

**Mémoire présenté pour la validation de la Formation
« Certificat d'Expertise Actuarielle »
de l'Institut du Risk Management
et l'admission à l'Institut des actuaires
le**

Par : Astou Kane DIALLO

Titre : Etude de la mortalité d'un portefeuille de prévoyance collective avec les données de DSN

Confidentialité : NON OUI (Durée : 1an 2 ans)
Les signataires s'engagent à respecter la confidentialité indiquée ci-dessus

Membres présents du jury de l'Institut des actuaires :

Entreprise : AXA France

Nom : Gilles CUVILLIER

Signature et Cachet :

Membres présents du jury de l'Institut du Risk Management :

Directeur de mémoire en entreprise :

Nom : Gilles CUVILLIER

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

Secrétariat :

Bibliothèque :

Signature(s) du candidat(s)

Résumé

Dans un contexte économique et réglementaire exigeant (Taux d'intérêt bas, réforme des retraites, Solvabilité 2...), les assureurs cherchent à allier à la fois rentabilité, compétitivité et croissance de chiffre d'affaires. L'un de leurs enjeux est d'avoir une bonne connaissance de leurs portefeuilles d'assurés pour pouvoir d'une part proposer des tarifs compétitifs, en cohérence avec leurs profils de clients, et d'autre part, mieux maîtriser leurs risques avec une estimation de leurs engagements au plus juste, notamment le capital économique réglementaire requis avec Solvabilité 2.

Dans ce mémoire, nous nous intéressons à la garantie décès d'un portefeuille de prévoyance collective d'entreprises. Cette garantie permet de protéger la famille de l'assuré en compensant sa perte de revenus en cas de décès.

Aujourd'hui, le tarif décès est établi sur la base d'indicateurs moyens de l'ensemble des salariés de l'entreprise (âge moyen, salaire moyen, part d'hommes et de femmes, nombre moyen d'enfants, etc.) et sur l'utilisation de tables de mortalité réglementaires avec des abattements adaptés.

L'objectif de notre étude est de construire des tables de mortalité d'expérience segmentées par sexe et catégorie socio-professionnelle (CSP) pour challenger les tables de tarification actuelles et, améliorer les hypothèses « Best Estimate » pour le risque de catastrophe en vie dans le cadre de solvabilité 2.

Les données utilisées sont celles des DSN (déclarations sociales nominatives) qui permettent à la fois de disposer de l'ensemble des assurés exposés au risque de décès mais aussi ceux sinistrés. Elles contiennent plusieurs informations provenant des fiches de paie des salariés des entreprises assurés et reflètent les événements survenus dans le mois (fin de contrat de travail, décès, démission, arrêt de travail...).

Dans une première étape, nous analyserons, la mortalité du portefeuille en estimant les taux bruts de mortalité avec les méthodes de Hoem et Kaplan Meier. Le modèle de Cox permettra d'évaluer l'influence du sexe et de la catégorie socio-professionnelle sur la mortalité avec la définition d'un groupe de référence.

Un modèle adéquat sera choisi pour la mortalité de ce groupe de référence après le lissage de ses taux bruts par la méthode de Whittaker-Henderson et l'ajustement à une loi de Makeham. Les taux de mortalité des autres groupes en seront déduits avec les coefficients estimés par le modèle de Cox.

La deuxième étape consistera à comparer les tables de mortalité d'expérience avec celles utilisées actuellement pour la tarification et, à proposer une optimisation de la tarification du décès à la maille groupe avec l'application de correctifs « âge actuariel ».

En dernière étape, nous évaluerons l'impact de l'utilisation de ces tables d'expérience segmentées dans le calibrage du choc du risque de catastrophe vie et le calcul du capital économique réglementaire associé.

Mots clés : Prévoyance collective, décès, DSN, table de mortalité, catégorie socio-professionnelle, Kaplan Meier, Cox, Whittaker-Henderson, Makeham, solvabilité 2

Abstract

In a tense economic and regulatory context (low interest rates, pension reform, Solvency 2, etc.), insurers are seeking to combine profitability, competitiveness, and revenues growth. One of their challenges is to reach a good knowledge of their insured portfolios, to offer competitive prices on one hand, in line with their customer profiles, and on the other hand, to better control their risks with a better estimate of their liabilities, in particular, the economic capital required with Solvency 2.

In this thesis, we are interested in death guarantee for a group protection portfolio. This guarantee protects the insured's family by compensating its loss of income in the event of death of the insured.

Today, death pricing is established on average indicators for all company's employees (average age, average salary, proportion of men and women, average number of children, etc.) and on regulatory mortality tables with appropriate adjustments.

The aim of our study is to build experience mortality tables segmented by sex and socio-professional category to challenge the current pricing tables and, to improve the Best Estimate assumptions for the life CAT risk, in the context of solvency 2.

The data used come from DSN ("déclaration sociale nominative") which allow to have the exposure to death risk (all insured) and the death claims. They contain several information arising from the pay slips and reflecting the events that have occurred during the month (end of employment contract, death, resignation, disability, etc.).

In a first step, we will analyze the mortality of the portfolio by estimating the mortality rates with the methods of Hoem and Kaplan Meier. The Cox model will allow to assess the influence of sex and socio-professional category on mortality with the definition of a reference group.

An adequate model will be chosen for the mortality of this reference group after smoothing its mortality rates by the Whittaker- Henderson method and fitting to a Makeham law. The mortality rates of the other groups will be derived therefrom with the Cox model estimated coefficients.

The second step will consist of comparing the experience mortality tables with those currently used for pricing and, to propose an optimization of death pricing at insureds group level with the application of actuarial age adjustments.

As a final step, we will assess the impact of using these segmented experience tables for the calibration of life CAT risk shock and calculate the associated regulatory economic capital.

Keywords: group protection, death, DSN, mortality table, socio-professional category, Kaplan Meier, Cox, Whittaker-Henderson, Makeham, solvency 2

Synthèse

Face à l'environnement économique actuel (taux d'intérêt bas, conséquences de la crise sanitaire...) et réglementaire, marqué par de nombreuses réformes (Solvabilité 2, réforme des retraites, ANI, réforme 100% santé...), l'équilibre entre rentabilité, risque et compétitivité est une équation complexe à résoudre pour les assureurs.

Dans ce contexte, la connaissance de la population assurée est essentielle pour optimiser la tarification, et estimer au plus juste les engagements de l'assureur notamment le niveau du capital réglementaire requis dans le cadre de solvabilité 2.

Sur le périmètre de la prévoyance collective, l'assureur dispose généralement d'informations peu granulaires lors de la souscription d'affaires nouvelles. En particulier, pour le risque décès, la tarification est basée sur des indicateurs moyens du groupe d'assurés (âge moyen, salaire moyen, nombre d'enfants moyen etc) et sur l'utilisation de tables de mortalité réglementaires ajustées avec des coefficients d'abattements.

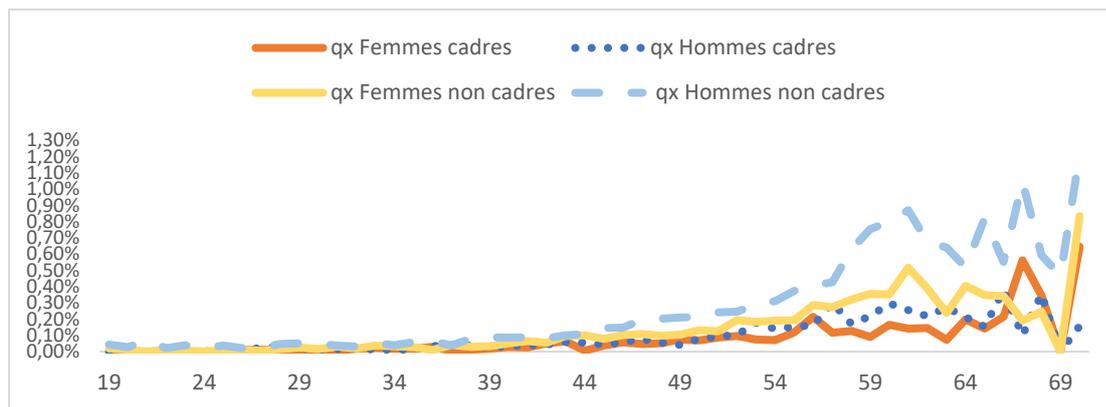
L'objectif de ce mémoire est d'étudier la mortalité d'un portefeuille de prévoyance collective d'entreprises avec la construction de tables de mortalité d'expérience, segmentées par sexe et catégorie socio-professionnelle à partir des données de DSN (déclaration sociale nominative).

En effet, la mise en place de la DSN, obligatoire pour les employeurs depuis 2017, a permis la collecte progressive de données détaillées concernant les salariés des entreprises en portefeuille, facilitant ainsi l'évaluation des risques.

L'étude s'articulera autour des axes suivants :

- Estimation des taux de mortalité par âge, sexe et catégorie socio-professionnelle (CSP)
- Comparaison des tables de mortalité d'expérience avec les tables de tarification actuelles
- Optimisation du tarif décès à la maille groupe avec l'introduction de correctifs « âge actuariel »
- Estimation de l'impact de l'utilisation de tables de mortalité segmentées dans le calibrage du choc du risque catastrophe et le calcul du capital économique réglementaire associé.

La première étape est d'estimer les taux bruts de mortalité par sexe et catégorie socio-professionnelle avec la méthode paramétrique de Hoem pour les âges allant de 19 ans à 70 ans. Ces âges sont les plus représentatifs et correspondent généralement aux âges d'entrée et de sortie de la vie active. Les résultats obtenus indiquent que les taux de mortalité des hommes non-cadres sont largement supérieurs à ceux des autres groupes notamment sur les âges de 45 à 70 ans. Ils sont en moyenne 2 fois plus élevés que ceux des hommes cadres et des femmes non-cadres. Les taux de mortalité observés sont très volatiles de 60 à 70 ans dû aux faibles expositions sur ces âges.



Taux de mortalité par sexe et CSP estimé par Hoem

Ces résultats conduisent à étudier l'influence du sexe et de la catégorie socio-professionnelle sur la mortalité du portefeuille avec le modèle relationnel de Cox. La mortalité de référence sera celle des hommes non-cadres car plus représentatifs en termes de données, à la fois sur l'exposition et le nombre de décès.

Les coefficients suivants sont obtenus avec le modèle de Cox :

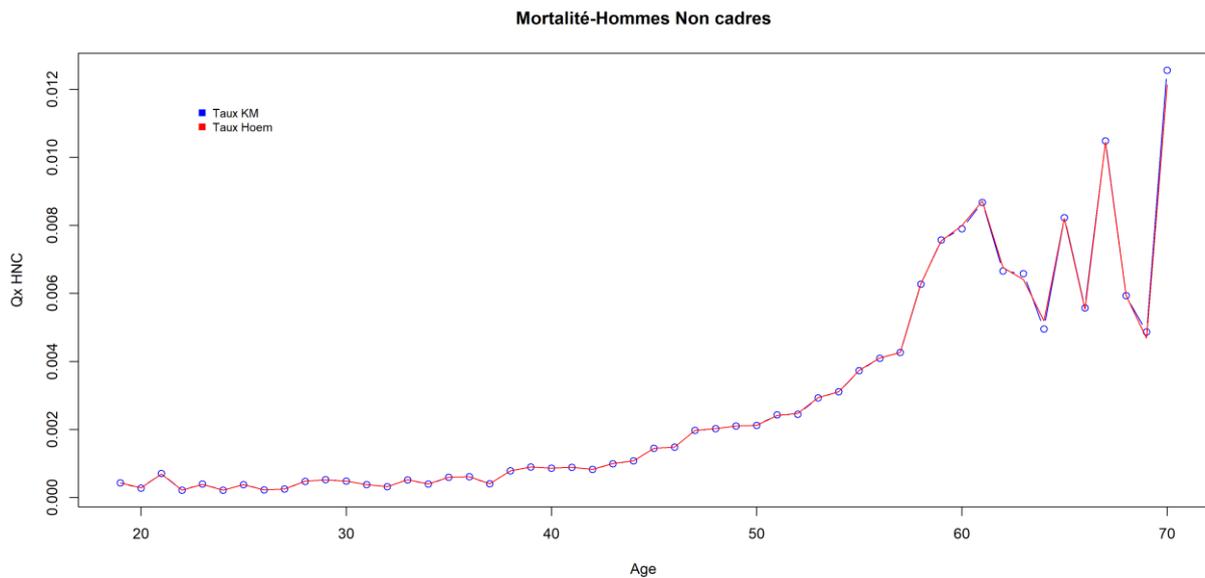
Variables explicatives	θ_0 estimé	exp (θ_0) estimé	Ecart-type	Probabilité critique p-value
femme	-0.53	0.59	0.030	<2e-16
cadre	-0.87	0.42	0.037	<2e-16

Leur interprétation indique que la fonction de survie des hommes non-cadres décroît plus rapidement que celle des autres groupes. Elle montre également que les femmes ont des probabilités conditionnelles de décès inférieures à celles des hommes d'environ 41 % (1-0.59) et les cadres ont des probabilités conditionnelles de décès inférieures à celles des non-cadres d'environ 58% (1-0.42).

