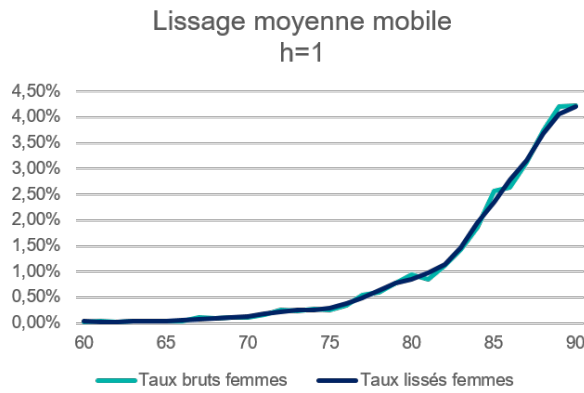
Les taux i_v, i_{u-1} et i_u seront obtenus suivant la même logique. Ce choix favorise la régularité.

En faisant varier h entre 1 et 4, les lissages obtenus sur les taux d'incidence bruts des hommes sont les suivants :



Plus le paramètre h est petit, plus les taux lissés sont fidèles aux taux bruts et épousent davantage les irrégularités. Plus h est grand, davantage les taux sont lissés mais s'éloignent des taux bruts sur les extrémités. Le lissage obtenu avec $h = 3$ semblent être le plus pertinent, il sera par la suite comparé aux résultats des autres méthodes.

Concernant le lissage des taux bruts d'incidence des femmes, ci-après les résultats obtenus :



Les remarques soulignées lors de l'utilisation de la méthode pour les taux d'incidence des hommes sont d'actualité. Le lissage obtenu avec le paramètre $h = 2$ nous semble être le plus pertinent en terme de précision et de régularité, il sera par la suite comparé aux lissages des autres méthodes.

Méthode de Whittaker-Henderson

La méthode de lissage de Whittaker-Henderson consiste à rechercher les valeurs ajustées qui minimisent une combinaison des critères de fidélité (ou précision) et de régularité.

Dans le cas d'une modélisation en dimension un, qui nous concerne, les critères de fidélité F et de régularité S ("Smoothness") sont définis comme suit :

$$F = \sum_{x=x_{min}}^{x_{max}} w_x (i_x - \hat{i}_x)^2 = (i - \hat{i})' W (i - \hat{i})$$

Et :

$$S = \sum_{x=x_{min}}^{x_{max}-z} (\Delta^z i_x)^2 = ((\Delta^z i)' (\Delta^z i) = i' K_z' K_z i$$

Avec :

- w_x les poids attribués aux taux bruts
- $W = \text{diag}(w_x)_{x_{\min} \leq x \leq x_{\max}}$ la matrice des poids
- i_x les taux lissés
- \hat{i}_x les taux bruts
- z paramètre du modèle, entier positif représentant le degré du polynôme auquel va s'ajuster la courbe lissée
- $\Delta^z i_x = \sum_{j=0}^z \binom{z}{j} (-1)^{z-j} i_{j+x}$ l'opérateur différence avant, composé z fois
- K_z matrice de taille (pz, p) , dont les termes sont les coefficients binomiaux d'ordre z et dont le signe alterne et commence positivement pour z pair, $\Delta^z i = K_z i$
- h réel positif contrôlant le poids accordé au critère de régularité dans le lissage.

Les valeurs lissées cibles sont celles qui minimisent la combinaison linéaire des deux critères à travers le paramètre h , à savoir :

$$\begin{aligned}
M &= F + h.S \\
&= (i - \hat{i})' W (i - \hat{i}) + h.i' K_z' K_z i \\
&= i' W i - 2i' \hat{i} + \hat{i}' W \hat{i} + h.i' K_z' K_z i
\end{aligned}$$

Pour obtenir le i qui minimise M , il faut annuler la dérivée de M par rapport à i :

$$\frac{\delta M}{\delta i} = 2W i - 2W \hat{i} + 2h.K_z' K_z i$$

$\frac{\delta M}{\delta i} = 0$ conduit à l'expression des taux lissés suivante :

$$i = (W + h.K_z' K_z)^{-1} W \hat{i}$$

En effet $h.K_z' K_z$ n'est pas inversible mais l'addition de la matrice W rend la somme inversible.

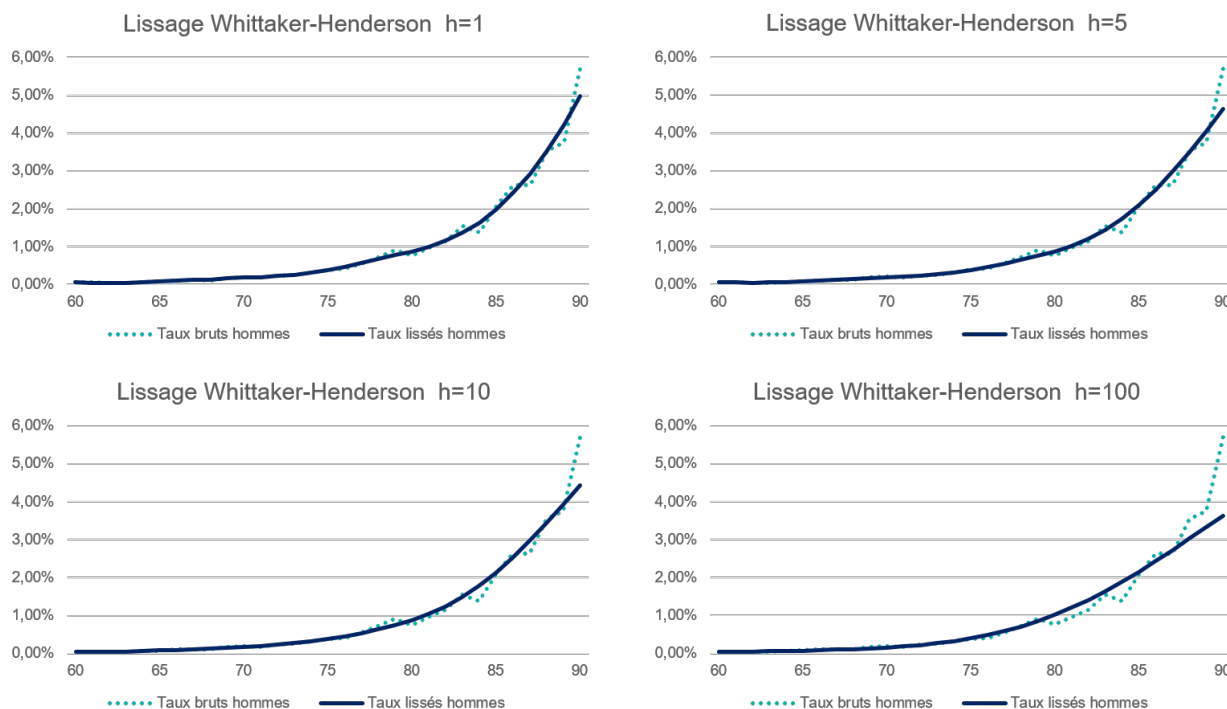
Pour la suite, nous faisons le choix de travailler avec $z = 2$ pour des raisons d'ordre pratique, soit $\Delta^2 i_x = \Delta(i_{x+1} - i_x) = i_{x+2} - 2i_{x+1} + i_x$.

Pour choisir les poids w_i , plusieurs options sont possibles :

- considérer tous les w_x égaux à 1 : cela revient à accorder la même crédibilité à tous les taux bruts \hat{i}_x , ce qui ne semble pas très pertinent étant donné qu'à certains âges, la faible exposition conduit à des irrégularités ;
- pondérer par l'exposition au risque de nos assurés à chaque âge : $w_x = n_x$;
- pondérer par l'exposition au risque normalisée à chaque âge : $w_x = \frac{n_x}{\bar{n}}$ avec \bar{n} l'exposition moyenne par âge. Cette option semble être plus pertinente en ce qu'elle permet

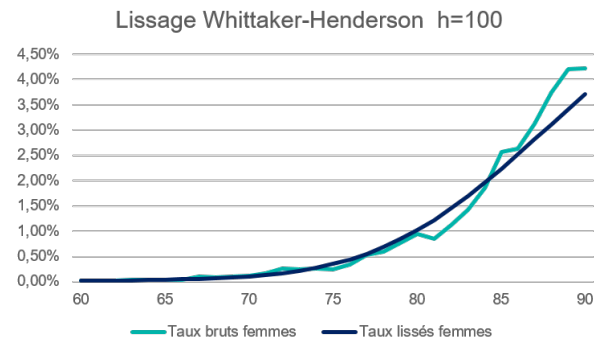
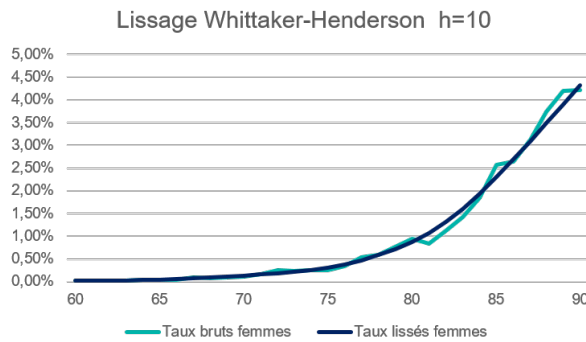
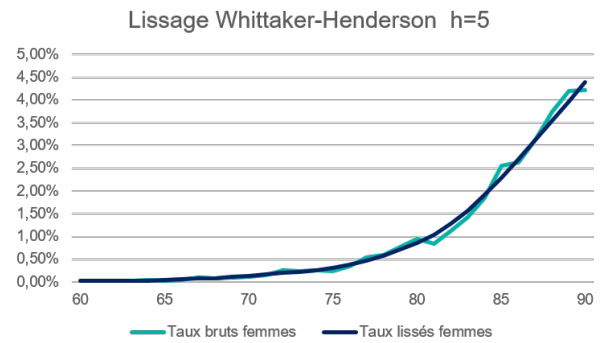
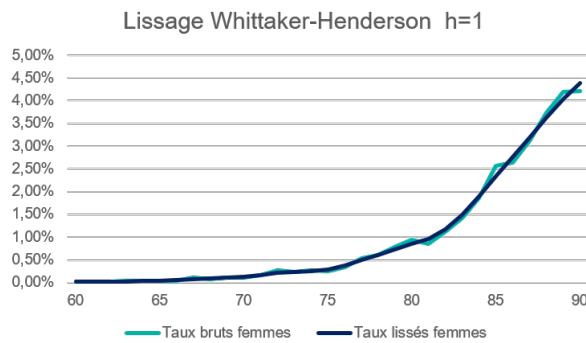
de contrôler le critère de fidélité, elle augmente l'influence de la fidélité aux âges où assez de données sont disponibles et la réduit pour les âges avec peu de données. C'est cette approche qui est retenue pour la suite.

Il faudra aussi choisir le paramètre h et les poids w_i . Plus h est grand, plus la régularité est privilégiée au détriment de la fidélité. Inversement, plus h est petit, plus on priorise la fidélité. Il s'agira donc de faire varier h et de choisir la valeur qui donne le lissage le plus pertinent. Un test avec les valeurs 1, 5, 10 et 100, choisis aléatoirement, donne les résultats suivants avec une simulation sur R :



Sans surprise, plus h est grand plus la courbe des taux ajustés est régulière et renonce aux disparités des taux bruts. La courbe obtenue avec $h = 100$ est lisse mais s'éloigne des taux bruts sur les âges élevés contrairement à celle obtenue avec $h = 1$ qui s'ajuste bien aux taux bruts mais présente toujours des irrégularités. Quant aux courbes des taux ajustés avec $h = 5$ et $h = 10$, elles sont régulières tout en étant fidèles aux taux bruts. Elles ne présentent pas de différence notoires sauf sur les âges qui présentent de réelles irrégularités. En privilégiant la régularité, les taux obtenus avec $h = 10$ semblent être plus adéquats.

Toujours sur R, pour $h \in \{1, 5, 10, 100\}$, voici les résultats du lissage de Whittaker-Henderson sur les taux d'incidence bruts des femmes :



Les taux ajustés avec $h = 100$ s'éloignent considérablement des taux bruts autour de 80 ans (près des âges moyen d'entrée en dépendance) et sur les âges élevés. Quant aux courbes des taux ajustés avec $h = 5$ et $h = 10$, elles sont régulières tout en étant plus ou moins fidèles aux taux bruts. Elles ne présentent quasiment pas de différence. En privilégiant la régularité, les taux lissés avec $h = 10$ semblent être plus pertinents. Ils seront comparés aux résultats des autres méthodes.

Lissage paramétrique de Gompertz-Makeham

Le modèle de Gompertz-Makeham est habituellement utilisé pour la construction de tables de mortalité comme modèle de référence. Il est défini par la fonction de risque instantanée suivante :

$$h(t) = a + b \times c^t$$

Le modèle suppose l'existence de deux causes principales à l'origine du risque. Appliqué à la dépendance :

- la vieillesse : modélisée par les paramètres b et c comme une croissance exponentielle avec l'âge,
- cause accidentelle : représentée par le paramètre a .

Étant donné que les taux d'incidence en dépendance augmentent globalement avec l'âge,

c doit être supérieur à 1 et b positif.

Or,

$$i_x = 1 - \frac{S(x+1)}{S(x)} \quad \text{et} \quad S(x) = \exp\left[-\int_0^x h(s)ds\right]$$

Donc,

$$\begin{aligned} 1 - i_x &= \exp\left[-\int_x^{x+1} h(s)ds\right] \\ &= \exp\left[-\int_x^{x+1} a + b \times c^s ds\right] \\ &= \exp(-a)\exp\left[-\frac{b}{\ln(c)}c^x(c-1)\right] \end{aligned}$$

En posant : $s = \exp(-a)$ et $g = \exp(-\frac{b}{\ln(c)})$,

$$1 - i_x = s \times g^{c^x(c-1)}$$

Soit,

$$i_x = 1 - s \times g^{c^x(c-1)}$$

C'est cette formule qui sera utilisée pour lisser les taux bruts. Avant de passer à l'ajustement, il est pertinent de vérifier l'adéquation du modèle de Gompertz-Makeham aux données utilisées.