Les différents tests de significativité des coefficients (rapport de vraisemblance, score, Wald) permettent de conclure que le vecteur de coefficient θ_0 estimé par le modèle de Cox est non nul et significatif. De plus, les hypothèses sous-jacentes du modèle de Cox notamment celle de hasard proportionnel est vérifiée avec le test des résidus de Schoenfeld.

La deuxième étape consiste à choisir un modèle pour la mortalité des hommes non-cadres et d'en déduire celles des autres groupes (hommes cadres, femmes cadres, femmes non-cadres) à partir des coefficients du modèle de Cox. Pour ce faire, nous avons estimé les taux bruts de mortalité des hommes non-cadres avec la méthode non paramétrique de Kaplan Meier puis les avons comparés à ceux obtenus avec la méthode paramétrique de Hoem.

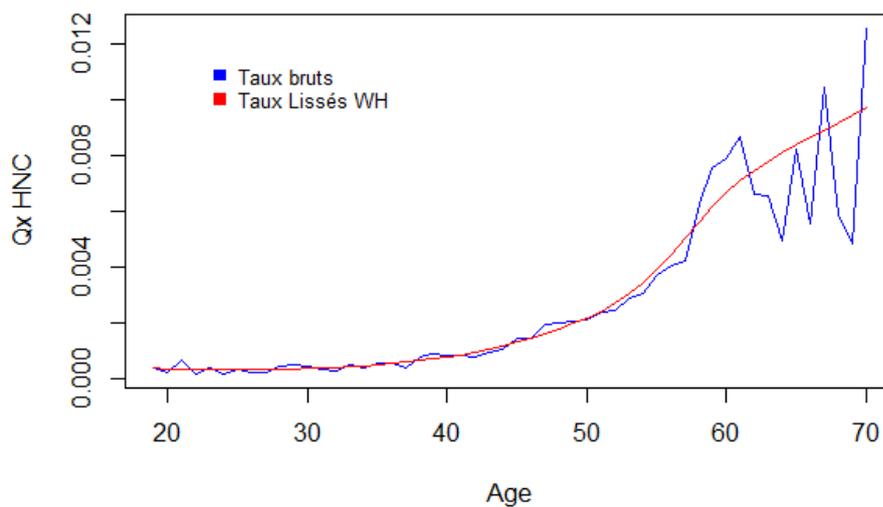
Les deux courbes sont presque confondues comme le montre le schéma ci-après :



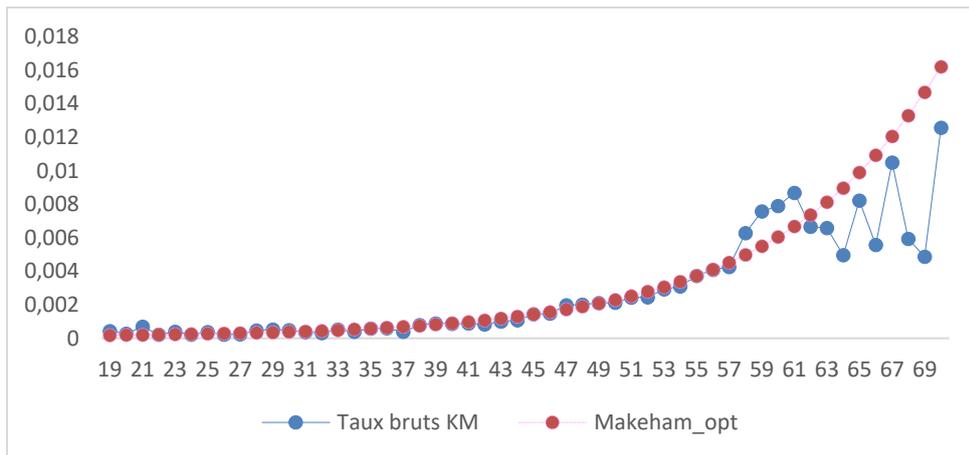
Kaplan Meier vs Hoem pour les hommes non-cadres

Toutefois, nous allons retenir le modèle de Kaplan Meier car non paramétrique et moins couteux en termes d'hypothèses. Les taux de mortalité obtenus par Kaplan Meier sont ensuite lissés afin de les rendre plus réguliers et croissants, sans s'éloigner des données brutes. Les méthodes de lissage utilisées sont celles de Whittaker-Henderson et l'ajustement à une loi de Makeham.

Lissage Whittaker Henderson



Lissage par Whittaker-Henderson (h=1, z=2)



Ajustement des taux bruts au modèle de Makeham pour les hommes non-cadres

Les différents tests statistiques appliqués (Test de changement de signe, test SMR, intervalle de confiance) permettent de valider l'estimation des taux de mortalité par la méthode non paramétrique de Kaplan Meier, qu'ils soient lissés par Whittaker-Henderson ou ajustés à la loi de Makeham. Cependant, nous ferons le choix du modèle totalement non paramétrique (Kaplan Meier lissé par Whittaker-Henderson).

En effet, même si globalement le modèle de Makeham s'ajuste bien aux données, nous remarquons une sous-estimation de la mortalité sur les âges très jeunes et une forte sur estimation sur les âges au-delà de 50 ans du fait de la structure du modèle qui contient un paramètre exponentiel dépendant de l'âge.

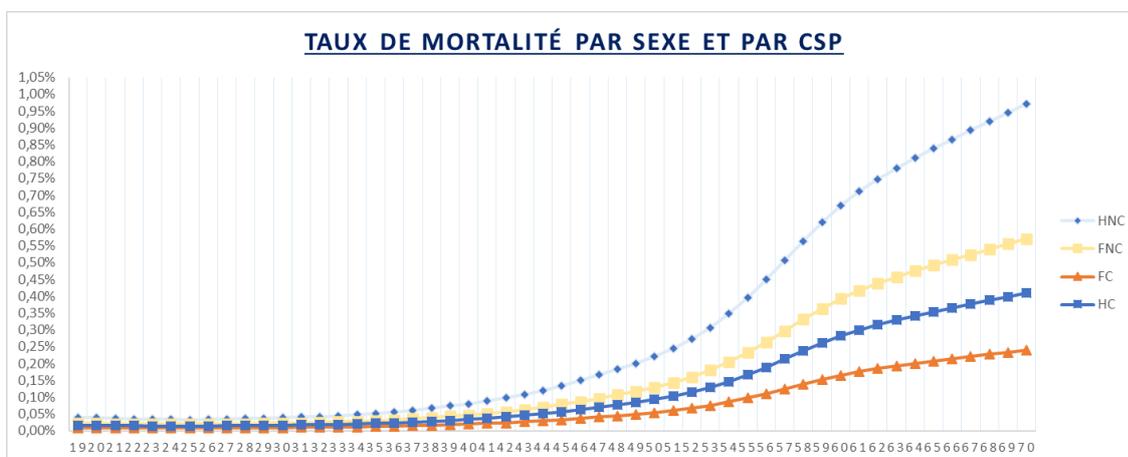
La formule suivante permet de déduire les taux de mortalité des autres groupes :

Pour chaque groupe i , avec $i = 2, 3, 4$ et chaque âge x :

$$q_x^i = 1 - \frac{S_i(x+1)}{S_i(x)} = 1 - \left(\frac{S_1(x+1)}{S_1(x)} \right)^{\exp(\theta_0^T Z^i)} = 1 - (1 - q_x^1)^{\exp(\theta_0^T Z^i)}$$

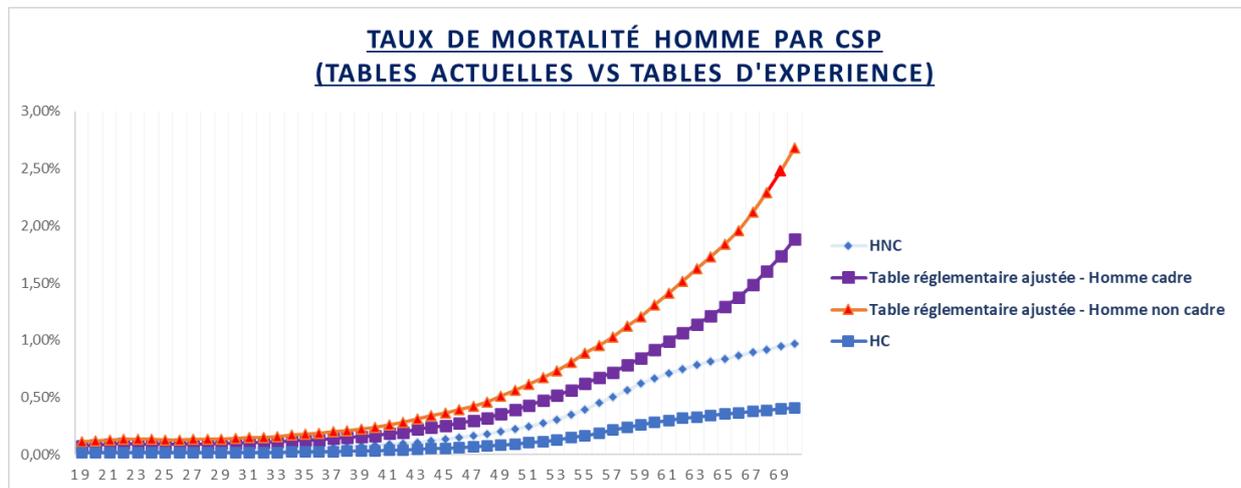
Avec q_x^1 la mortalité de notre groupe de référence qui est celle des hommes non-cadres et $Z_1^i = 1$ si sexe= femme et 0 sinon et $Z_2^i = 1$ si CSP=cadre et 0 sinon.

Nous obtenons les résultats suivants :

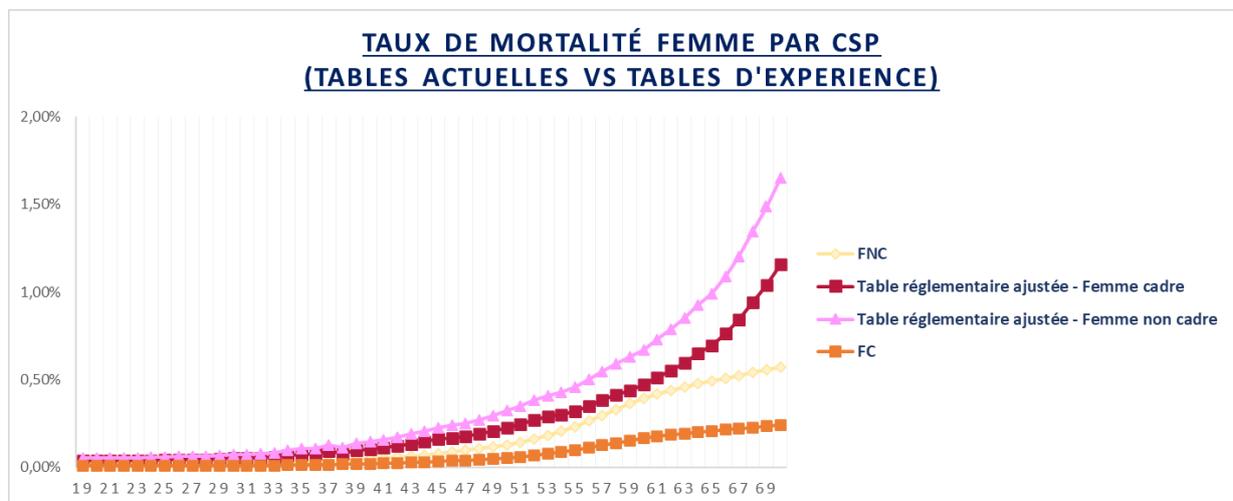


Taux de mortalité par sexe et par CSP retenus

Dans la troisième étape, nous comparons les tables de tarification actuelles (tables réglementaires ajustées avec des coefficients d'abattements) et les tables d'expérience. Cette comparaison montre que les taux de mortalité estimés par ces dernières sont largement en dessous de ceux appliqués pour le tarif. Ils sont en moyenne inférieurs de 60% chez les non-cadres et 70% chez les cadres.



Comparaison tables d'expérience vs tables réglementaires (Hommes)



Comparaison tables d'expérience vs tables réglementaires (Femmes)

De plus, nous remarquons que la mortalité des tables réglementaires ajustées pour les cadres est au-dessus de celle des non-cadres obtenue avec les tables d'expérience à la fois chez les hommes et les femmes. Ce qui indique une forte sur estimation de la mortalité avec les tables actuelles.

Toutefois, d'un point de vue opérationnel, des correctifs commerciaux sont appliqués aujourd'hui au tarif décès final pour tenir compte de l'environnement concurrentiel.

L'utilisation de ces tables d'expérience devrait générer un gain opérationnel avec la suppression de ces correctifs (simplification des process) et permettra d'améliorer notre positionnement concurrentiel avec un tarif au plus proche du risque.

En complément, puisque nous étudions des contrats collectifs, nous proposons d'optimiser le tarif à la maille groupe. Pour un individu d'âge x , la prime pure du décès est donnée par la formule suivante :

$$prime = q_x * Capital\ garanti .$$

Le capital garanti correspond au montant des engagements à la charge de l'assureur en cas de décès et dépend de plusieurs paramètres dont le salaire, la situation familiale, le nombre d'enfants à charge, la catégorie socio-professionnelle, etc.

Pour les contrats de prévoyance collective d'entreprise, la tarification est réalisée sur la tête moyenne du groupe (âge moyen arithmétique, salaire moyen, nombre d'enfants moyen etc). Cependant, il est important de prendre en compte la répartition des âges des salariés et des capitaux garantis associés pour éviter de biaiser la tarification. Pour cela, nous introduisons les correctifs « âge actuariel ».

L'âge actuariel est l'âge moyen x pour lequel la prime moyenne du groupe assuré est la plus proche du « juste prix » des risques engagés. Il est différent de l'âge moyen arithmétique.

Exemple de tarification avec prise en compte de l'âge actuariel

	Groupe 1			Groupe 2			Groupe 3		
	Ind1	Ind2	Ind3	Ind1	Ind2	Ind3	Ind1	Ind2	Ind3
Age	25	25	25	20	20	35	19	26	31
q_x	0,031%	0,031%	0,031%	0,026%	0,026%	0,057%	0,025%	0,033%	0,043%
Capital garanti en K€	63	63	63	32	32	144	28	72	114
Prime individuelle	19,73	19,73	19,73	8,21	8,21	81,67	7,19	23,31	48,60
Age moyen du groupe	25			25			25		
Prime moyenne du groupe sans prise en compte des spécificités individuelles	19,73			19,73			19,73		
Prime moyenne du groupe avec prise en compte des spécificités individuelles	19,73			32,70			26,37		
Age pour lequel la prime est la plus proche de la prime moyenne du groupe avec prise en compte des spécificités individuelles (ou Age actuariel)	25			29			28		

Dans le tableau ci-dessus, nous avons 3 groupes d'assurés avec un âge moyen arithmétique de 25 ans. Cependant nous constatons que la prime n'est pas la même en prenant en compte les spécificités de chaque âge individuel (q_x , capitaux garantis) dans chaque groupe. En effet, la prime moyenne du groupe ajustée des effets individuels est plus proche de celle d'un groupe d'âge moyen 29 ans pour le groupe 2 et 28 ans pour le groupe 3 au lieu de 25 ans. L'âge actuariel pour le groupe 2 est alors de 29 ans et celui du groupe 3 de 28 ans.

Pour déterminer les correctifs « âge actuariel » à appliquer dans la tarification, nous allons nous appuyer sur l'expérience du portefeuille. La méthode consiste, dans un premier temps, à construire un tableau de répartition des effectifs par âge, pour chaque âge moyen x , à partir des données des entreprises assurées. Celui-ci est complété par les capitaux moyens garantis par âge, sexe et catégorie socio-professionnelle afin d'avoir un profil moyen d'entreprise pour chaque âge moyen x .

Puis, le tarif est calculé en faisant la moyenne pondérée des coûts individuels considérant les particularités (capitaux garantis par sexe et CSP) pour chaque âge moyen x .

Enfin, l'âge actuariel est obtenu en recherchant l'âge pour lequel la prime est la plus proche de la prime moyenne du groupe calculée. Autrement dit, on recherche l'âge y pour lequel $q_y * Capital\ garanti_y \approx$ moyenne pondérée des coûts individuels.

Ces réajustements conduisent à augmenter le tarif de façon plus ou moins importante selon les âges moyens x (moins sur les âges jeunes et élevés et en moyenne + 46% chez les hommes non-cadres, +32% chez les femmes non-cadres, + 62% chez les hommes cadres, +46% chez les femmes cadres).

Nous obtenons les correctifs suivants par sexe et par CSP :

Age	Correctif âge actuariel moyen non-cadre	
	H	F
23 ans et moins	0	0
24 à 27 ans	3	2
28 à 32 ans	5	5
33 à 40 ans	6	5
41 à 43 ans	6	4
44 à 46 ans	5	3
47 à 50 ans	4	3
51 à 54 ans	2	1
55 ans et plus	0	0

Age	Correctif âge actuariel moyen cadre	
	H	F
23 ans et moins	0	0
24 à 27 ans	3	2
28 à 32 ans	6	5
33 à 40 ans	7	6
41 à 43 ans	6	6
44 à 46 ans	5	5
47 à 50 ans	4	3
51 à 54 ans	2	1
55 ans et plus	0	0

En cas de nouvelle souscription, une entreprise d'âge moyen x aura pour q_x celui de l'âge moyen $x +$ correctif âge actuariel.

Exemple : une entreprise composée à 100% d'hommes non-cadres d'âge moyen 40 ans se fera appliquer le q_x de 46 ans pour la tarification.

La dernière étape consiste à utiliser les tables d'expérience segmentées dans la calibration du choc « CAT mortality » pour le module de risque de catastrophe en vie. Ce choc correspond à une hausse additionnelle de la mortalité sur la première année de projection. Une étude interne indépendante de ce mémoire a permis d'estimer ce taux à 0,08% (hausse de q_x).

Cependant, sur le périmètre de la prévoyance collective, le modèle interne de l'entreprise fonctionne sous forme de S/C (ou *Loss ratio*) et la hausse additionnelle de 0,08% de mortalité doit être converti en impact S/C .

Ainsi, la formule retenue pour traduire ce choc est :

$$c\% = \left(\frac{q_x + \text{choc}}{q_x} \right) - 1,$$

q_x est le taux de mortalité moyen best estimate calibré sur le portefeuille.

Une hausse additive de 0,08% des taux de mortalité correspond à une augmentation de $c\%$ du loss ratio. Le S/C choqué correspond au $S/C_{de\ base} \times (1 + c\%)$.

Exemple : pour $q_x = 0,20\%$, nous avons $c\% = \left(\frac{0,20\% + 0,08\%}{0,20\%} \right) - 1 = 40\%$.

Le choc appliqué au S/C de base sera donc de 40%.

Aujourd'hui, le calcul du q_x moyen sur le portefeuille est basé sur des tables de mortalité prenant en compte uniquement l'effet âge et sexe des assurés et est estimé à 0,15%.

$$q_x \text{ moyen } BE = \sum_{x=age_{min}}^{x=age_{max}} \frac{CSR_x^H \times q_x^H}{CSR_{Total}^H + CSR_{Total}^F} + \frac{CSR_x^F \times q_x^F}{CSR_{Total}^H + CSR_{Total}^F},$$

avec $CSR = \text{capitaux garantis sous risque}$.

Le choix du q_x moyen BE se justifie par la relation entre celui-ci et la charge totale de sinistres S représentant le numérateur dans le calcul du S/C . Lorsque les paramètres sont bien calibrés,

$$S = q_x \text{ moyen } BE \times CSR_{Total}.$$

Pour affiner le calcul du choc « CAT mortality », nous allons recalculer le q_x moyen BE dans le calibrage du choc en prenant en compte la segmentation fine des q_x par CSP obtenus dans notre étude. La formule de calcul du q_x moyen BE est réécrite sous la forme suivante :

$$q_x \text{ moyen } BE = \frac{1}{CSR_{Total}} \sum_{x=age_{min}}^{x=age_{max}} CSR_x^{HC} \times q_x^{HC} + CSR_x^{FC} \times q_x^{FC} + CSR_x^{HNC} \times q_x^{HNC} + CSR_x^{FNC} \times q_x^{FNC},$$

Avec $CSR_{Total} = CSR_{Total}^{HC} + CSR_{Total}^{FC} + CSR_{Total}^{HNC} + CSR_{Total}^{FNC}$.

La prise en compte de l'effet catégorie socio-professionnelle conduit ainsi à un q_x moyen BE plus faible estimé à 0,13%. Ce qui entraîne une hausse mécanique du choc traduit en S/C de 16%.

A iso S/C , nous obtenons les montants de STEC¹ estimés suivants :

¹ STEC= Short Term Economic Capital → équivalent du SCR (solvency capital requirement) dans le modèle interne AXA France

Estimation des montants de STEC pour le risque « CAT mortality »

	Tables actuelles	Nouvelles tables d'expérience avec prise en compte de la CSP
Choc additif	0,08%	0,08%
Qx moyen BE	0,15%	0,13%
Choc traduit en impact S/C	54%	62%
Variation		16%
Montant estimé de STEC (en M€)	238	276
Variation (en M€)		38

Nous remarquons que la variation du STEC est sensible à la variation du q_x moyen BE. Une baisse de 0,02 point de ce paramètre fait augmenter le choc de 16% et le montant du STEC du même ordre de grandeur à iso « Loss ratio ».

Cependant, compte tenu du poids très réduit du STEC « CAT mortality » sur le STEC total vie et des effets de diversification entre risques (matrice de corrélation), l'impact sur les métriques S2 du bilan global de l'entreprise n'est pas matériel.

Synthesis

Faced with the challenging current economic environment (low interest rates, successive economic crises, health crisis, etc.) and regulatory environment, marked by many reforms (Solvency 2, pension reform, ANI, 100% health reform, etc.), the balance between profitability, risk and competitiveness is a complex equation to solve for insurers. In this context, knowledge of the insured population is essential to optimize pricing, and to accurately estimate the insurer's liabilities, particularly the level of regulatory capital required in the context of solvency 2.

Generally, in group protection, insurers do not have access to granular information when underwriting new business. In particular, for death risk, the pricing is based on average indicators of the group of insureds (average age, average salary, average number of children, etc.) and on the use of regulatory mortality tables adjusted with coefficients.

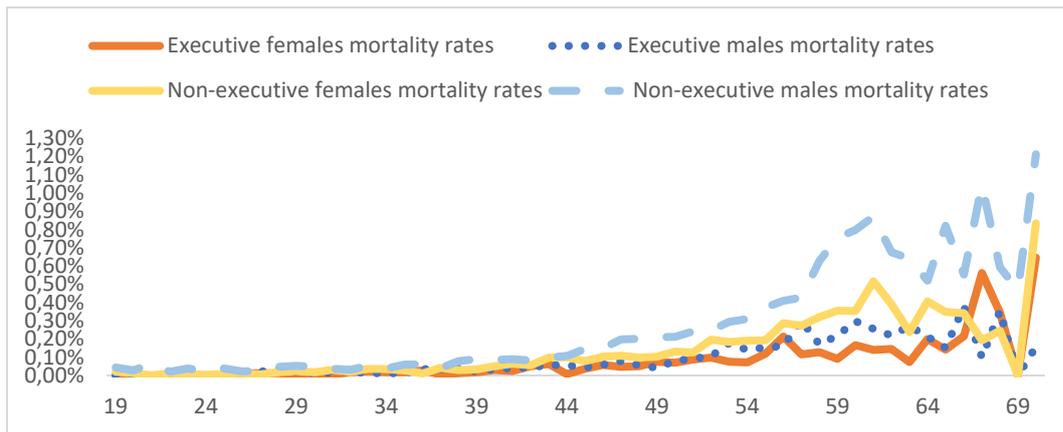
The aim of this thesis is to study the mortality of a group protection portfolio by building experience mortality tables segmented by gender and socio-professional category from DSN data (“déclaration sociale nominative”).