D'après la précédente écriture du taux d'incidence, en passant au logarithme :

$$\ln(1 - i_x) = \ln(s) + \ln(g)c^x(c-1)$$

Or, pour tout x au voisinage de 0, le développement limité à l'ordre 1 de $\ln(1+x)$ permet une approximation x par x . Comme nos taux d'incidences sont relativement proches de 0, $\ln(1 - i_x) \approx -i_x$. Soit :

$$-i_x = \ln(s) + \ln(g)c^x(c-1)$$

D'où

$$i_{x+1} - i_x = c^x(c-1)^2 \ln(g)$$

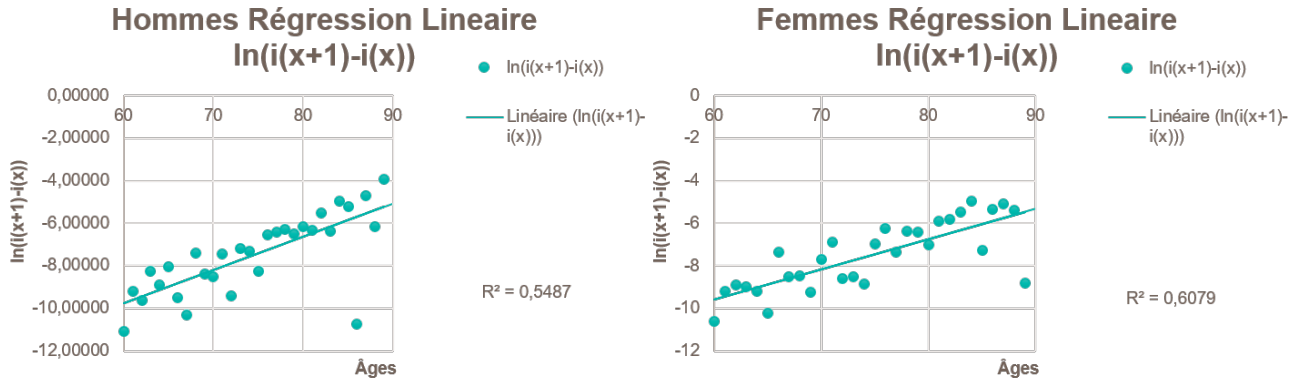
En passant au logarithme, il ressort que :

$$\ln(i_{x+1} - i_x) = x \ln(c) \ln((c-1)^2 \ln(g))$$

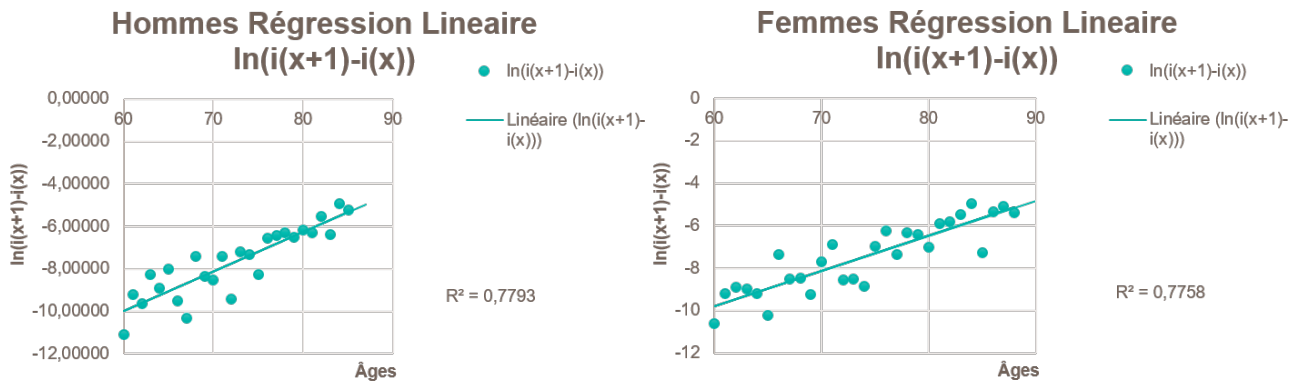
Il apparaît alors que $\ln(i_{x+1} - i_x)$ est une fonction linéaire de l'âge x de pente $\ln(c)$. C'est cette propriété qu'il faudra vérifier sur les données à l'aide d'une régression linéaire de $\ln(i_{x+1} - i_x)$ en fonction de x . Si le nuage de points $(x, y = \ln(i_{x+1} - i_x))$ est aligné sur une

droite, alors cela réconforte l'adéquation du modèle de Gompertz-Makeham aux taux bruts d'incidence.

Les graphes qui suivent représentent $\ln(i_{x+1} - i_x)$ en fonction de x pour les taux bruts des hommes et ceux des femmes.



Graphiquement, les points sont relativement alignés le long d'une droite à l'exception de quelques points qui s'éloignent et écrasent le R^2 . La régression pour les taux des hommes retourne un $R^2 = 0.55$ et celle des femmes un $R^2 = 0.61$. Pour améliorer l'adéquation à une droite, la plage de lissage est restreinte. Concernant les hommes, le lissage sera effectué sur les âges compris entre 60 et 86 ans et pour les femmes entre 60 et 89 ans. Les résultats des deux régressions sont les suivantes.



Avec les nouvelles plages de lissage, l'adéquation des données à une droite est améliorée et les coefficients de détermination s'améliorent considérablement, $R^2 = 0.78$ aussi bien pour les hommes que pour les femmes.

La fonction `summary()` du logiciel R retourne les caractéristiques suivantes pour la régression linéaire sur les taux d'incidence bruts des hommes (à gauche) et ceux des femmes (à droite) :

```

Call:
lm(formula = y ~ x, data = t)

Residuals:
    Min       1Q   Median       3Q      Max
-1.6499 -0.3548  0.1082  0.4487  1.1801

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -10.17307    0.31265  -32.539 < 2e-16 ***
x             0.18638    0.02024   9.207 2.41e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7742 on 24 degrees of freedom
Multiple R-squared:  0.7793,    Adjusted R-squared:  0.7701
F-statistic: 84.76 on 1 and 24 DF,  p-value: 2.408e-09

Call:
lm(formula = Fy ~ Fx, data = Ft)

Residuals:
    Min       1Q   Median       3Q      Max
-1.6311 -0.5220  0.2221  0.4538  1.4329

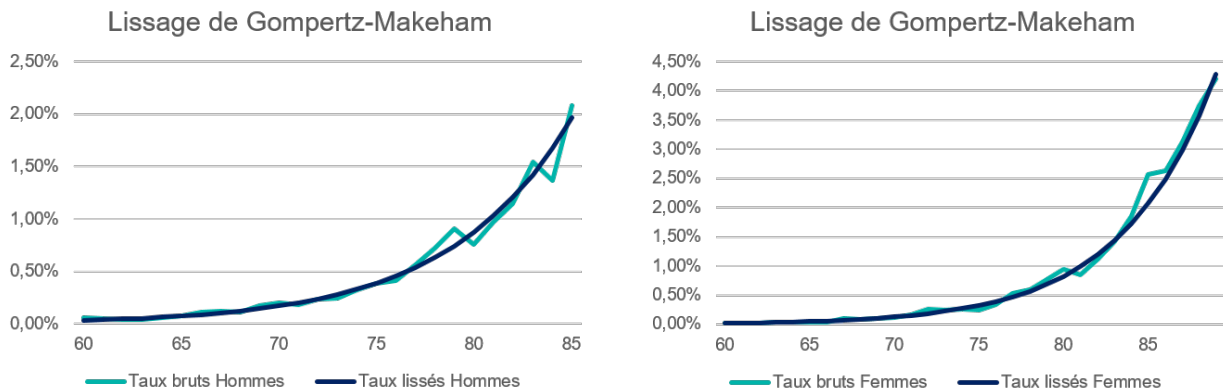
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -9.95418    0.29447  -33.803 < 2e-16 ***
Fx             0.16574    0.01714   9.667 2.92e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7725 on 27 degrees of freedom
Multiple R-squared:  0.7758,    Adjusted R-squared:  0.7675
F-statistic: 93.45 on 1 and 27 DF,  p-value: 2.92e-10

```

Les deux régressions présentent des R^2 ajustés de 0.77, la variabilité est donc expliquée en majeure partie par le modèle. La régression linéaire est donc pertinente. De plus, le test de Fisher conduit à rejeter l'hypothèse \mathcal{H}_0 de nullité du coefficient du modèle. En effet, les p-value des hommes et des femmes valent respectivement $2.41e^{-09}$ et $2.92e^{-09}$ qui sont négligeable devant 5%.

Par conséquent il peut être pertinent de modéliser les taux d'entrée en dépendance par une loi de Gompertz-Makeham. Le modèle sera à présent ajusté aux taux bruts. Les paramètres seront estimés par la méthode du maximum de vraisemblance.

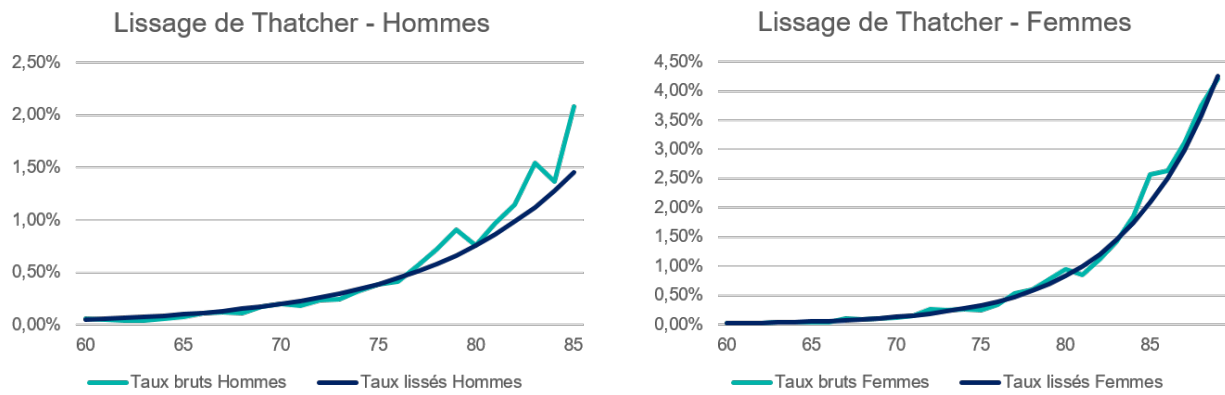


Lissage paramétrique de Thatcher

Le modèle de Thatcher est basé sur la fonction de hasard suivante :

$$h(t) = \frac{\alpha \exp(\beta x)}{1 + \alpha \exp(\beta x) + \gamma}$$

Les paramètres sont obtenus, de façon similaires à ceux du modèle de Gompertz-Makeham, par la méthode du maximum de vraisemblance. Appliqué aux taux bruts d'incidence des hommes et des femmes, le modèle donne les résultats qui suivent :



3.3.3 Comparaison des lissages

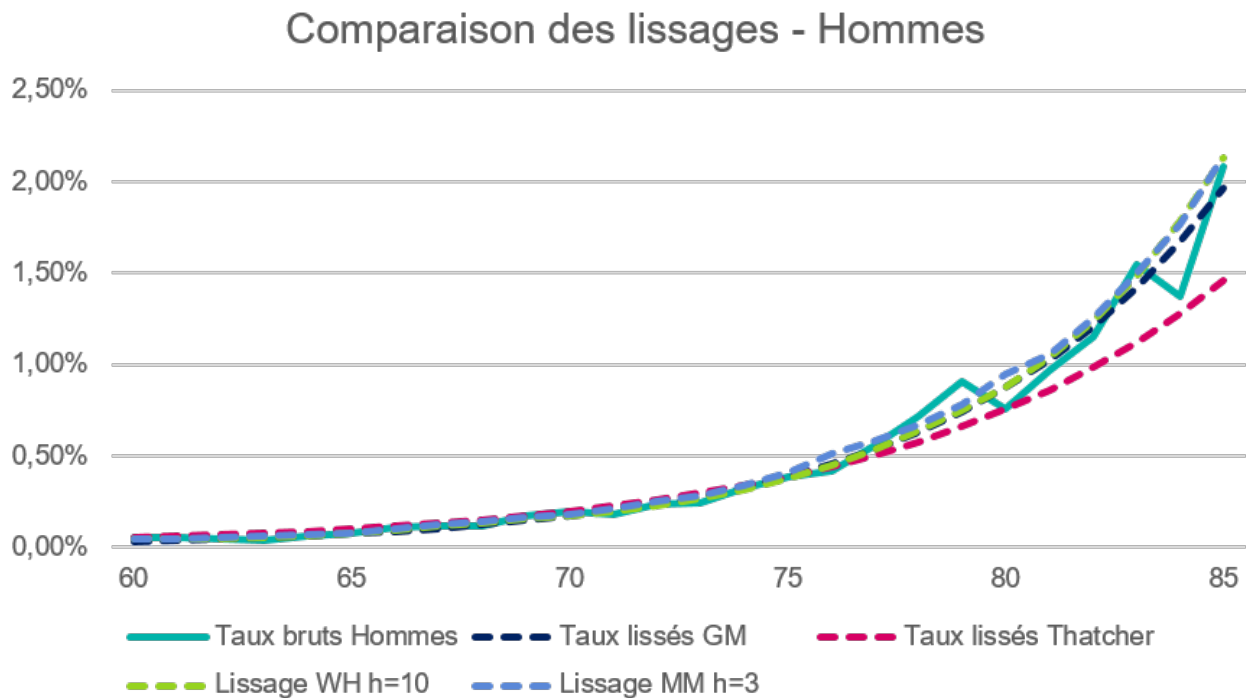


FIGURE 3.14 – Comparaison des lissages sur les taux d'incidence Hommes

Il est à noter que les taux lissés par le modèle de Thatcher sont largement inférieurs aux taux bruts et aux estimés des autres méthodes, il y a un fort risque de sous-estimation. Le modèle de Gompertz de Gompertz-Makeham donne des taux ajustés proches de ceux du lissage de Whittaker Henderson sauf sur les âges élevés où ses taux sont inférieurs. Quant à la méthode de la Moyenne Mobile, elle permet d'avoir des taux quasi-similaires à ceux de Whittaker Henderson surtout sur les âges élevés mais donne des taux sensiblement supérieurs entre 70 et 81 ans, elle est aussi moins lisse. Le lissage obtenu avec la méthode de Whittaker

Henderson semble être le plus pertinent.

Concernant les taux d'incidence des femmes, les différents lissages donnent des résultats assez proches, en particulier les taux obtenus avec Gompertz-Makeham et Thatcher sont quasi-identiques.

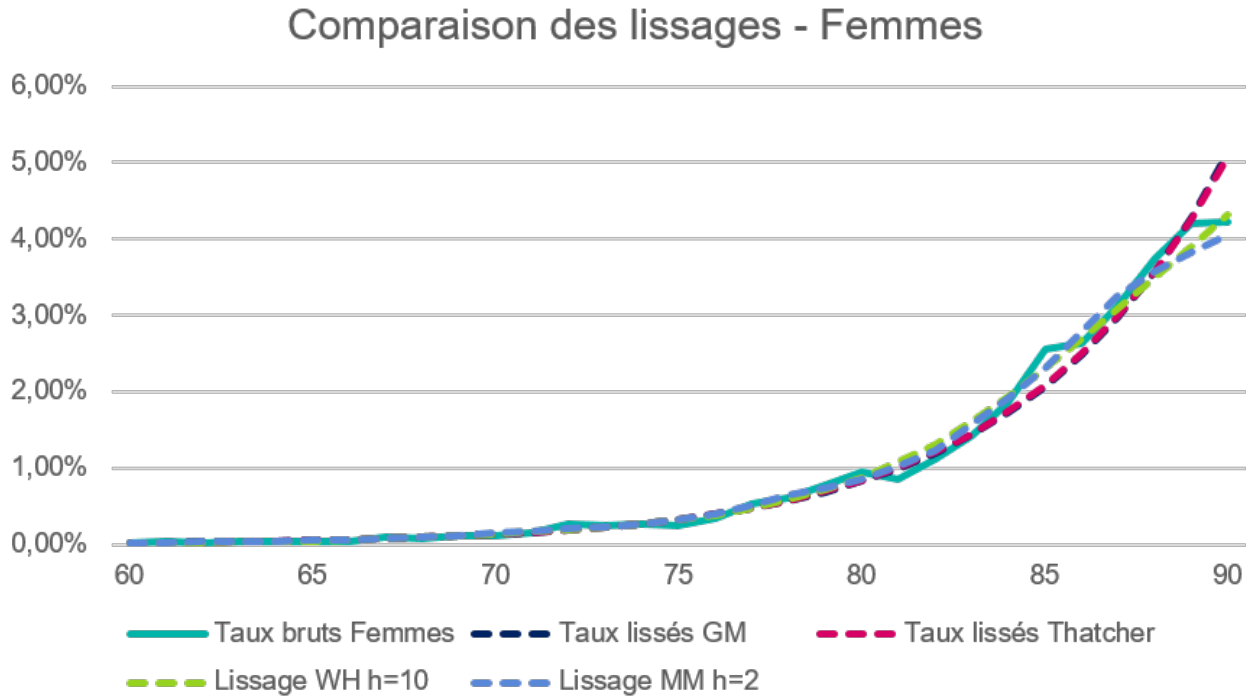


FIGURE 3.15 – Comparaison des lissages sur les taux d'incidence Femmes

Dans la suite, il s'agira de comparer les résultats des différentes méthodes d'abord par les critères de fidélité et de régularité, puis en faisant un backtesting.