Indeed, the mandatory implementation of DSN since 2017 for employers has enabled the gradual collection of detailed data concerning the employees of the companies in the portfolio, thus facilitating the risk assessment.

The study will be structured around the following axes:

- Estimation of mortality rates by age, gender and socio-professional category
- Comparison of experience mortality tables with current pricing tables
- Optimization of the death risk pricing on group level with the introduction of actuarial age coefficient
- Impact estimation of the use of segmented mortality tables in the calibration of the catastrophe risk shock and the calculation of the associated regulatory economic capital.

The first step is to estimate the death rates by gender and socio-professional category with the Hoem parametric method for ages ranging from 19 to 70 years. These ages are the most representative and usually correspond to the active life period. The results obtained indicate that the mortality rates of non-executive males (“homme non cadre”) are well above those of other groups, especially over 45 years old. In average, they are twice higher than executive males and non-executive females’ rates. The observed mortality rates are very volatile from 60 to 70 due to the low exposures over these ages.



Mortality rates estimated by Hoem

These results imply to study the influence of gender and socio-professional category on portfolio mortality with the Cox relational model. The non-executive males' mortality is the reference mortality because it is more representative in terms of data on both exposure and number of deaths.

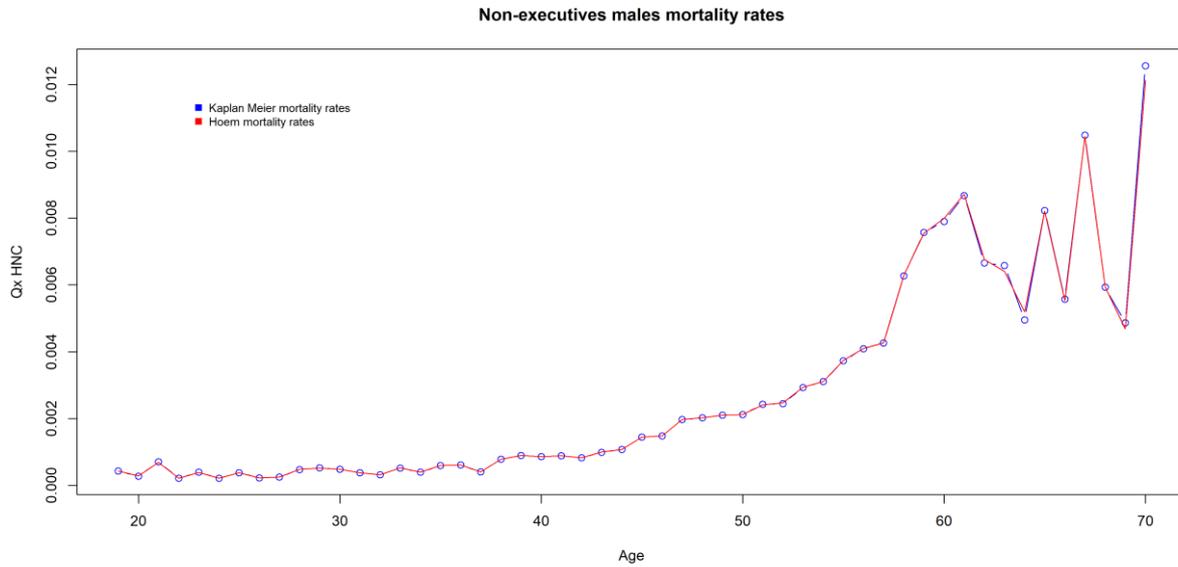
The following coefficients are obtained with the Cox model:

Variables	θ_0 estimated	$\exp(\theta_0)$ estimated	Standard deviation	Critical probabilities p-value
Female	-0.53	0.59	0.030	<2e-16
Executive	-0.87	0.42	0.037	<2e-16

Their interpretation indicates that the survival function of non-executive males decreases more rapidly than other groups. It also shows that females have dying conditional probabilities lower than males by ~41% (1-0.59) and executives have conditional probabilities of death lower than non-executives by ~58%(1-0.42).

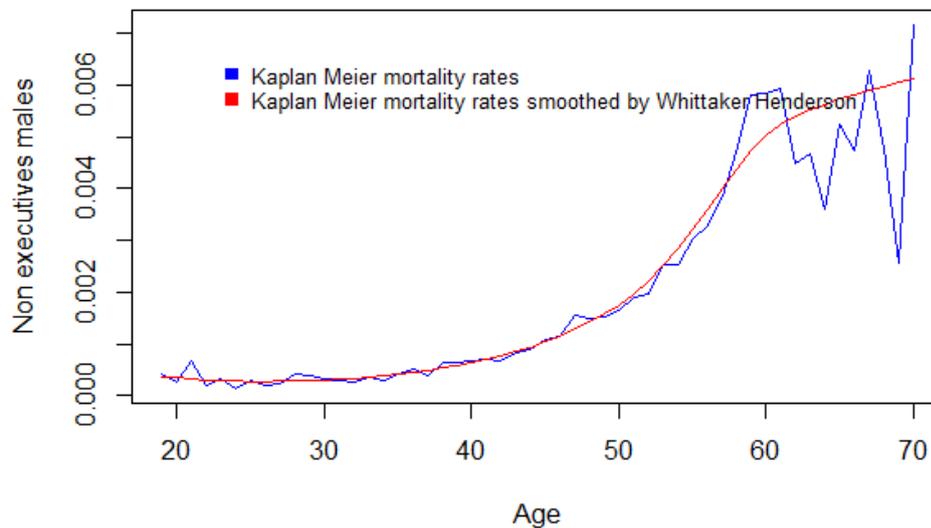
The various significance tests of the coefficients (likelihood ratio, score, Wald) conclude that the vector of coefficient θ_0 estimated by the Cox model is significant and non-equal to zero. Furthermore, the underlying hypotheses of the Cox model, in particular, proportional hazard is verified with the test of Schoenfeld residuals.

The second step is to choose a model for non-executive males' mortality and deduce other groups 'mortality rates (executive males, executive females, non-executive females) with Cox model coefficients. So, we estimated the non-executive males' mortality rates with the nonparametric method of Kaplan Meier and compared them with those obtained with the Hoem method. The two curves are similar as shown in the diagram below:

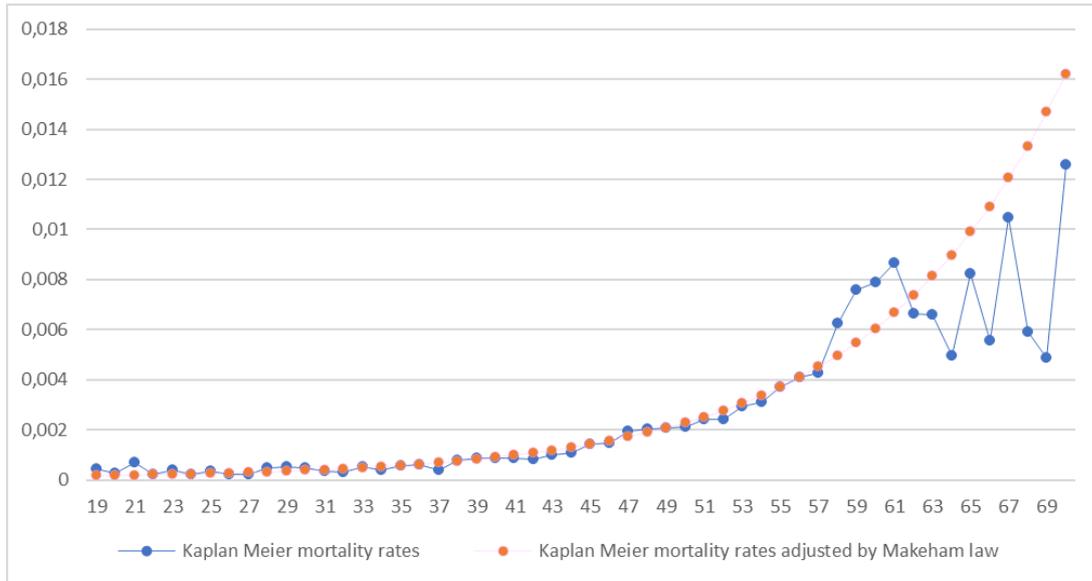


Kaplan Meier vs Hoem for non-executive males

However, we will use the Kaplan Meier model because it is nonparametric and less cumbersome in terms of assumptions. The mortality rates obtained by Kaplan Meier are then smoothed to make them more regular. The smoothing methods used are Whittaker Henderson and the Makeham law fitting.



Smoothing by Whittaker-Henderson ($h=1, z=2$)



Makeham Law fitting

The various statistical tests applied (sign test, SMR test, confidence interval) validate the estimation of the mortality rates by Kaplan Meier nonparametric method whether they are smoothed by Whittaker-Henderson or adjusted to the Makeham's law. However, we will choose the totally nonparametric model (Kaplan Meier smoothed by Whittaker-Henderson). Although the Makeham law fits well our data, we note an underestimation of the mortality on young ages and an overestimation over 50 years old due to the model structure containing an exponential parameter depending on age.

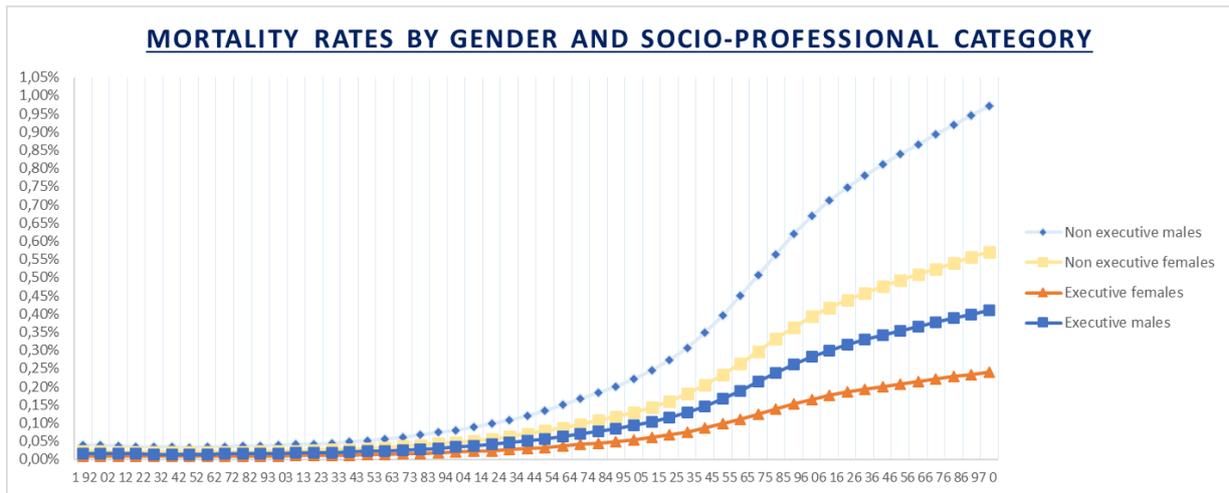
We can deduce the mortality rates of other groups with the following formula:

For each group i , $i = 2, 3, 4$ and each age x :

$$q_x^i = 1 - \frac{S_i(x+1)}{S_i(x)} = 1 - \left(\frac{S_1(x+1)}{S_1(x)} \right)^{\exp(\theta_0^T Z^i)} = 1 - (1 - q_x^1)^{\exp(\theta_0^T Z^i)} .$$

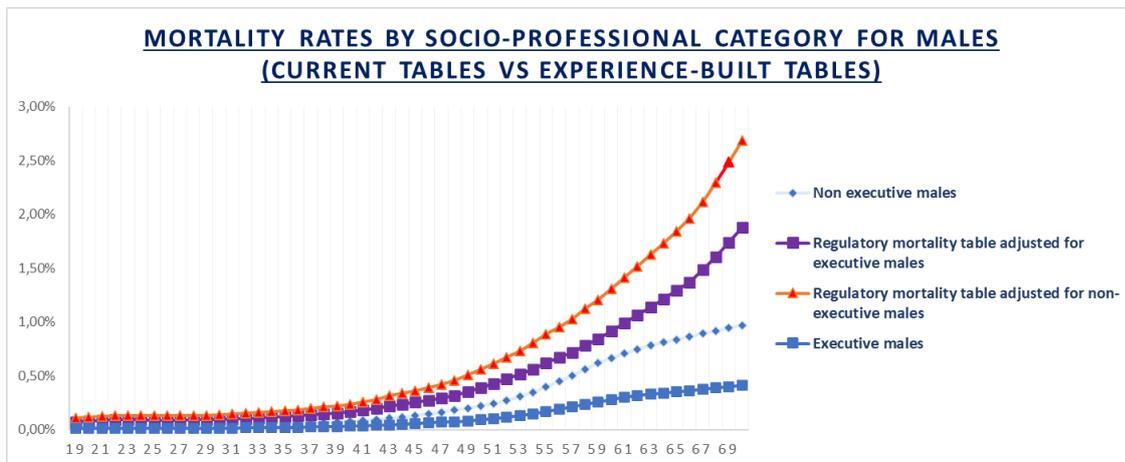
q_x^1 the mortality rate of non-executive males (reference group) and $Z_1^i = 1$ if gender= female and 0 otherwise and $Z_2^i = 1$ if socio-professional category =executive and 0 otherwise.

We obtain the following results:

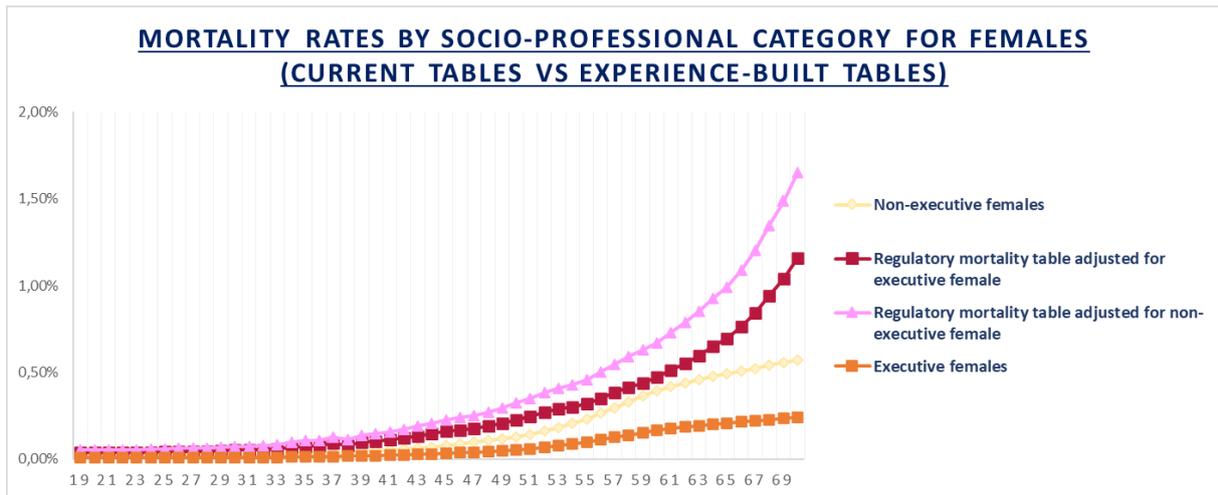


Mortality rates by gender and social-professional category

In the third step, we compare the current pricing tables and the experience-built tables. This comparison shows that the mortality rates estimated by these latter are widely lower than those applied for the pricing. In average, they are lower by 60% for non-executives and by 70% for executives.



Current tables vs experience-built tables for males



Current tables vs experience-built tables for females

Moreover, we notice that the regulatory mortality table adjusted for executives is higher than for experience-built table rates for non-executives. This implies a significant overestimation of the current pricing tables.

However, commercial adjustments are applied on final death premium to account for competitive environment.

Using the experience-built tables should generate an operational gain with the removal of those commercial adjustments (process simplification) and allow improvement of our competitive positioning with a more refined pricing.

In addition, we propose to optimize the pricing at group level.

For an insured of age x , the pricing of death guarantee is as follows:

$$premium = q_x * sum\ insured .$$

The sum insured corresponds to the insurer's liabilities in case of death and depends on several parameters including salary, family status, number of children, socio-professional etc.

For group protection contracts, the pricing is established on the average head (average age, average salary, average number of children etc.).

Nonetheless, it is essential to consider the repartition of age and associated sum insured to avoid bias in pricing. Therefore, we introduce actuarial age coefficients.

The actuarial age is the average age x for which the average group premium is the closest to the best estimate.

Example of pricing based on actuarial age

	Group 1			Group 2			Group 3		
	Ind1	Ind2	Ind3	Ind1	Ind2	Ind3	Ind1	Ind2	Ind3
Age	25	25	25	20	20	35	19	26	31
q_x	0,031%	0,031%	0,031%	0,026%	0,026%	0,057%	0,025%	0,033%	0,043%
Sum insured (K€)	63	63	63	32	32	144	28	72	114
Individual premium (€)	19,73	19,73	19,73	8,21	8,21	81,67	7,19	23,31	48,60
Average age of the group	25			25			25		
Average premium of the group without accounting for individuals' specificities (€)	19,73			19,73			19,73		
Average premium of the group accounting for individuals' specificities (€)	19,73			32,70			26,37		
Age for which the average group premium, accounting for individuals' specificities, is the closest (actuarial age)	25			29			28		

In the above table, we notice that premiums are different when considering individual specificities (q_x , sum insured) for each group. The actuarial ages for group 2 and 3 are respectively 29 and 28.

To determine the actuarial age coefficients to be applied in the pricing, we use portfolio's statistics. The principle is, first, build a table with the workforce age distribution for each average age x . It is then, completed with the average sum insured by age, gender, and socio-professional status to obtain an average group profile for x .

Afterward, for each average age x , the premium is based on the weighted average of individual premiums considering specificities (sum insured by gender and socio-professional status).

Those adjustments lead to increase the average death premium by +46% for non-executive males, +32% for non-executive females, +62% for executive males and +46% for executive females.

Finally, the actuarial age is obtained by searching the closest age y corresponding to the group premium ($q_y * sum\ insured_y \approx weighted\ average\ of\ individual\ premiums$).

Below are the results:

Age	Non-executive actuarial age coefficients	
	Male	Female
23 years old and younger	0	0
24 to 27	3	2
28 to 32	5	5
33 to 40	6	5
41 to 43	6	4
44 to 46	5	3
47 to 50	4	3
51 to 54	2	1
55 +	0	0

Age	Executive actuarial age coefficients	
	Male	Female
23 years and younger	0	0
24 to 27	3	2
28 to 32	6	5
33 to 40	7	6
41 to 43	6	6
44 to 46	5	5
47 to 50	4	3
51 to 54	2	1
55 +	0	0

In case of new business, a company with an average age x will have a mortality rate of $q_{x+actuarial\ age\ coefficient}$.

The final step consists of using the experience-built tables in the calibration of CAT mortality risk S2 shock. This shock corresponds to an additional increase of the mortality the first year of projection. An internal study estimates this shock at 0,08% (additional mortality).

However, in group protection perimeter, the solvency 2 internal model operates with Loss ratio and the additional increase of mortality rates must be converted in loss ratio impact.

Thus, the retained formula to convert this shock is:

$$c\% = \left(\frac{q_x + choc}{q_x} \right) - 1,$$

q_x is the best estimate average mortality rate calibrated on the portfolio

Increase the mortality rates of 0,08% leads to an impact of + $c\%$ on the loss ratio.

The formula of shocked loss ratio is $Loss\ ratio_{basis} \times (1 + c\%)$.

Example: For $q_x = 0,20\%$, we have $c\% = \left(\frac{0,20\% + 0,08\%}{0,20\%} \right) - 1 = 40\%$.

The shock applied to the basis loss ratio is 40%.

Today, the calculation of the best estimate average mortality rate is established on experience mortality table based only on gender and is estimated at 0,15%.

$$Best\ Estimate\ average\ q_x = \sum_{x=age_{min}}^{x=age_{max}} \frac{CSR_x^M \times q_x^M}{CSR_{Total}^M + CSR_{Total}^F} + \frac{CSR_x^F \times q_x^F}{CSR_{Total}^M + CSR_{Total}^F}.$$

avec $CSR = sum\ insured$.

The formula choice is explained by the relation between the average best estimate q_x and the total loss S :

$$S = Best\ Estimate\ average\ q_x \times total\ sum\ insured.$$

To refine the calculation of CAT mortality shock, we recalculate the best estimate average q_x by using the experience-built tables i.e., considering the social-professional category. The formula for calculating the q_x is rewritten as follows:

$$\text{Best Estimate average } q_x = \frac{1}{CSR_{Total}} \sum_{x=age_{min}}^{x=age_{max}} CSR_x^{EM} \times q_x^{EM} + CSR_x^{EF} \times q_x^{EF} + CSR_x^{NEM} \times q_x^{NEM} + CSR_x^{NEF} \times q_x^{NEF} .$$

With $CSR_{Total} = CSR_{Total}^{EM} + CSR_{Total}^{EF} + CSR_{Total}^{NEM} + CSR_{Total}^{NEF} = \text{total sum insured} .$

The new best estimate average q_x is estimated at 0,13% and is lower than the previous best estimate mortality rate (0,15%). This leads to a mechanical increase of the loss ratio shock by 16%. Based on equivalent loss ratio, we obtain the following STEC² amounts:

STEC amounts estimated for CAT mortality risk

	Current table	Table considering social-professional category
Shock	0,08%	0,08%
Best Estimate average Qx	0,15%	0,13%
Shock converted in loss ratio impact	54%	62%
Variation		16%
STEC amounts estimated (M€)	238	276
Variation (M€)		38

The variation of the STEC is sensitive to the variation of the Best Estimate average q_x . A decrease of 0,02 point of this parameter leads to an increase of 16% of the shock and the STEC amount. Nevertheless, given the reduced weight of CAT mortality STEC on total life STEC and the diversification effects between risks (correlation matrix), the impact on S2 metrics of the company's overall balance sheet will be not significant.

² STEC= short term economic capital → it is the equivalent of solvency capital requirement (SCR) on AXA internal model

Remerciements

Tout d'abord, je tiens à remercier Gilles CUVILLIER, Responsable des risques internationaux et Prévoyance Santé d'AXA France pour son tutorat et son accompagnement tout au long de la rédaction de ce mémoire. Je le remercie d'avoir pris le temps de relire tout le mémoire et pour ses conseils avisés.

Ensuite, je tiens à remercier la Direction Data Santé et collectives et la Direction Technique Marketing et Innovation France d'AXA de m'avoir facilité l'accès aux données DSN et pour leur collaboration.

Enfin, j'adresse toute ma gratitude à ma famille pour son soutien infaillible pendant toute la durée de la formation du CEA et la rédaction de ce mémoire.

Sommaire

RESUME	3
SYNTHESE	5
REMERCIEMENTS	23
INTRODUCTION	27
1 LA PREVOYANCE COLLECTIVE EN FRANCE	29
1.1 Présentation	29
1.2 Dynamique de marché	30
1.3 Offre de produit	30
1.4 Risques principalement couverts en prévoyance collective	31
1.4.1 L'incapacité temporaire de travail (ITT)	31
1.4.2 L'invalidité permanente	32
1.4.3 Le décès	32
1.5 Mise en place d'un contrat de prévoyance collective	36
1.5.1 La convention ou l'accord collectif d'entreprise	36
1.5.2 Le référendum	37
1.5.3 La décision unilatérale de l'employeur	37
1.6 Les obligations légales et conventionnelles en Prévoyance	37
1.6.1 La convention collective nationale des cadres	37
1.6.2 L'accord national de mensualisation	37
1.6.3 Les conventions collectives professionnelles	38
1.6.4 La loi Evin du 31 Décembre 1989	39
1.6.5 La loi du 8 Aout 1994	39
1.6.6 Les avantages des contrats collectifs de prévoyance	39
2 ANALYSE DE LA MORTALITE DU PORTEFEUILLE	41
2.1 Construction de la base de données	41
2.1.1 Qu'est-ce que la DSN ?	41
2.1.2 Quelles données de DSN sélectionnées ?	42
2.1.3 Traitement des données de DSN	43
2.1.4 Fiabilisation des décès dans les données de DSN	49
2.2 Analyse descriptive des données	49
2.3 Construction des tables de mortalité d'expérience	52

2.3.1	Introduction à l'analyse de survie.....	52
2.3.2	Modélisation des taux de mortalité	60
3	IMPACT DE L'UTILISATION DES TABLES DE MORTALITE D'EXPERIENCE	87
3.1	Impact dans la tarification	87
3.1.1	Comparaison des tables d'expérience et des tables de tarification	87
3.1.2	Optimisation du tarif décès à la maille groupe.....	88
3.2	Impact dans le calibrage du choc catastrophe en vie	94
3.2.1	Rappel de la directive Solvabilité 2.....	94
3.2.2	Le pilier 1 de Solvabilité 2	95
3.2.3	Le risque Catastrophe en vie (CAT).....	98
	CONCLUSION.....	101
	REFERENCES.....	103
	ANNEXES	104

Introduction

Dans le cadre de leurs activités, les assureurs doivent trouver un équilibre entre compétitivité et rentabilité. Le contexte économique et réglementaire actuel, marqué par différentes réformes (solvabilité 2, réforme des retraites, 100% santé, etc.), les amènent à mieux maîtriser leurs risques avec une optimisation de leur tarification et une évaluation, au plus juste, de leurs engagements.