Fidélité des taux lissés

Comme vu précédemment, la fidélité traduit la précision du lissage par rapport aux taux bruts. Autrement dit, elle mesure l'écart entre les taux lissés et les taux bruts. Le moyen le plus intuitif d'évaluer ce critère est de calculer la distance (valeur absolue) entre les taux bruts et ajustés. Plus la somme de cette distance est proche de 0, plus le critère de fidélité est satisfait. Nous calculons donc la somme suivante pour les différentes méthodes :

$$\sum_{x_{min}}^{x_{max}} |\hat{i}_x - \hat{i}_x^*|$$

Où \hat{i}_x^* représente les taux lissés.

Les tableaux suivants reprennent les valeurs obtenues avec les méthodes utilisées pour les hommes et les femmes.

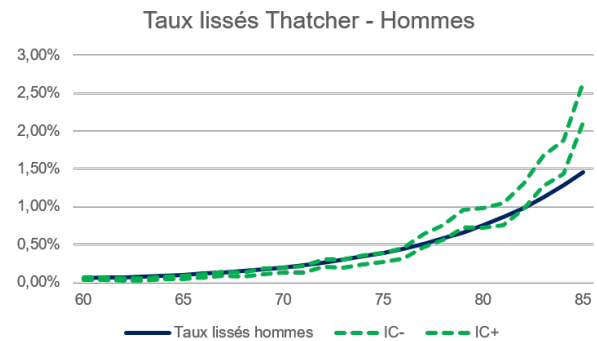
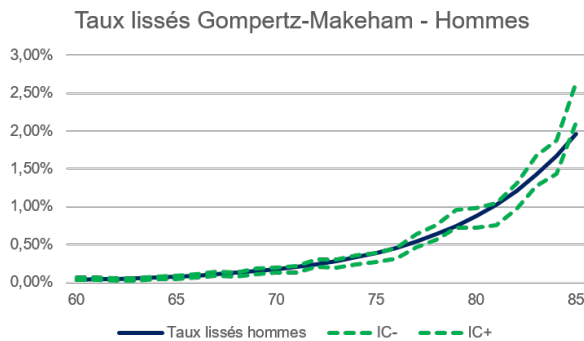
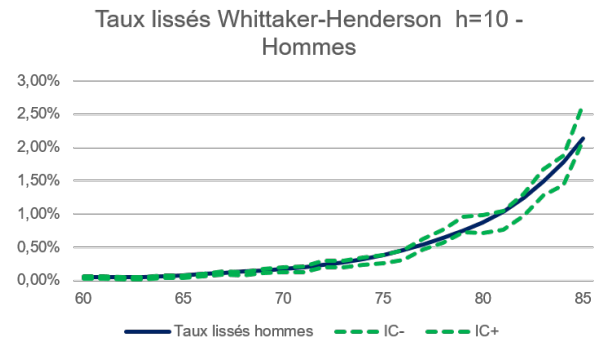
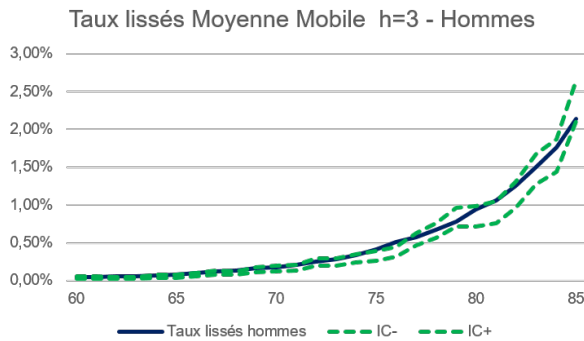
Lissages Hommes	Distance globale
Moyennes mobiles	0.014
Whittaker-Henderson	0.013
Gompertz-Makeham	0.020
Thatcher	0.020

Lissages Femmes	Distance globale
Moyennes mobiles	0.021
Whittaker-Henderson	0.021
Gompertz-Makeham	0.013
Thatcher	0.019

Il apparaît que pour les taux d'incidence des hommes, la méthode Whittaker-Henderson fournit les taux les plus fidèles suivie de près par celle des moyennes mobiles. Les taux ajustés par les modèles de Gompertz-Makeham et Thatcher sont moins fidèles. Cela s'explique par le fait que les taux bruts d'incidence des hommes présentent plusieurs irrégularités et les modèles paramétriques ont plus de mal à s'ajuster.

Concernant l'incidence des femmes, les taux ajustés par le modèle de Gompertz-Makeham sont les plus fidèles aux taux bruts suivis de ceux ajustés par le modèle de Thatcher. Les taux lissés par les méthodes des Moyenne Mobiles et de Whittaker-Henderson sont moins fidèles.

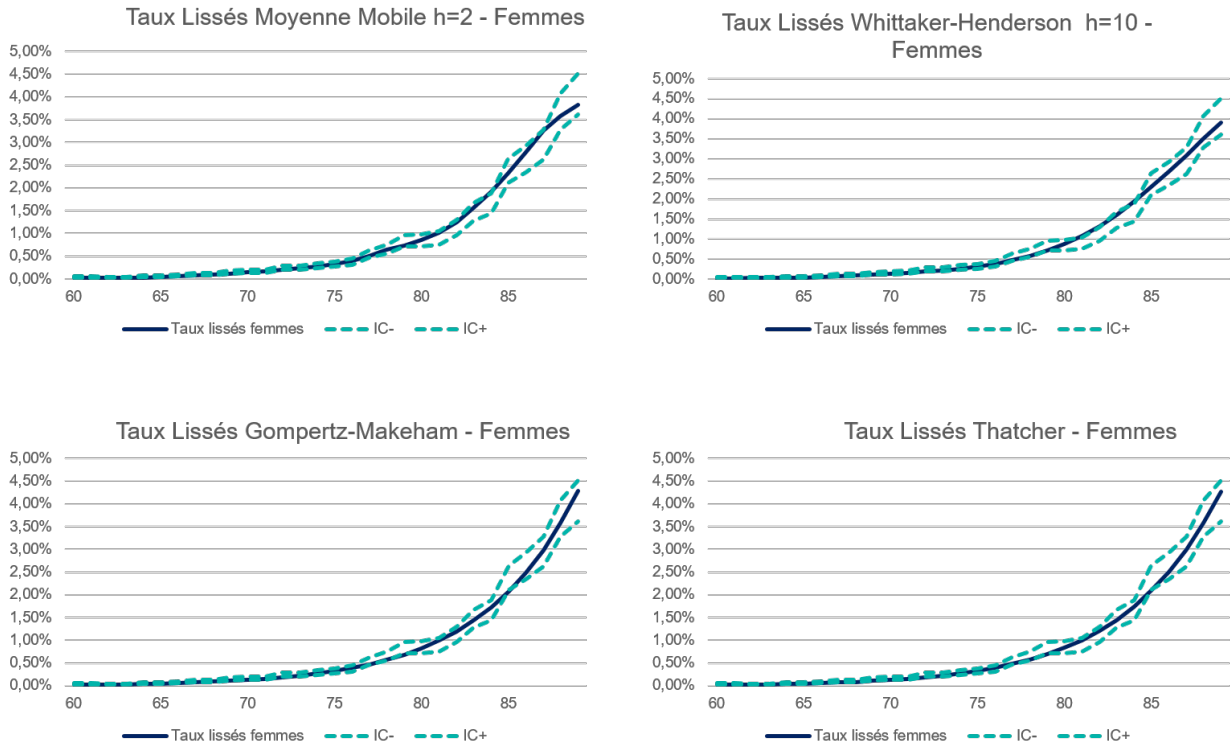
Le critère décrit ci-dessus évalue la fidélité globale des lissages, pour regarder la fidélité de plus près, il serait intéressant de regarder les taux lissés avec les intervalles de confiance asymptotiques des taux bruts. Avec des IC asymptotiques à 95%, nous obtenons les représentations graphiques suivantes pour les hommes.



À part les taux ajustés avec le modèle de Thatcher qui sortent carrément de l'intervalle

de confiance asymptotique., les taux lissés restent globalement dans l'intervalle de confiance asymptotique des taux bruts. Les quelques sorties notées sont de légères surestimation des taux bruts, règle de prudence.

Les mêmes représentations pour les lissages des taux bruts d'incidence en dépendance des femmes donnent les graphiques qui suivent.



Quant aux taux lissés d'incidence des femmes, ils restent globalement dans l'intervalle de confiance asymptotique des taux bruts. Quelques points sont en dehors de l'intervalle de confiance asymptotique des taux bruts mais en surestimant légèrement les taux bruts.

Régularité des taux lissés Concernant la régularité des taux lissés, elle est évaluée par la somme des carrés des différences d'ordre z des taux lissés. La courbe des taux ajustés est considérée comme régulière si cette somme est proche de zéro. Autrement dit, la courbe lissée est régulière si :

$$\sum_{x_{min}}^{x_{max}} (\Delta^z \hat{i}_x^*)^2 \rightarrow 0$$

Pour $z = 1$ et $z = 2$, la quantité à évaluer est respectivement :

$$\sum_{x_{min}}^{x_{max}} (\hat{i}_{x+1}^* - \hat{i}_x^*)^2 \quad \text{et} \quad \sum_{x_{min}}^{x_{max}} (\hat{i}_{x+2}^* - 2\hat{i}_{x+1}^* + \hat{i}_x^*)^2$$

Ci-après les résultats obtenus :

Lissages Hommes	$z = 1$	$z = 2$	Lissages Femmes	$z = 1$	$z = 2$
Moyennes mobiles	4,E-05	3,E-06	Moyennes mobiles	1,13E-04	5,69E-06
Whittaker-Henderson	4,00E-05	1,15E-06	Whittaker-Henderson	1,14E-04	1,20E-06
Gompertz-Makeham	3,13E-05	6,94E-07	Gompertz-Makeham	1,16E-04	3,17E-06
Thatcher	1,40E-05	2,08E-07	Thatcher	1,15E-04	2,96E-06

En terme de régularité des taux lissés, pour les hommes, la méthode Whittaker-Henderson fournit des taux légèrement plus lisses que ceux de la méthode des Moyennes Mobiles. Le modèle de Thatcher donne la courbe la plus lisse malgré l'inadéquation aux taux bruts, suivi du modèle de Gompertz-Makeham. Il nous paraît alors plus pertinent de lisser les taux bruts d'incidence des hommes avec la méthode de Whittaker-Henderson.

Concernant l'incidence des femmes, la régularité des taux lissés se valent à l'ordre 1, à l'ordre 2, les taux obtenus avec Whittaker-Henderson sont les plus réguliers suivi respectivement de Thatcher, Gompertz-Makeham et la méthode des Moyennes Mobiles.

Backtesting des lois Afin de décider des lois à retenir, il est pertinent de faire un backtesting sur les données. En effet, il est important que les lois estiment au plus-près la réalité. De plus, concernant l'incidence en dépendance, la prudence milite en faveur de la surestimation. Voici le résultat des backtesting des lois d'incidence en dépendance Hommes et Femmes.

Lissages Hommes	Observés/Attendus	Lissages Femmes	Observés/Attendus
Moyennes mobiles	94,50%	Moyennes mobiles	98,80%
Whittaker-Henderson	99,03%	Whittaker-Henderson	100,50%
Gompertz-Makeham	101,65%	Gompertz-Makeham	103,84%
Thatcher	102,23%	Thatcher	103,58%

Au regard du backtesting, pour les lois d'incidence hommes, celle obtenue avec un lissage de Whittaker-Henderson est celle qui estime le mieux la réalité. La loi surestime par ailleurs le risque de dépendance de moins de 1%. Les lois obtenues avec Gompertz-Makeham et Thatcher sont plus éloignées des observations et sous-estiment le risque de dépendance, ce qui n'est pas très prudent. La loi lissée avec la méthode des Moyennes Mobiles surestiment de plus de 5% l'entrée en dépendance, ce n'est pas non plus le cas idéal, cela conduirait à des PRC trop importantes. Dans la suite, c'est la loi lissée avec Whittaker-Henderson qui sera retenue pour les hommes.

Quant aux femmes, le backtesting des lois d'incidence révèle une sous-estimation de plus de 3% des lois obtenues avec Gompertz-Makeham et Thatcher. La loi obtenue avec un lissage de Whittaker-Henderson est celle qui estime le mieux l'entrée en dépendance avec une sous-estimation de 0,5% du risque. La méthode des Moyennes Mobiles surestime le risque d'entrée en dépendance de 1,20%. Il s'agit donc de choisir entre le lissage avec Moyennes Mobiles et celui de Whittaker-Henderson. Ces deux méthodes se valent sur les autres critères passés en revue plus tôt. Par prudence, la loi obtenue avec le lissage par Moyennes Mobiles sera retenue pour la suite.

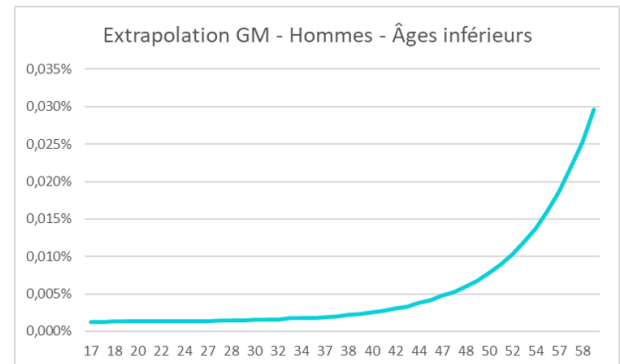
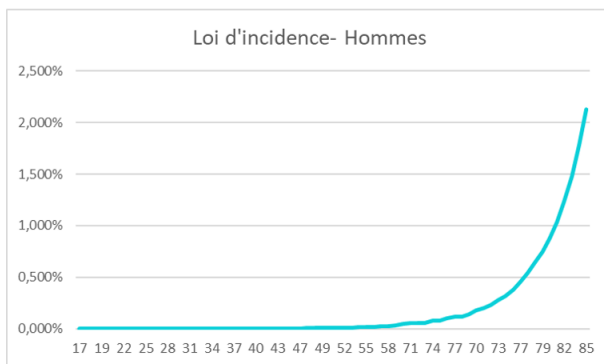
Les lois d'incidence lissées ont été construites pour des âges compris entre 60 et 85 ans

pour les hommes, et pour les femmes entre 60 et 89 ans. Pour rappel, au-delà le manque de données rendrait l'estimation non fiable. Il conviendra donc d'appliquer des méthodes d'extrapolation pour prolonger les taux. L'objectif est d'avoir les lois sur les âges [16;110]. En effet, les adhésions sont accessibles à partir de 16 ans. Il faudra alors extrapoler les lois sur les âges avant 60 ans dans un premier temps et ensuite sur les âges au-delà de 85 et 89 ans respectivement pour les hommes et les femmes.