Ainsi, une bonne connaissance de leurs portefeuilles d'assurés reste un enjeu majeur, pour pouvoir proposer des tarifs attractifs aux clients, afin de garder ceux les plus rentables déjà présents en portefeuille, mais également en attirer de nouveaux.

Cela est d'autant plus important pour les contrats collectifs d'entreprise où en général, l'assureur dispose d'informations limitées et peu granulaires au moment de la souscription.

L'objectif de ce mémoire est d'étudier la garantie décès d'un portefeuille de prévoyance collective avec la construction de tables de mortalité d'expérience segmentées par sexe et catégorie socio-professionnelle (CSP) à partir des données de DSN (déclarations sociales nominatives).

Les données de DSN sont des fichiers construits à partir des bulletins de paie des salariés et envoyés mensuellement aux organismes administratifs (CPAM³, URSSAF⁴, centre des impôts...) pour la gestion de la protection sociale des salariés.

Depuis leur mise en place obligatoire en 2017, les organismes d'assurance peuvent disposer de ces données concernant les salariés des entreprises assurées sous réserve de leur accord. Ces données englobent un panel d'informations utiles pour une meilleure connaissance des risques.

Après une présentation de la prévoyance collective et des aspects réglementaires encadrant sa mise en place, la première étape de l'étude sera de traiter et d'analyser les données de DSN avec leurs spécificités. Puis, débutera l'analyse de la mortalité du portefeuille avec l'estimation des taux de mortalité bruts par âge, sexe, et catégorie socio-professionnelle avec la méthode paramétrique de Hoem. L'influence du sexe et de la catégorie socio-professionnelle sur la mortalité sera évaluée par le modèle de Cox avec la définition d'un groupe dont la mortalité sera considérée comme celle de référence.

La deuxième étape sera de choisir un modèle adéquat pour définir la mortalité de ce groupe de référence. Une comparaison des taux de mortalité estimés par la méthode de Hoem et celle non paramétrique de Kaplan Meier sera effectuée et les taux bruts retenus sont lissés par des méthodes paramétriques et non paramétriques (Whittaker-Henderson, Makeham).

Les taux de mortalité des autres groupes seront déduits de ceux du groupe de référence si les hypothèses et coefficients du modèle de Cox sont statistiquement validés.

³ CPAM= caisse primaire d'assurance maladie

⁴ URSSAF= Union de recouvrement des cotisations de Sécurité sociale et d'allocations familiales

La troisième étape consistera, à challenger la tarification actuelle avec une comparaison des tables de mortalité d'expérience construites avec celles utilisées actuellement dans la tarification correspondant à des tables réglementaires ajustées avec des coefficients d'abattement.

Une optimisation de la tarification du décès à la maille groupe sera proposée avec l'introduction de correctifs « âge actuariel ». En effet, le tarif du décès est calibré sur la tête moyenne du groupe d'assurés (âge moyen arithmétique, salaire moyen, nombre d'enfants moyens etc) et il est important de prendre en compte les disparités dans le groupe afin d'éviter de biaiser le vrai coût des engagements.

En dernière étape, nous évaluerons l'impact de l'utilisation de ces tables d'expérience segmentées par CSP dans le calcul du capital économique réglementaire requis pour le risque de catastrophe en vie dans le cadre de solvabilité 2.

1 La Prévoyance collective en France

1.1 Présentation

Les contrats de prévoyance collective sont des contrats de groupe à adhésion obligatoire ou facultative à destination d'entreprises ou d'associations pour la couverture respective de leurs salariés et adhérents. L'objectif étant de compenser la perte de revenus de la famille des assurés en cas de maladie, d'incapacité temporaire de travail, d'invalidité et de décès de l'assuré. En effet, les risques précités sont soit couverts de façon partielle par la sécurité sociale, soit non couverts par celle-ci. Ce qui peut avoir de lourdes conséquences financières sur le niveau de vie des assurés et leurs familles. Le graphique ci-dessous illustre de façon simplifiée le schéma de l'assurance collective entre les différentes parties concernées.

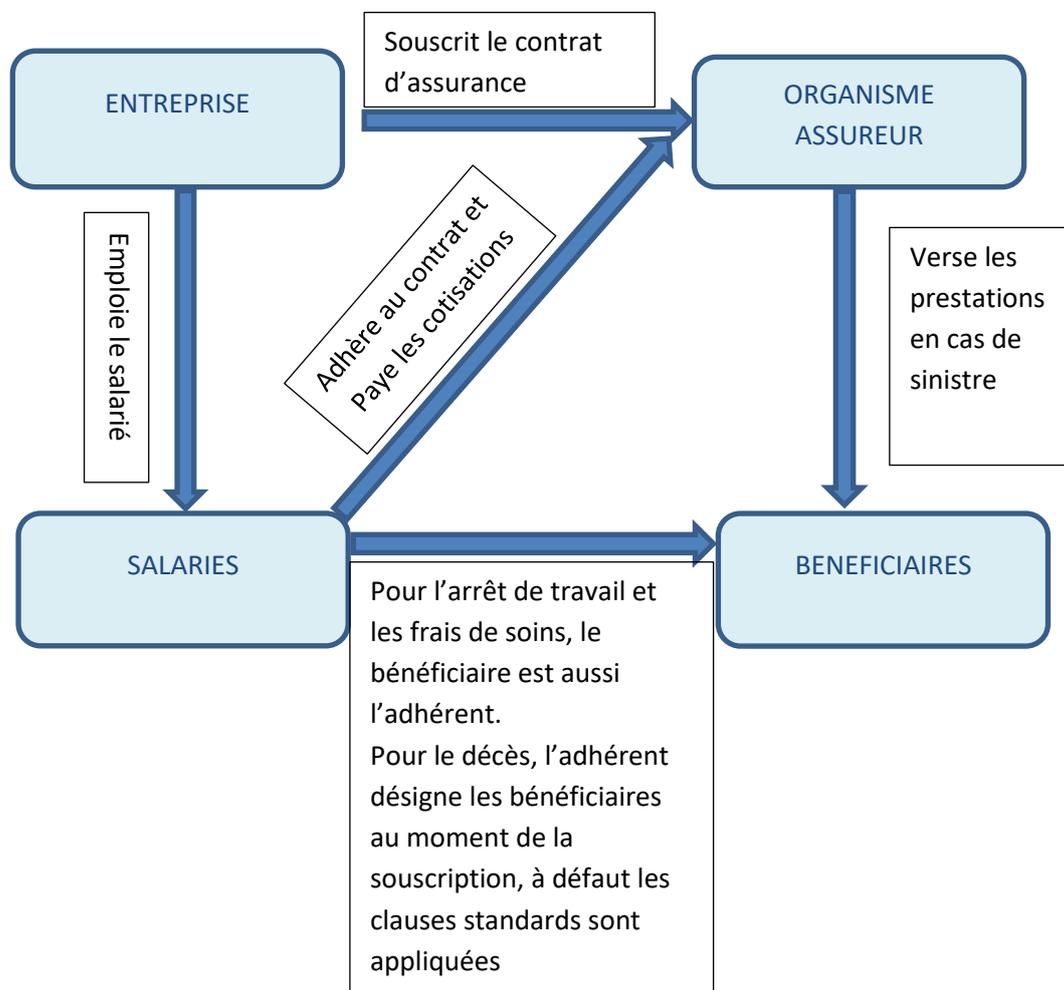


Figure 1: Schéma simplifié de l'assurance collective

1.2 Dynamique de marché

En France, un nombre important d'individus ne bénéficie pas d'une couverture en prévoyance notamment chez les non-cadres. En effet, contrairement à la santé avec l'accord national interprofessionnel (ANI), il n'y a pas d'obligation de couverture en prévoyance collective. Il s'agit donc d'un marché à fort potentiel de développement.

A fin 2018, les cotisations en prévoyance collective représentent 12,2 Md€. Les parts de marché sont détenues principalement par les institutions de prévoyance avec près de la moitié des cotisations (49%) et les sociétés d'assurance avec 45% des cotisations.

La présence importante des institutions de prévoyance s'explique par les clauses de désignation qui existaient auparavant dans les accords de branche. Ils imposaient, pour la plupart, aux entreprises concernées par la convention collective, des institutions de prévoyance comme assureurs en prévoyance et santé. Depuis 2013, les clauses de désignation ont été supprimées des accords de branches par le conseil constitutionnel.

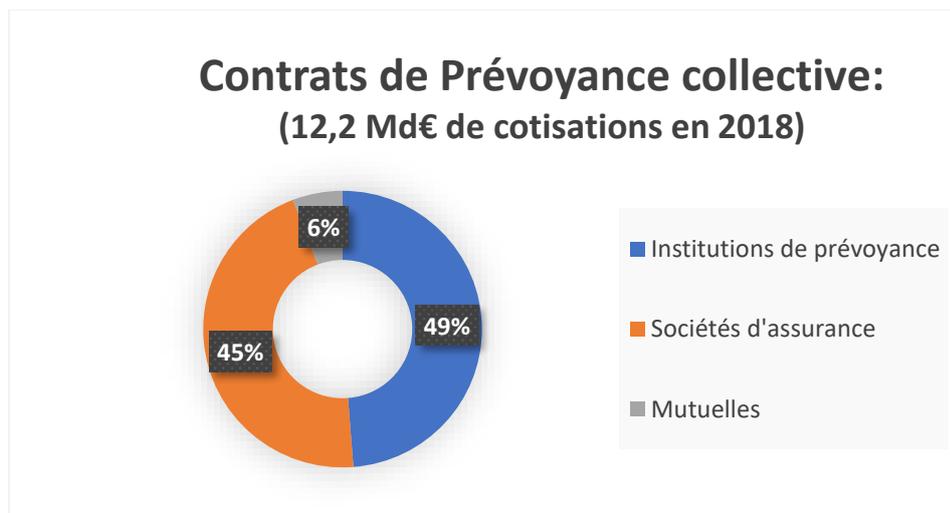


Figure 2: Répartition des cotisations de prévoyance collective 2018 par acteur

1.3 Offre de produit

Généralement, les offres de produits proposées aux entreprises sont en fonction de leur taille, de leur secteur d'activité et de la catégorie socio-professionnelle des salariés.

Tableau 1: Offres de Produit

Cible	Gamme	Commentaires
TPE/PME	Produits standard/quasi-standard	Plusieurs formules modulables selon le souhait de l'entreprise et cohérentes avec les spécificités métiers et/ou CCN
Grandes entreprises	Offre sur mesure	Tarifcation sur mesure selon le cahier des charges de l'entreprise et en adéquation avec les spécificités métiers et/ou CCN

1.4 Risques principalement couverts en prévoyance collective

Comme énoncé dans le paragraphe précédent, les contrats de prévoyance collective couvrent principalement des garanties en cas d'arrêt de travail et de décès.

Généralement, les cotisations et les garanties sont exprimées en fonction du salaire (Tranche A/Tranche B/Tranche C) ou du PASS (Plafond annuel de la sécurité sociale) :

- La tranche A représente la partie du salaire limitée au plafond de la sécurité sociale (PASS)
- La tranche B représente la partie du salaire entre 1 et 4 fois le PASS
- La tranche C représente la partie du salaire entre 4 et 8 fois le PASS

Pour les garanties en cas d'arrêt de travail, les prestations servies complètent celles versées par la sécurité sociale qui sont plafonnées.

1.4.1 L'incapacité temporaire de travail (ITT)

La garantie incapacité temporaire de travail permet à l'adhérent de compenser sa perte de revenus en cas d'arrêt de travail dû à une atteinte de son intégrité physique : maladie professionnelle, maladie non professionnelle, accident de la vie ou du travail.

Elle est temporaire et peut durer jusqu'à 3 ans en cas d'affection longue durée. De plus, elle cesse nécessairement soit à la guérison, soit à la détermination d'un taux d'invalidité ou au décès de l'assuré.

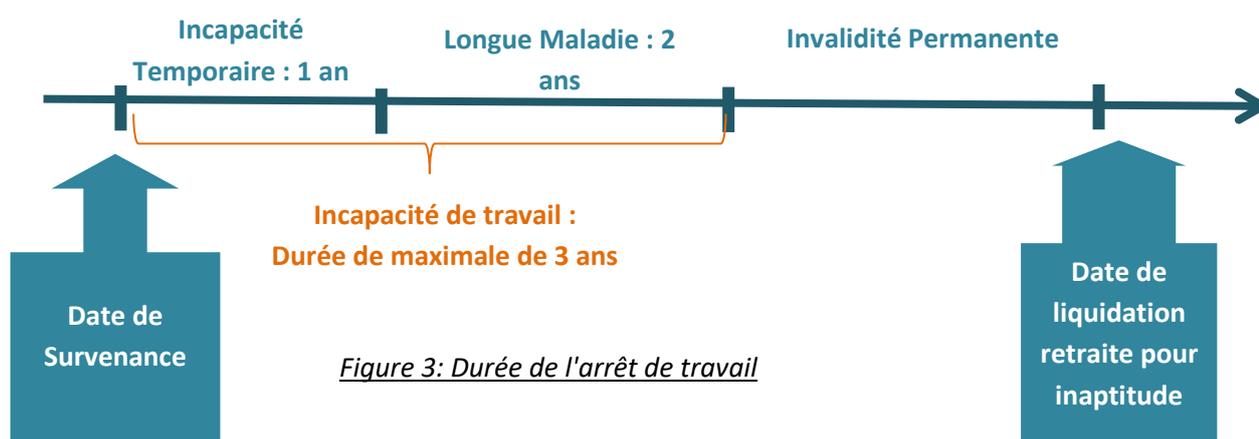


Figure 3: Durée de l'arrêt de travail

1.4.2 L'invalidité permanente

La garantie invalidité permanente permet à l'assuré de compenser la diminution durable de sa capacité de travail et de gain à la suite d'une maladie ou d'un accident. C'est la sécurité sociale qui apprécie l'état d'invalidité de l'assuré, soit lors de la consolidation ou de la stabilisation de son état, soit à l'expiration de la période d'incapacité et de longue maladie.

Elle définit 3 niveaux d'invalidité :

- ❖ L'invalidité de 1^{ère} catégorie : la personne invalide est capable d'exercer une activité rémunérée
- ❖ L'invalidité de 2^{ème} catégorie : la personne invalide est en incapacité d'exercer une profession quelconque
- ❖ L'invalidité de 3^{ème} catégorie : la personne invalide est en incapacité d'exercer une profession quelconque et doit avoir recours à l'assistance d'une tierce personne pour effectuer les actes ordinaires de la vie.

Le montant de la pension d'invalidité est fonction du classement dans une de ces trois catégories et est versé sous forme de rente d'invalidité servie jusqu'à la date de liquidation des droits de la retraite. La pension d'invalidité se transforme alors en pension retraite pour inaptitude.

1.4.3 Le décès

Il existe plusieurs options de garanties proposées en cas de décès de l'assuré. Elles peuvent être souscrites cumulées ou en substitution (cf exemple tableau de garanties).

1.4.3.1 Les différentes options de garanties

- Le décès en capital

Il s'agit de la garantie décès la plus courante dans les contrats de prévoyance. Le versement du capital est effectué aux bénéficiaires désignés, à la suite du décès de l'assuré. Il peut s'agir du conjoint, des parents, des enfants ou d'autres héritiers. Ce capital peut être versé par anticipation en cas de perte totale et irréversible d'autonomie de l'assuré. Dans certains cas, le capital est doublé en cas de décès accidentel ou de décès simultané du conjoint de l'assuré. Il peut également être majoré par enfant à charge en cas de mariage ou de lien par pacte civil de solidarité.

- La rente éducation

La garantie rente éducation est versée aux enfants de l'assuré. Elle est en général temporaire jusqu'à 21 ans ou 26 ans sous condition de la poursuite des études et viagère pour les enfants handicapés. Elle peut être doublée pour les orphelins de père et de mère.

- La rente de conjoint

Elle est versée de façon temporaire ou viagère selon le contrat, au conjoint, au concubin ou partenaire lié par un pacte civil de solidarité de l'assuré. Elle a pour but de compenser les droits de réversion du conjoint non acquis entre la date du décès de l'assuré et son 65^{ème} anniversaire.

- Les frais d'obsèques

La garantie est versée en général sous forme de forfait, en fonction du plafond mensuel de la sécurité sociale pour couvrir les dépenses liées aux frais d'obsèques à la suite du décès de l'assuré.

Le tableau ci-après montre un exemple de garanties qui pourraient être proposées pour la couverture en cas de décès :

Tableau 2: Exemple de tableau de garanties en décès

Garanties	
DECES OU PERTE TOTALE ET IRREVERSIBLE D'AUTONOMIE (PTIA) TOUTES CAUSES + RENTE EDUCATION (en % des tranches A, B et C du salaire annuel brut)	
En cas de décès de l'assuré, le bénéficiaire a le choix entre l'OPTION 1 et l'OPTION 2	
OPTION 1 (Capital + Rente éducation)	
<u>CAPITAL :</u>	
<ul style="list-style-type: none"> • Assuré sans personne à charge : <ul style="list-style-type: none"> - Célibataire, veuf, divorcé - Marié, pacsé • Assuré avec personne(s) à charge : <ul style="list-style-type: none"> - Célibataire, divorcé - Veuf - Marié, pacsé - Majoration par enfant à charge 	<p>200%</p> <p>100%</p> <p>100%</p> <p>100%</p> <p>100%</p> <p>20%</p>
<u>RENTE EDUCATION :</u>	
<ul style="list-style-type: none"> - Jusqu'à 5 ans inclus - De 6 ans à 13 ans inclus - De 14 ans à 25 ans inclus (si poursuite d'études) 	<p>5%</p> <p>10%</p> <p>15%</p>
(La rente éducation est versée viagèrement pour les enfants bénéficiaires des allocations pour personnes handicapées)	
OPTION 2 (Capital minoré + Rente éducation majorée)	
<u>CAPITAL :</u>	
<ul style="list-style-type: none"> - Célibataire, divorcé - Veuf - Marié, pacsé - Majoration par enfant à charge 	<p>50%</p> <p>50%</p> <p>50%</p> <p>20%</p>

RENTE EDUCATION :	
- Jusqu'à 5 ans inclus	10%
- De 6 ans à 13 ans inclus	15%
- De 14 ans à 25 ans inclus (si poursuite d'études)	20%
(La rente éducation est versée viagèrement pour les enfants bénéficiaires des allocations pour personnes handicapées)	
En cas de PTIA, l'option 1 s'applique en tout état de cause	
DECES SIMULTANE OU POSTERIEUR DU CONJOINT OU PARTENAIRE PACSE (en % des tranches A, B et C du salaire annuel brut)	
En cas de décès simultané ou postérieur du conjoint ou partenaire pacsé	100%
Majoration par enfant à charge	20%
FRAIS D'OBSEQUES	
En cas de décès de l'assuré	100% du PMSS
PREDECES	
En cas de prédécès du conjoint, du partenaire pacsé ou d'un enfant à charge	100% du PMSS
RENTE DE CONJOINT (en % des tranches A et B du salaire annuel brut)	
En cas de décès de l'assuré, versement d'une rente au profit du conjoint ou partenaire pacsé :	
• Rente viagère (x étant l'âge de l'assuré à la date du décès)	$0,80\% \times (65 - x)$ avec un minimum de 5 ans
• Rente temporaire (x étant l'âge de l'assuré à la date du décès)	$0,40\% \times (x - 25)$

1.4.3.2 Tarification du décès : rappel de la réglementation française

L'article A. 132-18 du code des assurances encadre les aspects techniques de la tarification et du provisionnement pour les contrats d'assurance sur la vie. Il stipule que les assureurs doivent utiliser les tables réglementaires en vigueur ou des tables d'expériences établies en fonction des spécificités de leur portefeuille, pour établir leur tarif en cas de décès.

Lorsque des tables d'expérience sont utilisées, celles-ci doivent être certifiées par un actuair externe et indépendant de l'entreprise.

Pour le cas particulier des contrats collectifs résiliables annuellement, le tarif peut être établi avec les tables réglementaires ajustées avec une méthode forfaitaire si celle-ci est justifiée (en fonction par exemple des CSP, du sexe...).

Art. A. 132-18 du code des assurances :

Les tarifs pratiqués par les entreprises d'assurance sur la vie et de capitalisation et par les fonds de retraite professionnelle supplémentaire comprennent la rémunération de l'entreprise et sont établis d'après les éléments suivants :

1° Un taux d'intérêt technique fixé dans les conditions prévues à l'article A. 132-1.

2° Une des tables suivantes :

a) Tables homologuées par arrêté du ministre de l'Économie et des finances, établies par sexe, sur la base de populations d'assurés pour les contrats de rente viagère, et sur la base de données publiées par l'Institut national de la statistique et des études économiques pour les autres contrats ;

b) Tables établies ou non par sexe par l'entreprise d'assurance et certifiées par un actuaire indépendant de cette entreprise, agréé à cet effet par l'une des associations d'actuaire reconnues par l'autorité mentionnée à l'article L. 310-12.

Les tables mentionnées au b sont établies d'après des données d'expérience de l'entreprise d'assurance, ou des données d'expérience démographiquement équivalentes.

Lorsque les tarifs sont établis d'après des tables mentionnées au a, et dès lors qu'est retenue une table unique pour tous les assurés, celle-ci correspond à la table appropriée conduisant au tarif le plus prudent.

Pour les contrats en cas de vie autres que les contrats de rente viagère, les tables mentionnées au a sont utilisées en corrigeant l'âge de l'assuré conformément aux décalages d'âge ci-annexés.

Pour les contrats de rentes viagères, en ce compris celles revêtant un caractère temporaire, et à l'exception des contrats relevant du chapitre III du titre IV du livre 1er, le tarif déterminé en utilisant les tables mentionnées au b ne peut être inférieur à celui qui résulterait de l'utilisation des tables appropriées mentionnées au a.

Pour les contrats collectifs en cas de décès résiliables annuellement, le tarif peut être établi d'après les tables mentionnées au a avec une méthode forfaitaire si celle-ci est justifiable.

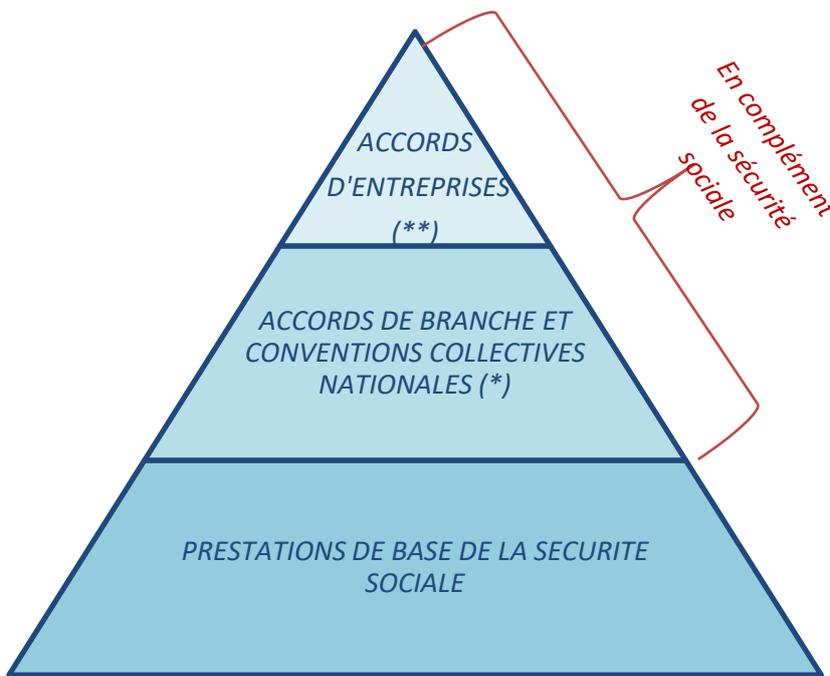
1.5 Mise en place d'un contrat de prévoyance collective

La mise en place de régime de prévoyance collective au niveau de l'entreprise peut se faire selon différentes procédures :

1.5.1 La convention ou l'accord collectif d'entreprise

Il est négocié entre l'employeur et les organisations syndicales représentatives des salariés au niveau de l'entreprise. Il permet de rendre obligatoire, pour tous les salariés concernés, l'affiliation à la couverture mise en place qu'il s'agisse des contrats de travail en cours ou ceux à venir. Les salariés ne peuvent en aucun cas renoncer aux effets des clauses de celui-ci.