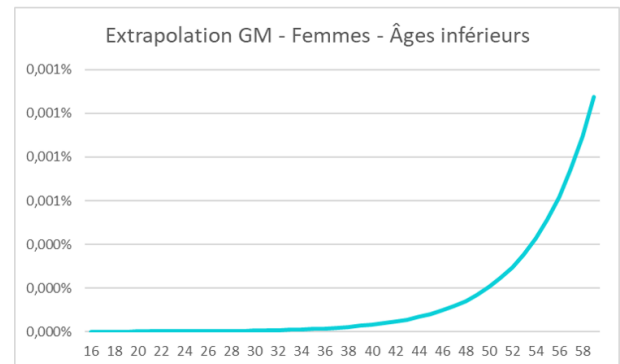
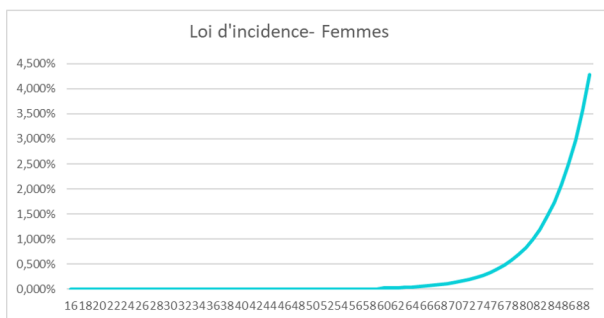
3.3.4 Extrapolation sur les âges inférieurs

Pour l'extrapolation sur les âges inférieurs, il est à noter que les lissages paramétriques étaient en bonne adéquation avec les taux bruts jusqu'à 75 ans pour les hommes et 80 ans pour les femmes. C'est cette piste qui sera exploitée.

Pour les hommes, la méthode de Gompertz-Makeham fournit des taux très proches des taux bruts entre 60 et 75 ans. La méthode de Thatcher fournit des taux plus éloignés. Il nous alors semblé pertinent d'utiliser les taux estimés par Gompertz-Makeham pour extrapoler les taux d'incidence jusqu'à 16 ans. Voici le résultat obtenu.



Pour les femmes, comme nous l'avons souligné précédemment, les méthodes de Gompertz-Makeham et Thatcher fournissent des taux d'incidence similaires. Nous avons alors réutilisé les taux estimés par la méthode de Gompertz-Makeham pour extrapoler les taux d'incidence jusqu'à 16 ans. Ci-après le résultat ainsi obtenu.



Nos lois sont à présents construites à partir de 16 ans et ce jusqu'à 85 et 89 ans respectivement pour les hommes et les femmes, nous allons les prolonger jusqu'à 110 ans.

3.3.5 Prolongement des lois sur les âges élevés

Les méthodes de Gompertz-Makeham et Thatcher utilisées pour le lissage permettent toutes deux de faire un prolongement de nos lois. En effet, leurs paramètres étant déjà estimés, il suffit d'appliquer leur formule sur nos plages d'extrapolations. Nous obtenons les résultats suivants pour les hommes et les femmes.

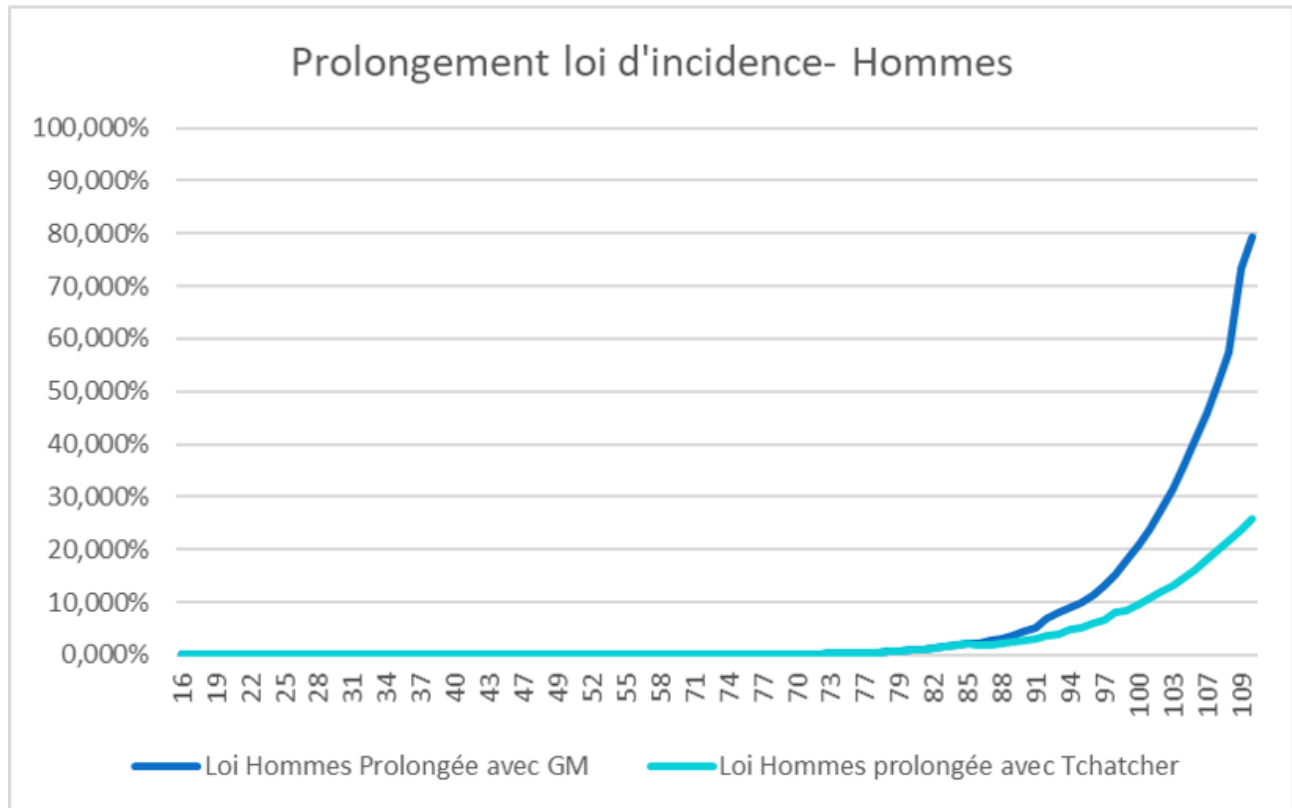


FIGURE 3.16 – Prolongement lois d'incidence Hommes

Nous remarquons dans les deux cas que le prolongement estimé par la méthode de Gompertz-Makeham traduit une augmentation exponentielle de l'incidence en dépendance sur les âges élevés. Cela s'apparente plus à la mortalité. Les taux d'entrée en dépendance ne devraient pas "exploser" autant. La méthode de Thatcher semble être plus adéquate pour nos prolongements. Néanmoins, pour être sûr de ne pas sous-estimer l'entrée en dépendance sur la plage de prolongement, nous avons comparé nos lois ainsi obtenues et deux lois d'incidence utilisées en interne. Il s'agit de lois de tarification utilisées l'une pour la tarification des garanties dépendances obligatoires (adossées à des contrats collectifs obligatoires) et l'autre pour la tarification des garanties dépendances facultatives. Ces lois n'étant pas différenciés par sexe, nous avons reconstitués une loi unique d'incidence en respectant la proportion Hommes/Femmes de notre portefeuille. Les résultats de cette comparaison sont présentés ci-après.

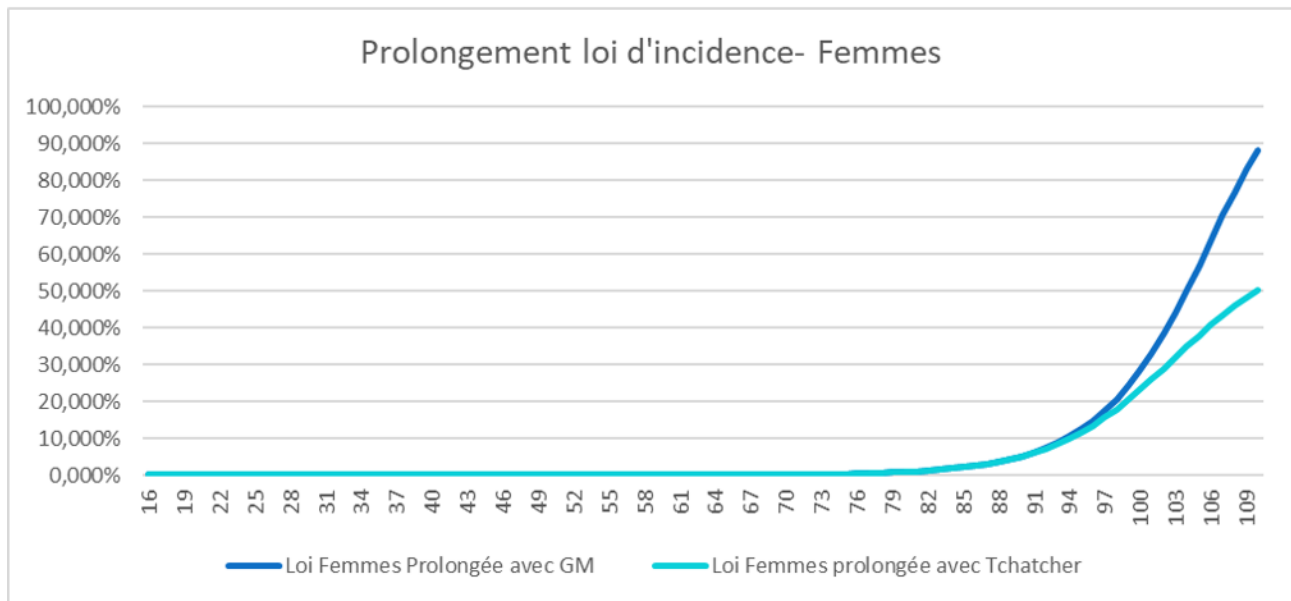


FIGURE 3.17 – Prolongement lois d'incidence Femmes

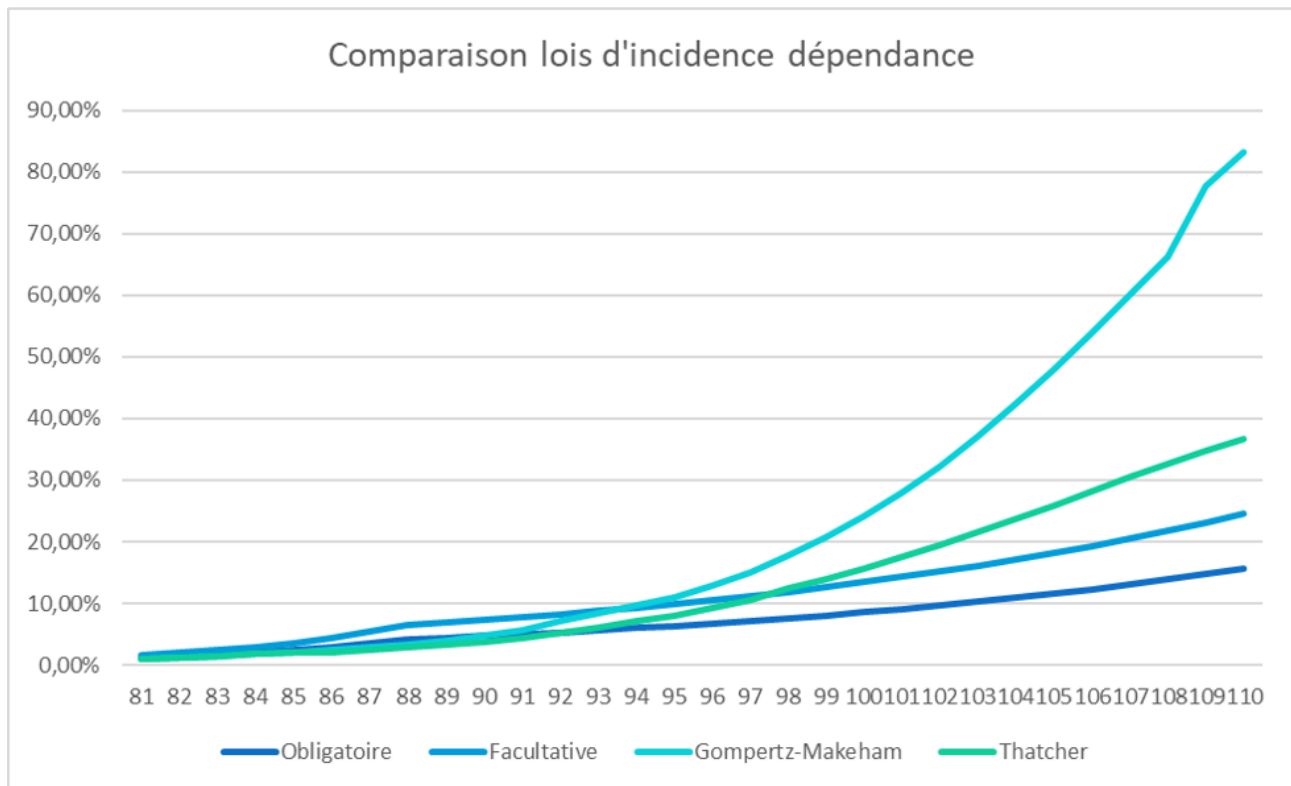


FIGURE 3.18 – Comparaison lois d'incidence avec des lois de tarification internes

Nous notons que le prolongement obtenu par Gompertz-Makeham (que nous avons gardé volontiers à titre illustratif) se démarque de toutes les lois montrant ainsi son inadéquation. Cependant la loi prolongée par la méthode de Thatcher est plus prudente que la loi obliga-

toire et même facultative sur les âges élevés. Entre 80 et 95 ans, la loi facultative est plus prudente que notre loi, cela peut s'expliquer par le risque d'antisélection sur des produits d'assurance facultatives ou à l'inverse de sous-déclaration sur des garanties annexes dans des contrats collectifs obligatoires. La loi d'incidence prolongée par Thatcher est donc celle qui nous paraît plus adéquate au risque d'être trop prudent. Et là encore, le fait qu'elle soit plus prudente que les lois obligatoires et facultatives utilisée en interne peut s'expliquer par le fait que ces lois ne soient pas à jour et que la dépendance soit de plus en plus connue par le grand public et les sinistres déclarées de plus en plus.

Nous avons finalement construit nos lois d'incidence en dépendance pour les hommes et les femmes sur une plage d'âges compris entre 16 et 110 ans. La loi ainsi obtenue a été comparé a deux lois utilisées en tarification en interne. Le résultat est satisfaisant. Cependant, tous nos assurés ne passent pas par l'état de dépendance. Certains décèdent directement en étant valides. Nous allons dans la setion suivante étudier cette mortalité.

3.4 Lois de mortalité des valides

Notre population assurée n'a pas la même mortalité de la population générale du fait qu'elle présente des particularités liées à la catégorie socio-professionnelle ou une répartition par âge et sexe différente entre autres. Ce décalage par rapport à la mortalité générale s'accroît si nous nous focalisons sur nos assurés valides. C'est pour cette raison que nous nous sommes proposés de construire une loi de mortalités des valides. Nous allons estimer les taux bruts entre 60 et 90 ans puis procéder au lissage avant de prolonger les lois obtenus.

3.4.1 Taux bruts

Notons :

- e_x^c l'exposition centrale observée entre les âges x et $x + 1$
- n_x^1 le nombre de décès autonomes observés entre les âges x et $x + 1$

Rappelons que l'exposition centrale est le temps total passé par l'ensemble des individus dans l'état d'autonomie. En effet, les individus dépendants ne sont plus soumis au risque de mortalité des valides.

Nous disposons des informations suivantes :

- La date de naissance notée DN,
- La date d'adhésion notée DA,
- La date de sortie notée DS,
- La date d'entrée en dépendance notée DE,
- La date de décès notée DC, si elle est connue,
- Le motif de sortie notée MS

Le motif de sortie MS peut être :

- 1 : Chute suite à décès avéré enregistré sur les garanties Décès statutaires
- 2 : Chute sans décès avéré sur les garanties Décès statutaires
- 3 : Chute sans adhésion retrouvée à une garantie Décès statutaire
- 4 : Non sorti

En restreignant notre intervalle d'étude entre les dates *Deb* et *Fin*, nous avons pour les n_c cotisants :

- L'âge de début d'observation x_i :

$$x_i = \max(DA_i, Deb) - DN_i$$

- L'âge de fin d'observation y_i :

$$y_i = \min(DS_i, DE_i, Fin) - DN_i$$

Sont décédés en étant de validité les assurés n'étant pas rentré en dépendance et dont une date de décès est connue ou le motif de sortie renseigné comme 1. En effet, pour les personnes dont le décès est avéré, on prend $DC=DS$ si DC inconnue, c'est le cas de tous les assurés valides qui quitte le portefeuille avec un MS renseigné 1.

À partir de là, nous calculons :

-

$$e_x^c = \sum_{i=1}^{n_c} \max[0, \min(y_i, x + 1) - \max(x_i, x)]$$

-

$$n_x^1 = \sum_{i=1}^{n_c} \mathbf{1}_{\{DC < Fin\}} \mathbf{1}_{[x, x+1]}(y_i)$$

Le taux de mortalité des valides à l'âge x s'écrit alors :

$$i_x = \frac{n_x^1}{e_x^c}$$

Nous obtenons ainsi les lois brutes de mortalités des valides.

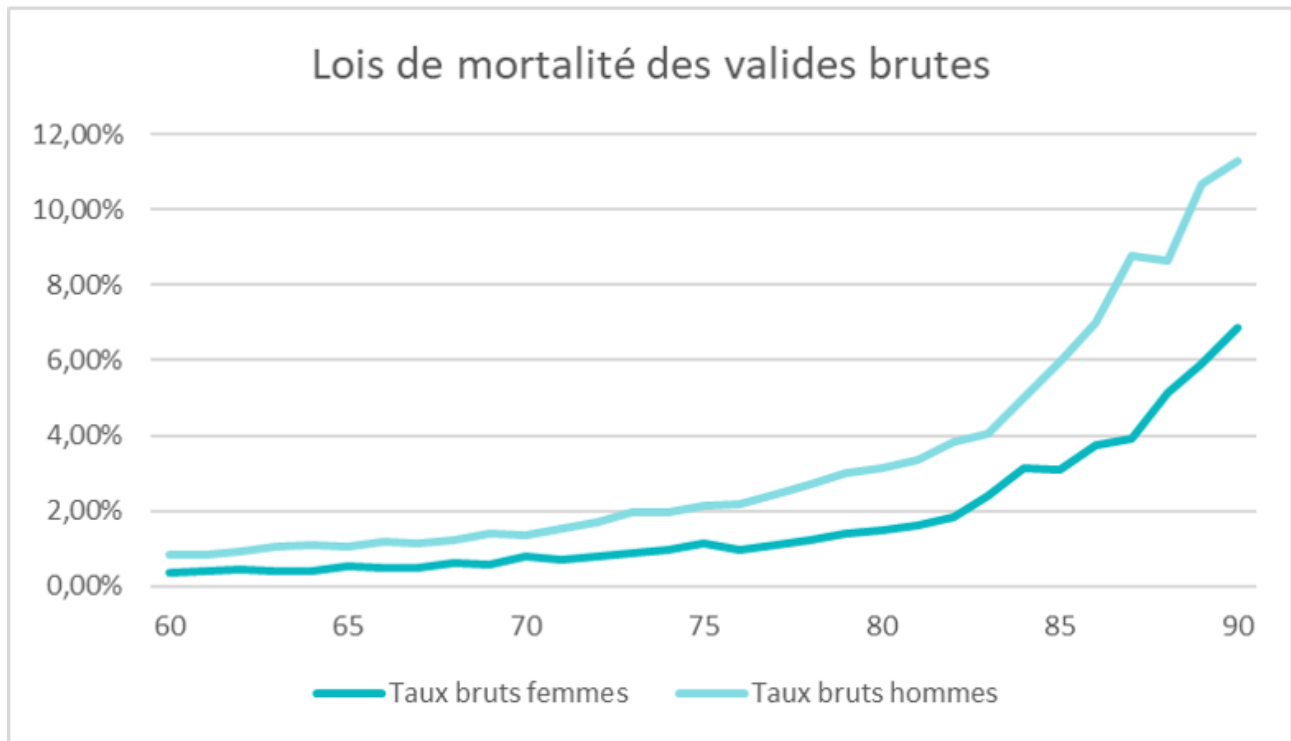


FIGURE 3.19 – Taux bruts de mortalité des valides par âge

La mortalité des valides augmentent avec l'âge de manière quasi-exponentielle. Les taux de mortalité sont plus élevés chez les hommes que chez les femmes avec une accentuation de l'écart avec l'âge.

3.4.2 Lissage

Comme pour les lois d'incidence, nous avons appliqué différentes méthodes de lissage à nos taux bruts. Nous pouvons comparer les résultats obtenus.

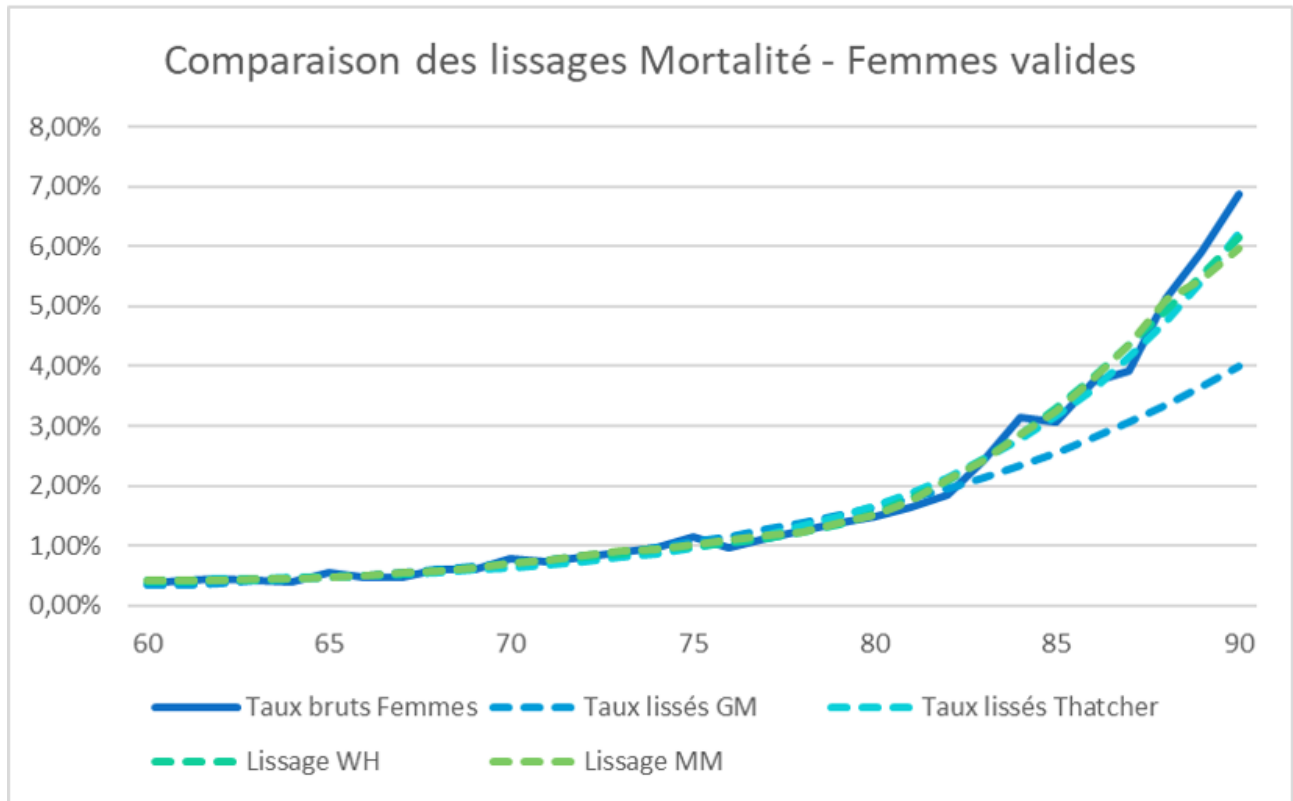
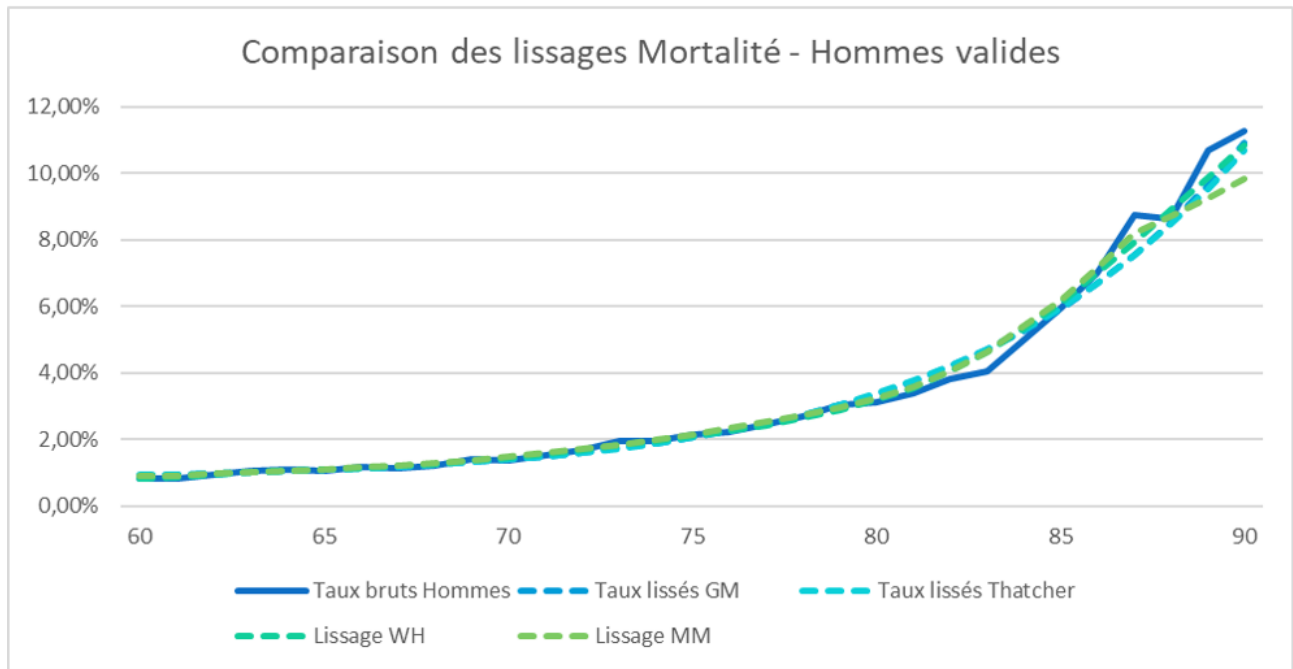


FIGURE 3.20 – Comparaison des lissages sur les taux de mortalité des valides

Globalement les méthodes de lissage donnent des résultats assez similaires. Pour les Hommes les méthodes de Thatcher et des moyennes mobiles se démarquent en sous-estimant particulièrement les taux bruts sur les âges élevés. Chez les femmes le lissage obtenu par Gompertz-Makeham s'éloigne des taux bruts alors que les autres lissage sont très proches. Les critères de fidélité et de régularité ont permis de choisir la méthode de Whittaker-Henderson à la fois pour les femmes et les hommes. Nous avons ensuite fait un backtesting dont les résultats sont les suivants.

Lissages Hommes	Observés/Attendus	Lissages Femmes	Observés/Attendus
Moyennes mobiles	97,41%	Moyennes mobiles	98,64%
Whittaker-Henderson	99,47%	Whittaker-Henderson	99,41%
Gompertz-Makeham	99,13%	Gompertz-Makeham	108,23%
Thatcher	99,06%	Thatcher	92,93%

Nous remarquons qu'au global les méthodes de lissage "éligibles" surestiment légèrement la mortalité des valides et que la méthode de Whittaker-Henderson est effectivement la plus adéquate parmi celles testées.

3.4.3 Extrapolation des lois de mortalité des valides

Nous allons prolonger nos lois de mortalité entre 90 et 110 ans en raison du manque de données robustes sur cette plage. Nous avons remarqué que les méthodes de lissage paramétrique ne fournissaient pas une bonne estimation de la mortalité des valides. Nous allons donc exploiter les tables réglementaires et obtenir notre prolongement à partir de là. Par souci de simplicité et de robustesse, nous allons travailler avec une loi de mortalité unique reconstruite à partir des lois par sexe en respectant la répartition de notre portefeuille. La table réglementaire sera également reconstruite de la sorte en partant des tables par sexe

Nous commençons par abattre la table THFOO-02* ainsi obtenue de 25% et la comparer avec la mortalité de notre portefeuille. On remarque que l'écart entre les deux lois se creuse progressivement avant d'atteindre son maximum autour de 80 ans, ceci constitue la plage la plus exposée à l'entrée en dépendance. A partir de 85 ans l'écart se stabilise relativement jusqu'à 90 ans.

On pourrait s'attendre à ce que l'évolution reprenne par la suite mais cela nous paraît très peu probable. Par prudence on pourrait choisir de maintenir l'écart entre les deux taux à leur niveau constaté à 90 ans. C'est-à-dire considérer que les valides meurent moins que la population générale de la table réglementaire abattue et que l'écart à 90 ans est le même au-delà. Comme compromis, il nous semble pertinent d'utiliser l'écart moyen pour décider du prolongement. La table 75% THF00-02* sera ainsi abattue de 15%. Nous obtenons la loi suivante en comparaison à la table réglementaire.

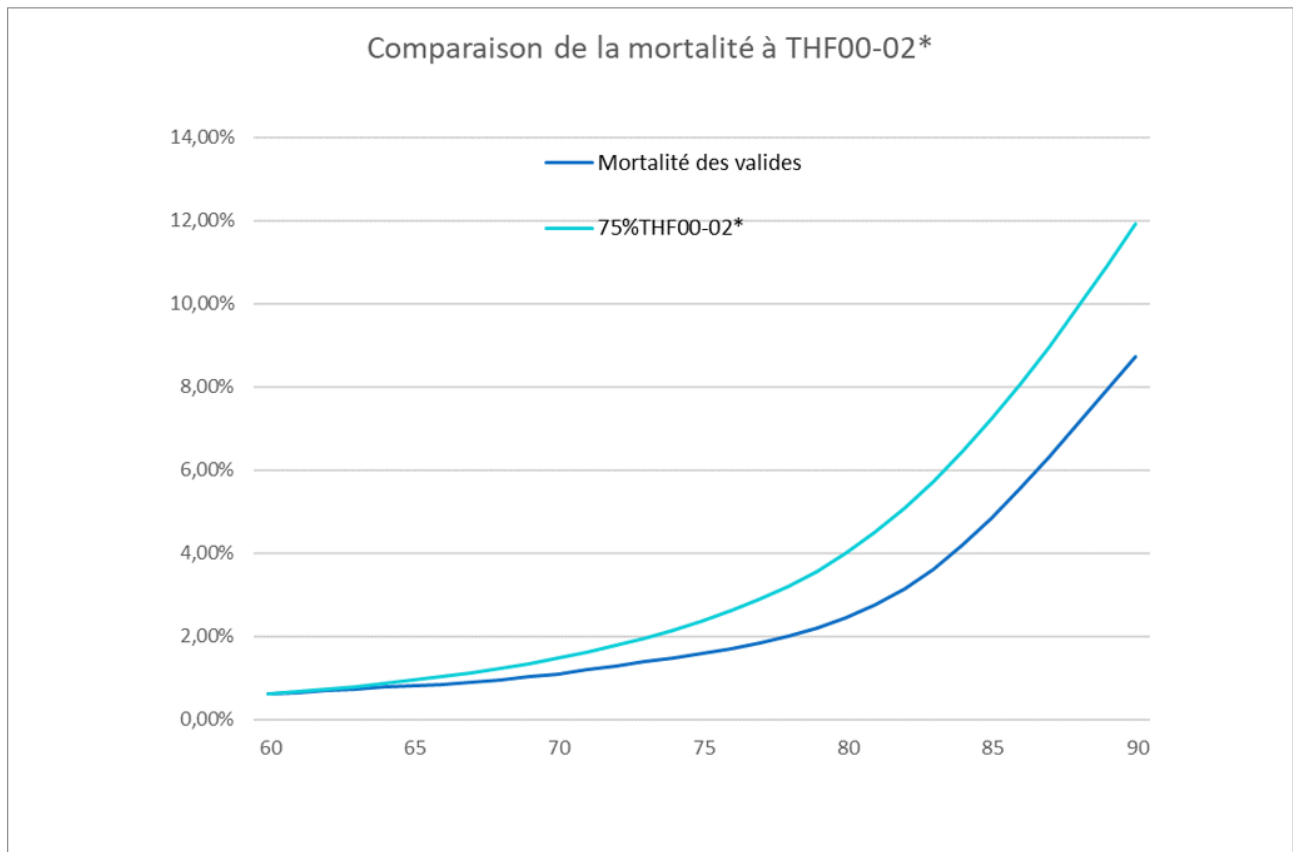


FIGURE 3.21 – Comparaison loi de mortalité des valides et table THF00-02

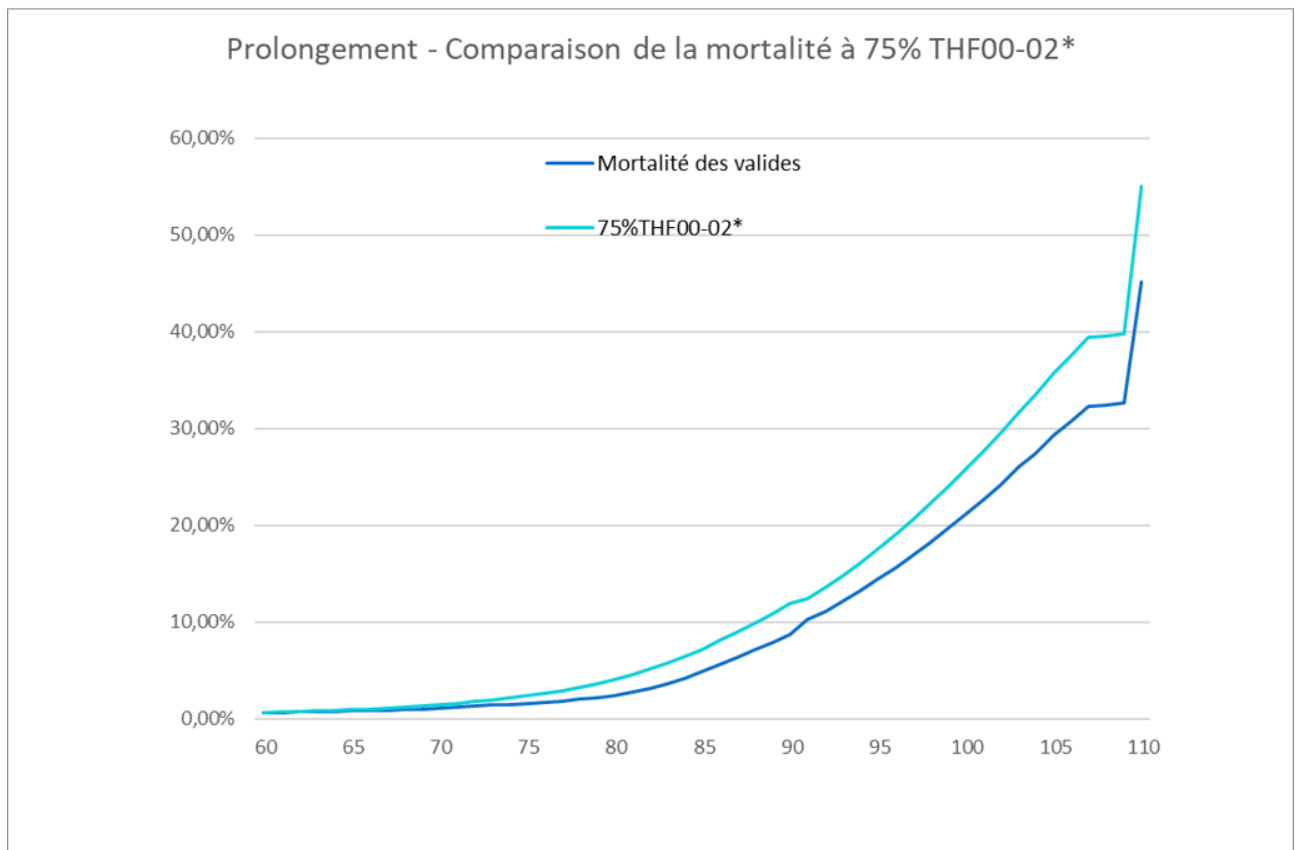


FIGURE 3.22 – Prolongement loi de mortalité des valides

Chapitre 4

Impacts sur le provisionnement

Dans ce chapitre, nous allons mesurer l'impact des lois construites sur la Provision pour Sinistres en Cours (PSC) et Provision pour Risque en Cours (PRC). Chacune de ces provisions sera présentées puis recalculées et l'écart entre le montant obtenu et les comptes sera quantifié vu que nous ne disposons ni des lois de tarification ni des lois de provisionnement utilisé par le principal Réassureur.

4.1 Impact sur la PSC

Abusivement appliquée PM de rentes du fait de sa méthode de calcul, la Provision pour Sinistres en Cours pour but de garantir le versement des rentes des personnes dépendantes auxquelles une indemnisation au moins a eu lieu (dans les deux mois précédant la date de calcul pour notre cas). Elle est calculée tête par tête à la fin de chaque année et on suppose que l'état de dépendance est irréversible, seule la mortalité est une cause de sortie.

Soient :

- ${}_k p_x^a$ la probabilité pour un assuré autonome de rester vivant encore k années sachant qu'il est vivant et autonome à l'âge x
- ${}_k p_x^d$ la probabilité pour un assuré dépendant de rester vivant encore k années sachant qu'il est vivant et dépendant à l'âge x
- i_x la probabilité pour un individu d'âge x d'entrer en dépendance
- l_x^a le nombre de survivants autonomes d'âge x
- l_x^d le nombre de survivants dépendants d'âge x
- w l'âge de fermeture de nos tables
- $v = \frac{1}{1+j}$ le facteur d'actualisation au taux technique j
- R le montant de la rente annuelle en cas de dépendance

Notons a_x^d le capital constitutif de rente annuelle de 1€ payable à terme échu à un assuré dépendant d'âge x en m fractions annuelles.

$$a_x^d = \sum_{k=1}^{w-x} {}_k p_x^d * v^k + \frac{m-1}{2m}$$

Ainsi, pour un assuré entré en dépendance à l'âge x et ayant passé k années dans cet état, la PSC s'écrit :

$$PSC_{x+k} = a_{x+k}^d * R$$

À partir de là nous avons pu quantifier l'impact de la loi de maintien en dépendance sur la PSC.

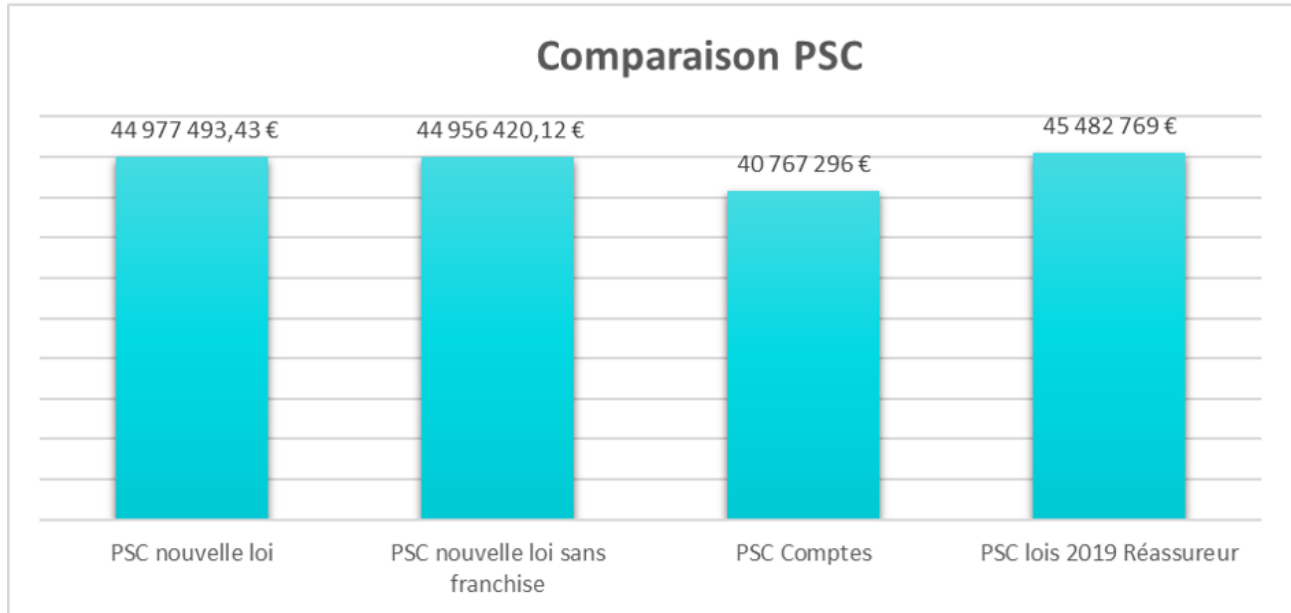


FIGURE 4.1 – Impact de la loi de maintien sur la PSC

Nous remarquons dans un premier temps l'impact de très faible de la franchise sur la PSC, elle permet de réduire la PSC de 0,05%. Cette information pourra permettre de statuer sur sa réelle utilité lors du comité de pilotage.

Les lois à jour du principal Réassureur et notre loi de maintien conduisent à des PSC très proche avec un écart négligeable de 1%. Quant au montant enregistré dans les comptes, notre loi surestime la PSC de 10%.

4.2 Impact sur la PRC

La Provision pour Risques Croissant (PRC) est a pour objectif de pérenniser un équilibre long terme du régime dans un contexte où les cotisations sont constantes dans le temps alors que le risque de survenance des sinistres croît avec l'âge de l'assuré. En effet, les premières années de vie du contrat, alors que l'assureur reçoit des primes plus importantes que le risque réel couvert, un excédent est mis en réserve sous forme de provision qui sera alors utilisée par la suite lorsque la tendance sera inversée.

Autrement dit, la valeur actuelle de l'engagement de l'assuré décroît progressivement avec le paiement successif des cotisations tandis que le risque de survenance de la dépendance (et

donc d'indemnisation) croit très fortement avec l'âge.

La PRC correspond alors à la différence des engagements de l'assureur et de l'assuré. À la souscription du contrat, les valeurs actuelles probables des engagements de l'assureur et de l'assuré sont égales. Au cours du temps, ces valeurs évoluent. Ainsi, le surplus provenant de la différence entre la prime payée et le niveau de risque réel alimente la PRC et sera utilisé lorsque les primes de l'année ne suffisent plus à couvrir le risque de l'année.

Pour un assuré ayant souscrit son contrat à l'âge x et âgé actuellement de $x + k$ ans, la PRC s'écrit :

$$PRC_k(x) = Engagemen_{x+k}^{Assureur} - Engagemen_{x+k}^{Assure}$$

L'engagement de l'assureur correspond au versement d'une rente viagère à l'assuré lorsqu'il rentre en état de dépendance. Quant à l'assuré son engagement consiste au paiement d'une prime annuelle tant qu'il est en vie et reste valide.

En supposant que l'entrée en dépendance survient en milieu d'année

$$a_{x+1/2}^d = \frac{a_x^d + va_{x+1}^d}{2}$$

L'engagement de l'assureur s'écrit alors :

$$Engagemen_{x+k}^{Assureur} = \sum_{k=1}^{w-x} {}_k p_x^a * i_{x+k} * v^k * a_{x+k+1/2}^d$$

Et celui de l'assuré :

$$Engagemen_{x+k}^{Assure} = P(x) \left(\sum_{k=1}^{w-x} {}_k p_x^a * v^k + \frac{m-1}{2m} \right)$$

La PRC s'obtient ainsi en faisant la différence.

En comparant la PRC obtenue d'une part au montant de PRC dans les comptes et d'autres part au montant de PRC estimé par le Réassureur, nous avons les variations suivantes.

Au global les lois de mortalité des valides, d'incidence et de maintien en dépendance sous estiment la PRC estimée par les lois du Réassureur et davantage la PRC des comptes. N'ayant pas les lois utilisés pour le provisionnement ni celles du Réassureur, nous ne pouvons pas quantifier l'impact isolé de la mise à jour de chacune de ces lois. Cependant, vue la PRC estimée assez proche de celle du Réassureur, il nous semble probable que cet écart vienne des lois d'incidence en dépendance et de mortalité des valides.

Par exemple, notre choix de prolongement de la loi d'incidence conduit à surestimer vraisemblablement l'incidence en dépendance sur les âges élevés. Ce qui a pour impact une baisse de l'engagement de l'assuré (qui arrête de cotiser pour étant dépendant) et une hausse de l'engagement de l'assureur (qui serait amené à payer plus de rentes), d'où l'augmentation de la PRC.

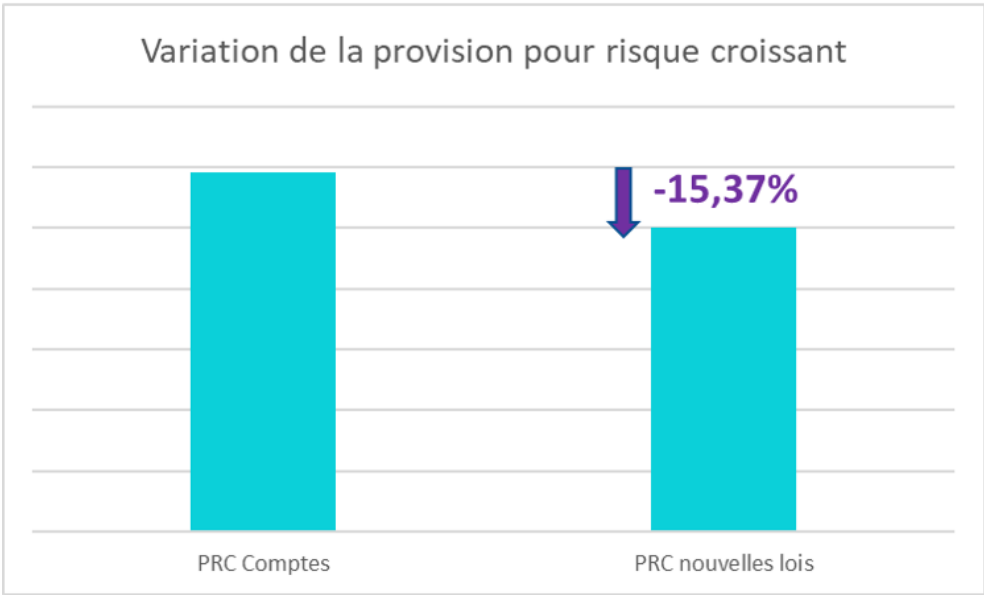
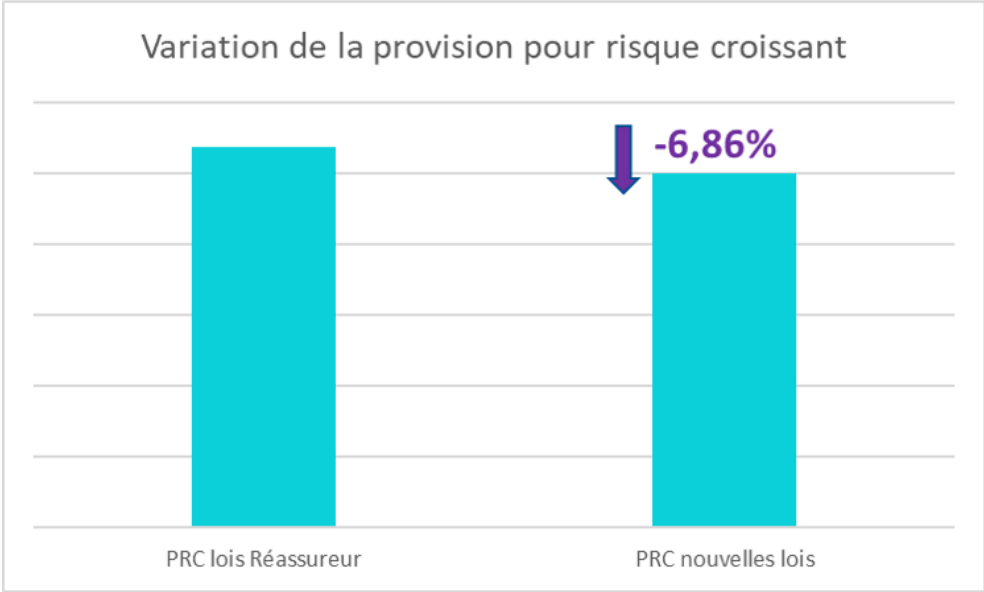


FIGURE 4.2 – Impact des lois d’expérience sur la PRC

Conclusion

Ce mémoire s'est donné pour objectif la construction et l'exploitation de lois d'expérience d'un portefeuille d'assurance dépendance.

Il n'existe pas une mais des définitions de la dépendance ou perte d'autonomie qui varient en fonction des acteurs et des zones géographiques. La prise en charge est tout aussi variée, que ce soit par les pouvoirs publics ou les organismes d'assurance. Face à une prise en charge des pouvoirs publics insuffisantes, les produits d'assurance dépendance se multiplient et la modélisation du risque devient une nécessité. Afin de nous assurer de la bonne qualité et de la pertinence des données utilisées un traitement a été effectué au préalable sur les données brutes en faisant des tests de cohérence et en nous basant sur notre connaissance du risque assuré et du contrat qui régit la garantie. Les statistiques descriptives ont permis de souligner les limites des données notamment l'historique relativement récent des sinistres. Nous avons néanmoins construits nos lois d'expérience sur la base des données retenues et avons utilisés différentes méthodes pour les prolonger.

Nous avons construite une loi de maintien avec l'estimateur de Kaplan Meier sur chacune des classes jugée pertinentes à travers le modèle de Cox et les tests de rapport de maxima de vraisemblance de Wald et du logrank. En raison du manque de données à partir de la cinquième année de dépendance, les lois de maintien ont été retenues sur les quatre premières années puis prolongées en appliquant des taux de majorations aux tables réglementaires. Les coefficients utilisés diminuent avec l'âge en se rapprochant de 1, ce qui revient à dire que la survie des dépendants se rapproche de celle de la population générale avec l'âge.

Ensuite, nous avons estimé des lois d'incidence bruts pour hommes et femmes et avons comparé les méthodes de lissage non paramétriques des Moyennes Mobiles et de Whittaker-Henderson, puis les lissages paramétriques de Gompertz-Makeham et Thatcher sur nos deux lois d'incidence. Les critères de fidélité et de régularité ont permis de faire des choix qui ont été réconforté par des backtestings. Ces lois ont été construites pour des âges compris entre 60 et 85 ans pour les hommes, et pour les femmes entre 60 et 89 ans car au-delà le manque de données rendrait l'estimation non fiable. Elles ont été extrapolées sur les âges avant 60 ans dans un premier temps et ensuite sur les âges au-delà de 85 et 89 ans respectivement pour les hommes et les femmes. Sur les âges inférieurs les deux méthodes testées donnent des résultats assez similaires et nous avons retenues le prolongement obtenu par la méthode de Gompertz-Makeham. À l'inverse sur les âges élevés, le prolongement obtenu par la méthode de Gompertz-Makeham traduit une croissance exponentielle qui ne nous a pas paru adapté à l'incidence en dépendance, notre choix a porté sur le prolongement de la méthode de Thatcher plus adapté à notre connaissance au risque de perte d'autonomie.

Puis, nous avons construit de façon similaire aux lois d'incidence, deux lois de mortalités des assurés autonomes. Les lois bruts ont été lissées avec les mêmes méthodes que les lois d'incidence. Ces lois construites sur une plage d'âges compris entre 60 et 90 ans ont été prolongées en se basant sur l'écart de mortalité observé par rapport aux tables réglementaires TH00-02 et TF00-02 entre 80 et 90 ans.

Enfin, nous avons mesurer l'impact des lois établies sur la Provision pour Sinistres en Cours (PSC) et la Provision pour Risques en Cours (PRC). En effet les lois construites peuvent permettre de projeter le portefeuille mais aussi de calculer des provisions sur la base de l'expérience qu'on a acquiert, avec donc plus de précisions que les lois de tarification. Sur la PSC qui n'est impactée que par la loi de maintien en dépendance, la provision calculée présente un écart négligeable de 1% à la baisse par rapport à la provision estimée avec les derniers lois du Réassureur. Quant à la PRC, elle est impactée par les trois lois construites avec une baisse de 6,86% par rapport à la PRC calculée par le Réassureur avec ses dernières lois. N'ayant pas les lois utilisés par le Réassureur nous ne pouvons pas quantifier l'impact isolé de chacune de nos lois.

Pour aller plus loin, il serait intéressant de mettre à jour ces lois avec un historique de données plus profond. Le lissage et surtout le prolongement des lois pourraient être améliorés en utilisant des techniques d'apprentissage automatique se basant sur des données agrégées sur un ensemble de portefeuilles. Ainsi les données du portefeuilles seront utilisées pour la constructions des lois brutes puis le lissage et le prolongement se fera en utilisant des données plus générales.

Bibliographie

- [1] Institut des ACTUAIRES. *Lignes directrices mortalité de la Commission d’Agrément*. 2006.
- [2] Alexandre ARMBRUSTER. *Estimation du Best Estimate sur le risque dépendance*. Mémoire DUAS, 2010.
- [3] Benjamin BALTESAR. *Construction d’une table de mortalité sur un portefeuille de temporaire décès*. Mémoire ISFA, 2013.
- [4] Rida BEZZA. *Construction de tables d’expériences pour le risque dépendance*. Mémoire ISUP, 2018.
- [5] Guillaume BIESSY. *L’Assurance dépendance*. 2020.
- [6] Guillaume BIESSY. *Modélisation en assurance dépendance*. 2020.
- [7] Roch GIORGI. *Extensions du modèle de Cox : Variables dépendantes du temps, Effets non linéaires*. Faculté de Médecine, Aix-Marseille Université.
- [8] INSEE. *Site de L’Institut national de la statistique et des études économiques*. www.insee.fr. avril 2020.
- [9] Chloé LE GUEN. *Construction de tables de mortalité d’expérience à la CPRPSNCF*. Mémoire ISFA, 2015.
- [10] Dae-Jin LEE. “Smoothing mixed models for spatial and spatio-temporal data”. Thèse de doct. Universidad Carlos III de Madrid, 2010.
- [11] Dominique LIBAULT. *Rapport Concertation Grand âge et autonomie*. Ministère des Solidarités et de la Santé, 2019.
- [12] Ning LIN. *Etudes des lois de Maintien et de passage en dépendance*. Mémoire ISFA, 2012.
- [13] Clive LOADER. *Local Regression and Likelihood*. Springer, 1999.
- [14] Annaël LUZON. *Mise à jour de la loi d’incidence et de maintien d’un contrat dépendance et Impacts sur le provisionnement*. Mémoire ISUP, 2019.
- [15] Frédéric PLANCHET. *Méthodes de lissage et d’ajustement*. Support de cours 2019-2020, 2019.
- [16] Aki QUASHIE et Michel DENUIT. *Modèles d’extrapolation de la mortalité aux grands âges*. Institut des Sciences Actuarielles Institut de Statistique, 2005.
- [17] Claire ROGIER. *Construction de lois d’entrée et de maintien en dépendance en vue du suivi d’un contrat*. Mémoire ISFA, 2015.

- [18] Philippe SAINT PIERRE. *Introduction à l'Analyse des durées de survie*. Université Pierre et Marie Curie, 2015.
- [19] Marie-Bérangère VIVILLE. *Comparaison de méthodes d'ajustement de la mortalité des rentiers dans un but prospectif*. Mémoire ISFA, 2008.

Annexe

Codes R utilisés

```
library(readxl)
library(survival)

Incidence.surv= Surv( Incidences$Age deb observation',incidence$Age fin observation',event = Incidences$Fin observation pour dépendance')
class(Deb)
w=survfit(Incidence.surv~1,Incidence)
plot(w)

survdiff(Surv(Incidence$Age fin observation',incidence$Fin observation pour dépendance')~Incidence$Genre)

for (i in 00:90)
{if }

taux_b=read_excel("C:/Users/bsdiack/Desktop/Taux_bruts_incidence.xlsx")
library(locfit)

#Moyennes mobiles
hommes=unlist(taux_b[,2])
age=unlist(taux_b[,1])

MM1=filter(hommes, rep(1/1,1))
MM1= unlist(MM1)
plot(age,hommes, type = "n", xlab = "Age", ylab = "Taux d'incidence")
lines(age,hommes, col = "blue")
lines(MM1, age, col = "red")
MM1

#####
#Définition de la fonction Whittaker-Henderson#
#####

W_H=fonction(Tbruts,Poids,z,h)#
{
  #Déclaration de la taille des variables#
  p<-nrow(Tbruts)
  W=matrix(0,p,p)
```

```
#####
#Définition de la fonction Whittaker-Henderson#
#####

W_H=fonction(Tbruts,Poids,z,h)#
{
  #Déclaration de la taille des variables#

  p<-nrow(Tbruts)
  W=matrix(0,p,p)
  v=matrix(0,z+1)
  K=matrix(0,p-z,p)
  M=matrix(0,p,p)
  qlisse=matrix(0,p)
  Tlisse=matrix(0,p)
  Ecart=matrix(0,p)

  # Construction du vecteur des poids W#

  {
    for (i in 1:p)
    {
      W[i,i]=Poids[i,1]
    }
  }

  #Construction de la matrice K#
  for (k in 0:z)
  {
    v[(k+1),1]=(-1)^(z-k)*factorial(z)/(factorial(k)*factorial(z-k))
  }

  for (j in 1:(p-z))
  {
    for (i in 1:(z+1))
    {
      K[j,i+j-1]=v[i,1]
    }
  }

  #Calcul des taux lissés#
  M=W+h*t(K)%*%K
  qlisse=solve(M)%*%W%*%Tbruts
```

```

#Calcul des taux lissés#
M=H*t(K)*X
qlisse=solve(M)*%D%*Tbruts

for (i in 1:p)
{
  Tlisse[i,1]=qlisse[i,1]
  Ecart[i,1]=Tlisse[i,1]/Tbruts[i,1]
}

Tlisse<-Tlisse
Ecart<-Ecart

par(mfrow=c(1,2))
plot(Tbruts,xlab="Age",main="Taux Bruts",adj=0.5,font=2,col='red',type="l")
plot(Tlisse,xlab="Age",main="Taux Lissés",adj=0.5,font=2,col='red',type="l")
}

#####
#Import de la matrice à lisser et de la matrice des poids#
#####
library(readxl)
Tbruts=data.matrix(read_excel("C:/Users/bsdiack/Desktop/MTauxBruts.xlsx",col_names=FALSE))
Poids=data.matrix(read_excel("C:/Users/bsdiack/Desktop/poids.xlsx",col_names = FALSE))

#####
#Appel de la fonction (les ordres de lissage et de sensibilité sont fixés)#
#####
W_H(Tbruts,Poids,2,1)
write.table(Tlisse, "c:/Users/bsdiack/Desktop/MTauxLissesMH1.txt", sep="\t")

W_H(Tbruts,Poids,2,10)
write.table(Tlisse, "c:/Users/bsdiack/Desktop/MTauxLissesMH10.txt", sep="\t")

W_H(Tbruts,Poids,2,100)
write.table(Tlisse, "c:/Users/bsdiack/Desktop/MTauxLissesMH100.txt", sep="\t")

W_H(Tbruts,Poids,2,1000)
write.table(Tlisse, "c:/Users/bsdiack/Desktop/MTauxLissesMH1000.txt", sep="\t")

W_H(Tbruts,Poids,2,5)

```

```

W_H(Tbruts,Poids,2,5)
write.table(Tlisse, "c:/Users/bsdiack/Desktop/MTauxLissesMH5.txt", sep="\t")

FTbruts=data.matrix(read_excel("C:/Users/bsdiack/Desktop/MTauxBrutsF.xlsx",col_names=FALSE))
FPoids=data.matrix(read_excel("C:/Users/bsdiack/Desktop/FPoids.xlsx",col_names = FALSE))

#####
#Appel de la fonction (les ordres de lissage et de sensibilité sont fixés)#
#####
W_H(FTbruts,FPoids,2,1)
write.table(Tlisse, "c:/Users/bsdiack/Desktop/FMFTauxLissesMH1.txt", sep="\t")

W_H(FTbruts,Poids,2,10)
write.table(Tlisse, "c:/Users/bsdiack/Desktop/FMFTauxLissesMH10.txt", sep="\t")

W_H(FTbruts,Poids,2,100)
write.table(Tlisse, "c:/Users/bsdiack/Desktop/FMFTauxLissesMH100.txt", sep="\t")

W_H(FTbruts,Poids,2,5)
write.table(Tlisse, "c:/Users/bsdiack/Desktop/FMFTauxLissesMH5.txt", sep="\t")

##### CHUTE #####
W_H(Tbruts,Poids,2,1)
write.table(Tlisse, "c:/Users/bsdiack/Desktop/MTauxLissesMH1.txt", sep="\t")

W_H(Tbruts,Poids,2,10)
write.table(Tlisse, "c:/Users/bsdiack/Desktop/MTauxLissesMH10.txt", sep="\t")

W_H(Tbruts,Poids,2,100)
write.table(Tlisse, "c:/Users/bsdiack/Desktop/MTauxLissesMH100.txt", sep="\t")

W_H(Tbruts,Poids,2,1000)
write.table(Tlisse, "c:/Users/bsdiack/Desktop/MTauxLissesMH1000.txt", sep="\t")

W_H(Tbruts,Poids,2,5)
write.table(Tlisse, "c:/Users/bsdiack/Desktop/MTauxLissesMH5.txt", sep="\t")

```

```
##### Gompertz et Makeham ###
# Lissage par Makeham
# Taux conditionnel de survie : Makeham / h(x)=a+b*c^x
qxMakeham = fonction (a,b,c,x) {1 - exp (-a)* exp( -b/ log (c)*(c -1) *c^x)}
# Ajustement par moindres 'ecarrs

#Ajustement par moindres carrés
getValeursInitiales=fonction(qxBrut,xMin,xMax){
#Détermination des valeurs initiales pour Makeham (on les utilise aussi pour Thatcher)
#On exploite le fait que ln(q(x+1)-q(x)) est aligné sur une droite de pente ln(c)
#Cette fonction n'est pas utilisée pour le modèle de Brass

q=pmax(qxBrut,10^-10) #on enlève les valeurs nulles

x=as.vector(xMin:xMax)
y=log(pmax(abs(q[(xMin+1):(xMax+1)]-q[(xMin+2):(xMax+2)]),10^-10))
t=data.frame(cbind(x,y))

d=lm(y~x,data=t)

p=as.vector(1:3)

p[1]=2*10^-4 #a=Taux accident de base
p[3]=exp(d$coefficients[2]) #c=exp(pente)
p[2]=max(10^-10,-log(p[3])*exp(d$coefficients[1])/(p[3]-1)^2) #b

return(p)
}

getEcart=fonction(qxBrut,expo,xMin,xMax,a,b,c,pond){
#Ecart quadratique entre les taux estimés et les taux bruts

#Pondérations par l'exposition éventuelle
w=pmin(1,as.vector(1:120))
if (pond){expo=expo[1:120]}
w=expo/(qxBrut*(1-qxBrut))
w[(is.na(w))|(is.infinite(w))]=0 #Pour les cas où qxBrut=0 ou 1, on met un poids nul

s=0
```

```
w=expo/(qxBrut*(1-qxBrut))
w[(is.na(w))|(is.infinite(w))]=0 #Pour les cas où qxBrut=0 ou 1, on met un poids nul

s=0
for (x in xMin:xMax){
s=s+w[x+1]*(qxMakeham(a,b,c,x)-qxBrut[x+1])^2
}
s
}

qxAjuste = fonction ( qxBrut ,expo ,xMin ,xMax , pond ){
#eDetermination de l' ajustement
pInitial = getValeursInitiales (qxBrut ,xMin , xMax )
f = fonction (p){ getEcart ( qxBrut ,expo ,xMin ,xMax ,p[1] ,p[2] ,p[3] , pond )}
argmin=constrOptim(pInitial,f,ui=rbind(c(1,0,0),c(0,1,0),c(0,0,1)),ci=c(0,0,0),control=list(maxit=10^3),method="Nelder-Mead")$par
q=as.vector(0:120)
for (x in 0:120) {
q[x +1]= qxMakeham ( argmin [1] , argmin [2] , argmin [3] ,x)
}
return (q)
}

Tbruts=data.matrix(read_excel("C:/Users/bsdiack/Desktop/MTauxBruts.xlsx",col_names=FALSE))
Expo_H=data.matrix(read_excel("C:/Users/bsdiack/Desktop/Expohhh.xlsx",col_names=FALSE))
Result=qxAjuste(Tbruts,Expo_H,60,90,pond = TRUE)
write.table(Result, "c:/Users/bsdiack/Desktop/MortLisseGM.txt", sep="\t")

FTbruts=data.matrix(read_excel("C:/Users/bsdiack/Desktop/MTauxBrutsF.xlsx",col_names=FALSE))
Expo_F=data.matrix(read_excel("C:/Users/bsdiack/Desktop/Fpoids.xlsx",col_names=FALSE))
FResult=qxAjuste(FTbruts,Expo_F,60,90,pond = TRUE)
write.table(FResult, "c:/Users/bsdiack/Desktop/FMortLisseGM.txt", sep="\t")

TResult= TqxAjuste(Tbruts,Expo_H,60,90,pond = TRUE)
write.table(TResult, "c:/Users/bsdiack/Desktop/MortLisseT.txt", sep="\t")

TFResult= TqxAjuste(FTbruts,Expo_H,60,90,pond = TRUE)
write.table(TFResult, "c:/Users/bsdiack/Desktop/FMortLisseT.txt", sep="\t")
```

```

#Taux conditionnel de survie : Thatcher / h(x)=alpha*exp(beta*x)/(1+alpha*exp(beta*x))+gamma
#le paramètre beta correspond à ln(c) dans le modèle de Makeham
qxThatcher=function(a,b,g,x){
  1-exp(-g)*((1+a*exp(b*(x+1)))/(1+a*exp(b*x)))^(-1/b)
}

TgetEcart=function(qxBrut,expo,xMin,xMax,a,b,c,pond){
  #Ecart quadratique entre les taux estimés et les taux bruts

  #Pondérations par l'exposition éventuelle
  w=pmin(1,as.vector(1:120))
  if (pond){expo=expo[1:120]}
  w=expo/(qxBrut*(1-qxBrut))
  w[(is.na(w))|(is.infinite(w))]=0 #Pour les cas où qxBrut=0 ou 1, on met un poids nul

  s=0
  for (x in xMin:xMax){
    s=s+w[x+1]*(qxThatcher(a,b,c,x)-qxBrut[x+1])^2
  }
  s
}

TqxAjuste = function ( qxBrut ,expo ,xMin ,xMax , pond ){
  #eDtermination de l' ajustement
  pInitial = getValeursInitiales (qxBrut ,xMin , xMax )
  f= function (p){ TgetEcart ( qxBrut ,expo ,xMin ,xMax ,p[1] ,p[2] ,p[3] , pond )}
  argmin=constrOptim(pInitial,f,ui=rbind(c(1,0,0),c(0,1,0),c(0,0,1)),ci=c(0,0,0),control=list(maxit=10^3),method="Nelder-Mead")$par
  q=as.vector(0:120)
  for (x in 0:120) {
    q[x +1]= qxThatcher( argmin [1] , argmin [2] , argmin [3] ,x)
  }
  return (q)
}

TResult=TqxAjuste(Tbruts,Expo_H,60,85,pond = TRUE)
TFResult=TqxAjuste(FTbruts,Expo_F,60,89,pond = TRUE)

chisq.test(x=FResult[60:89],y=FTbruts[60:89])

y_Thatc= TResult[60:85]
Tt=data.frame(cbind(x,y_Thatc))
Thatcher_lm=lm(y_Thatc~x, data = t)
summary(Thatcher_lm)

```

```

y= Result[60:85]
t=data.frame(cbind(x,y))
Mak_lm=lm(y~x, data = t)
summary(Mak_lm)

install.packages("AICcmoDavg")
library(AICcmoDavg)

### MAINTIEN

Maintien_dep=read_excel("C:/Users/bsdiack/Desktop/Base_rentiers.xlsx")
Deb=Maintien_dep$`Age entrée en jours`
Fin=Maintien_dep$`Age sortie en jours`
Age=Maintien_dep$`Classe âges`
Maintien_test=read_excel("C:/Users/bsdiack/Desktop/Base_rentiers.xlsx")
Age_entree_dep=Maintien_test$`Classe âges`

AgeDebJours_KM=as.numeric(Deb)
AgeFinJours_KM=as.numeric(Fin)
Sexe= Maintien_dep$Homme

w_testSurv=Surv(time=(AgeFinJours_KM-AgeDebJours_KM)/365.25 , event = Maintien_dep$DC)
w_test= survfit(w_testSurv ~1, data =Maintien_dep , type = "kaplan-meier")
w_test
library(survminer, quietly = TRUE)
ggsurvplot(w_test,data = Maintien_dep, conf.int = TRUE ,risk.table = TRUE,main = "Courbe de survie",submain = "Estimateur de Kaplan-Meier",caption = "Maintien en dépendance")

KM_sexe=survfit(w_testSurv ~ Sexe, data = Maintien_dep,type="kaplan-meier")
ggsurvplot(KM_sexe,data = Maintien_dep, conf.int = TRUE ,risk.table = TRUE,main = "Courbe de survie",submain = "Estimateur de Kaplan-Meier",caption = "Maintien en dépendance par sexe",legend.labs =c("M", "F"))

KM_age=survfit(w_testSurv ~ Age_entree_dep, data = Maintien_dep,type="kaplan-meier")
ggsurvplot(KM_age,data = Maintien_dep, conf.int = TRUE ,risk.table = TRUE,main = "Courbe de survie",submain = "Estimateur de Kaplan-Meier",caption = "Maintien en dépendance", legend.labs=c("[00;89]", "[80;89]"))

cox_global=coxph(Surv(time=(AgeFinJours_KM-AgeDebJours_KM)/365.25 , Maintien_dep$Censure)~ test_age + Homme, data=Maintien_dep)
summary(cox_global)

cox_global_bis=coxph(Surv(time=(AgeFinJours_KM-AgeDebJours_KM)/365.25 , Maintien_dep$Censure)~ test_age*Homme, data=Maintien_dep)
summary(cox_global_bis)
plot(cox.zph(cox_global_bis))
ggcoxzph(cox.zph(cox_global_bis))

```

```

#### Ajustement cox###
cox=coxph(Surv(time=(AgeFinJours_KM-AgeDebJours_KM)/365.25 , Maintien_dep$Censure)~Sexe, data=Maintien_dep)
cox
summary(cox)
res.c=cox.zph(cox)
plot(res.c)

####TEST####
### P doit être supérieur à 5% qui est notre seuil#
test <- cox.zph(test_c)
test

##### AGE ###
cox_age=coxph(Surv(time=(AgeFinJours_KM-AgeDebJours_KM)/365.25 , Maintien_dep$Censure)~ Age_entree_dep , data=Maintien_dep)
cox_age$coefficients
summary(cox_age)
cox_age

ggcoxzph(cox.zph(cox_age))
plot(cox.zph(cox_age))
cox.zph(cox_age)

cox_age_bis=step(cox_age)
ggforest(cox_age,Maintien_dep)

cox_age_test=coxph(Surv(time=(AgeFinJours_KM-AgeDebJours_KM)/365.25 , Maintien_dep$Censure)~ test_age, data=Maintien_dep)
summary(cox_age_test)

cox_test=coxph(Surv(time=(AgeFinJours_KM-AgeDebJours_KM)/365.25 , Maintien_dep$Censure)~ Age+sexe , data=Maintien_dep)
ggforest(cox_test)
plot(cox.zph(cox_age))
plot(cox.zph(cox_age_test))
anova(cox_age,cox_age_test)
plot(basehaz(cox))

Maintien_homme_inf80= subset(Maintien_test, Sexe == "Hommes" & Age_entree_dep=="Moins de 80")
Maintien_homme_sup80= subset(Maintien_test, Sexe == "Hommes" & Age_entree_dep=="Plus de 80")
Maintien_homme_80_89= subset(Maintien_test, Sexe == "Hommes" & Age_entree_dep=="[80;89]")

Maintien_femme_inf80= subset(Maintien_test, Sexe == "Femmes" & Age_entree_dep=="Moins de 80")
Maintien_femme_sup80= subset(Maintien_test, Sexe == "Femmes" & Age_entree_dep=="Plus de 80")
Maintien_femme_80_89= subset(Maintien_test, Sexe == "Femmes" & Age_entree_dep=="[80;89]")

```

```

#Femmes entrées en dep avant 80 ans
W_F_1=subset(w_testSurv, Sexe == "Femmes" & Age_entree_dep=="Moins de 80")
KM_F_1=survfit(W_F_1~1, data = Maintien_femme_inf80,type="kaplan-meier")
ggsurvplot(KM_F_1,data = Maintien_femme_inf80, risk.table = FALSE,main = "Courbe de survie",submain = "Estimateur de Kaplan-Meier",caption = "Maintien en dépendance des femmes entrées en dépendance avant 80 ans")
taux_F_1=getQx(KM_F_1,1,7)
taux_F_1

#Femmes entrées en dep entre 80 et 89 ans
W_F_2=subset(w_testSurv, Sexe == "Femmes" & Age_entree_dep=="[80;89]")
KM_F_2=survfit(W_F_2~1, data = Maintien_femme_80_89,type="kaplan-meier")
ggsurvplot(KM_F_2,data = Maintien_femme_80_89, risk.table = FALSE,main = "Courbe de survie",submain = "Estimateur de Kaplan-Meier",caption = "Maintien en dépendance des femmes entrées en dépendance entre 80 et 89 ans")
taux_F_2=getQx(KM_F_2,1,7)
taux_F_2

#Femmes entrées en dep après 89 ans
W_F_3=subset(w_testSurv, Sexe == "Femmes" & Age_entree_dep=="Plus de 80")
KM_F_3=survfit(W_F_3~1, data = Maintien_femme_sup80,type="kaplan-meier")
ggsurvplot(KM_F_3,data = Maintien_femme_sup80, risk.table = FALSE,main = "Courbe de survie",submain = "Estimateur de Kaplan-Meier",caption = "Maintien en dépendance des femmes entrées en dépendance après 89 ans")
taux_F_3=getQx(KM_F_3,1,7)
taux_F_3

#Hommes entrés en dep avant 80 ans
W_H_1=subset(w_testSurv, Sexe == "Hommes" & Age_entree_dep=="Moins de 80")
KM_H_1=survfit(W_H_1~1, data = Maintien_homme_inf80,type="kaplan-meier")
ggsurvplot(KM_H_1,data = Maintien_homme_inf80, risk.table = FALSE,main = "Courbe de survie",submain = "Estimateur de Kaplan-Meier",caption = "Maintien en dépendance des Hommes entrées en dépendance avant 80 ans")
taux_H_1=getQx(KM_H_1,1,7)
taux_H_1

#Hommes entrés en dep entre 80 et 89 ans
W_H_2=subset(w_testSurv, Sexe == "Hommes" & Age_entree_dep=="[80;89]")
KM_H_2=survfit(W_H_2~1, data = Maintien_homme_80_89,type="kaplan-meier")
ggsurvplot(KM_H_2,data = Maintien_homme_80_89, risk.table = FALSE,main = "Courbe de survie",submain = "Estimateur de Kaplan-Meier",caption = "Maintien en dépendance des Hommes entrées en dépendance entre 80 et 89 ans")
taux_H_2=getQx(KM_H_2,1,7)
taux_H_2

#Hommes entrés en dep après 89 ans
W_H_3=subset(w_testSurv, Sexe == "Hommes" & Age_entree_dep=="Plus de 80")
KM_H_3=survfit(W_H_3~1, data = Maintien_homme_sup80,type="kaplan-meier")
ggsurvplot(KM_H_3,data = Maintien_homme_sup80, risk.table = FALSE,main = "Courbe de survie",submain = "Estimateur de Kaplan-Meier",caption = "Maintien en dépendance des Hommes entrées en dépendance après 89 ans")
taux_H_3=getQx(KM_H_3,1,7)
taux_H_3

```