Lorsqu'une entreprise est soumise à une convention collective (cf paragraphe 1.6.3), la mise en place d'une convention ou d'un accord collectif signé au niveau de l'entreprise ne peut qu'adapter, les dispositions de la convention collective, dans un sens plus favorable aux salariés comme illustré dans le schéma ci-dessous.



(*) Obligations pour l'employeur de proposer aux salariés une couverture minimale égale à celle de la Convention collective nationale.

() La couverture la plus favorable pour l'assuré entre l'accord de branche et l'accord d'entreprise est appliquée**

Figure 4: Fonctionnement de la couverture en prévoyance

1.5.2 Le référendum

En cas de référendum, l'accord est soumis à la ratification de la majorité des salariés concernés. L'accord approuvé par référendum porte les mêmes effets que celui collectif. Il est notamment intéressant d'avoir recours au référendum en cas d'impossibilité légale de conclure un accord collectif (absence de délégués syndicaux) ou encore en cas de mésentente entre l'employeur et les organisations syndicales.

1.5.3 La décision unilatérale de l'employeur

La décision unilatérale de l'employeur n'a pas la force obligatoire de l'accord collectif ou du référendum. L'employeur n'a pas la possibilité d'imposer le paiement par les salariés présents au moment de la mise en place d'une partie de la cotisation.

Cependant, pour les salariés embauchés après la mise en place de la couverture, l'affiliation sera obligatoire puisqu'elle fait partie des conditions d'embauche.

1.6 Les obligations légales et conventionnelles en Prévoyance

1.6.1 La convention collective nationale des cadres

L'article 7 de la convention collective nationale des cadres du 14 Mars 1947 prévoit :

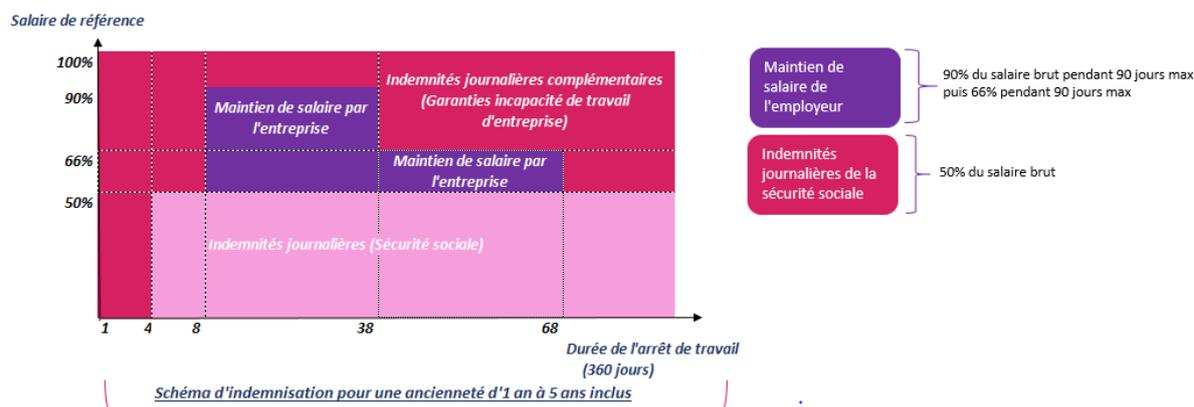
- une obligation de couverture en prévoyance des salariés cadres avec un financement à 100%, à la charge de l'employeur, à hauteur de 1,5% de la tranche A du salaire brut annuel minimum
- l'affectation de la moitié de cette somme (0,75% TA) à des garanties en cas de décès

En cas de non-respect de ces dispositions, l'employeur devra verser une somme égale à 3 fois le plafond annuel de la sécurité sociale aux ayants droit du salarié cadre en cas de décès de celui-ci.

1.6.2 L'accord national de mensualisation

En application de l'accord national du 10 Décembre 1977 étendu par la loi du 19 janvier 1978, les entreprises ont l'obligation de garantir aux salariés ayant un an d'ancienneté, une indemnité minimum en cas d'arrêt de travail, exprimée en pourcentage de la rémunération brute. Le taux et la durée varient en fonction de l'ancienneté dans l'entreprise ou dans la profession. La prise en charge de cette obligation peut être confié à un organisme assureur.

Elle prévoit également le versement d'indemnités en cas de licenciement et de départ à la retraite.



Ancienneté	Indemnisation à 90%	Indemnisation à 66%
1 an à 5 ans inclus	30 jours	30 jours
6 ans à 10 ans inclus	40 jours	40 jours
11 ans à 15 ans inclus	50 jours	50 jours
16 ans à 20 ans inclus	60 jours	60 jours
21 ans à 25 ans inclus	70 jours	70 jours
26 ans à 30 ans inclus	80 jours	80 jours
31 ans et plus	90 jours	90 jours

Figure 5: Schéma de la loi de mensualisation

1.6.3 Les conventions collectives professionnelles

Les conventions collectives professionnelles sont négociées entre organisations représentatives des employeurs et celles des salariés au niveau de la branche professionnelle ou du secteur d'activité. Elles permettent de fixer, au-delà des obligations légales et réglementaires, un statut collectif des salariés au niveau des branches ou secteurs d'activité, et incluent le plus souvent des garanties obligatoires en matière de prévoyance.

Il peut s'agir de garanties en cas de décès, en cas d'incapacité de travail et d'invalidité. Le montant des garanties peut être différent suivant les catégories professionnelles (cadres, agents de maîtrise, employés, ouvriers...). La convention peut imposer une répartition de la cotisation entre l'employeur et le salarié.

Les employeurs qui entrent dans le champ professionnel et territorial de la convention doivent appliquer ces obligations, y compris ceux n'adhérant pas aux organisations syndicales signataires, dès lors que la convention collective a été étendue.

Les accords pouvaient comporter des clauses de désignation sur la couverture en prévoyance et santé : seuls certains assureurs, institutions de prévoyance ou mutuelles pouvaient assurer le régime de protection sociale de la branche d'activité. Ces clauses avaient pour objectif de mutualiser les résultats au sein des branches.

Cependant, avec la mise en place de l'ANI en 2013, ces clauses ont été supprimées par le conseil constitutionnel car jugées contraires à la constitution. En revanche, il peut y avoir une recommandation de certains acteurs. Cette levée de la clause de désignation a permis une ouverture sur le marché de la prévoyance collective et favorisé ainsi la concurrence.

Désormais, l'entreprise est libre de choisir son assureur avec une obligation de respecter les minimums conventionnels.

1.6.4 La loi Evin du 31 Décembre 1989

La loi Evin renforce la protection des assurés et le contrôle des organismes d'assurance. Elle prévoit notamment :

- l'obligation de garantir les états pathologiques antérieurs à l'adhésion au contrat collectif,
- le maintien de la garantie décès pour les assurés en arrêt de travail ou d'invalidité,
- le maintien des garanties en cours, en cas de résiliation ou de non renouvellement du contrat particulièrement les rentes en cours de versement (indemnités journalières, rente d'invalidité, rente de conjoint ou rente éducation),
- le maintien de la couverture des frais médicaux pour les anciens salariés (invalides, chômeurs, retraités) et les ayants droit d'assurés décédés,
- l'obligation de l'assureur en matière d'information notamment sur les comptes de résultat du contrat.

1.6.5 La loi du 8 Aout 1994

Elle complète la loi Evin et prévoit notamment, en cas de changement d'organisme assureur, la prise en charge de la revalorisation des rentes en cours de service et le maintien de la garantie décès pour les assurés en arrêt de travail ou d'invalidité à un niveau au moins égal à celui du contrat résilié.

1.6.6 Les avantages des contrats collectifs de prévoyance

Les contrats collectifs présentent de nombreux avantages à la fois pour l'employeur mais aussi les salariés. Ils permettent aux employeurs de couvrir des garanties plus étendues que celles résultantes des obligations légales et conventionnelles, et sont des éléments importants de la politique de gestion du personnel.

De plus, ils permettent de proposer des tarifs de groupe moins élevés que les tarifs individuels avec une mutualisation du risque et des coûts (notamment gestion et distribution).

Par ailleurs, l'accord national interprofessionnel (ANI) garantit, en cas de chômage, le maintien de la couverture santé et prévoyance complémentaire pendant au moins la durée du dernier contrat de travail et au maximum pendant 12 mois : c'est la **portabilité des droits**.

En termes d'avantages fiscaux, ils permettent de bénéficier de la déduction fiscale des cotisations dans les limites fixées par l'article 83 du code général des impôts (CGI) et de l'exonération des charges sociales sur la contribution de l'employeur dans les limites fixées par l'article L 242.1 du code de la sécurité sociale.

2 Analyse de la mortalité du portefeuille

L'objectif de cette section est de construire notre base de données d'étude, de l'analyser et la fiabiliser pour ensuite l'utiliser dans la construction des tables de mortalité.

2.1 Construction de la base de données

Pour construire notre base de données d'étude, nous allons nous appuyer à la fois sur les données DSN transmises par les entreprises assurées afin d'avoir la liste exhaustive des assurés mais également des bases de données internes, notamment celles de sinistres pour fiabilisation.

Les contrats étudiés sont ceux en assurance directe et bénéficiant de la garantie décès.

Tous les traitements de données sont effectués avec le logiciel SAS.

2.1.1 Qu'est-ce que la DSN ?

La DSN (Déclaration Sociale Nominative) est un fichier, que chaque entreprise doit adresser aux organismes et administrations de protection sociale (*CPAM, Urssaf, AGIRC ARRCO, Organismes complémentaires, Pôle emploi, Centre des impôts, Caisses régimes spéciaux, etc.*) de façon dématérialisée et à une fréquence mensuelle. Il permet la gestion de la protection sociale des salariés et contient l'ensemble des informations requises à cet effet. Il est construit à partir des bulletins de paie des salariés et reflète leurs situations au moment de leur production avec le détail des événements survenus (Ex : maladie, maternité, changement d'un élément du contrat de travail, fin de contrat de travail, décès, etc.).



Figure 6: Fonctionnement de la DSN

L'envoi des DSN aux administrations de protection sociale est obligatoire depuis 2017. Cependant, il ne l'est pas pour les assureurs. L'entreprise assurée n'est donc pas tenue de fournir ses fichiers à ces derniers.

Depuis leur mise en place, AXA France, par sa collaboration avec les entreprises assurées, reçoit les fichiers mensuels de la DSN pour près de trois quarts de ses clients en assurance directe.

Les données sont reçues et traitées par des équipes de datascientist qui s'occupent de leur cryptage pour respecter les règles RGPD⁵ et sont stockées dans un système d'information sous forme de plusieurs tables.

2.1.2 Quelles données de DSN sélectionnées ?

Comme énoncé précédemment, l'avantage des données DSN est qu'elles fournissent des informations sur la situation de l'assuré à l'instant où le bulletin de paie est émis.

Cela permet de limiter les risques de biais sur la qualité des informations puisqu'avec le flux DSN, nous disposons à la fois de la liste exhaustive des assurés exposés au risque de décès, mais aussi du nombre de décès survenus. Le tableau suivant représente l'ensemble des variables provenant des DSN et sélectionnées pour l'étude de la mortalité.

Tableau 3: Variables DSN sélectionnées

NOM DE VARIABLE	DESCRIPTIF
VARIABLES DSN	
CLE	Identifiant unique de l'individu construit à partir de plusieurs variables : CLE= Nom crypté + Prénom crypté + Code Département de Naissance + Date de Naissance + Sexe
SIRET	Numéro de SIRET de l'entreprise
SALARIE	Identifiant du salarié dans une entreprise donnée SALARIE= CLE + SIRET
CD_PTT_ETB	Département d'implantation de l'établissement
DT_NAI	Date de naissance du salarié
CD_SEXE	Sexe du salarié (1= Homme, 2= Femme)
CD_PTT_SAL	Département de résidence du salarié
CSP	Catégorie socio-professionnelle (ouvrier, cadres, agents de maîtrise, ...)
DT_DEB_CTR	Date de début du contrat de travail
DT_FIN_CTR	Date de fin du contrat de travail
MTF_FIN_CTR_TRV	Code permettant d'identifier le motif de la fin du contrat de travail : (66= Décès, 31= fin de CDD, 14=licenciement, 39= départ à la retraite, 59= Démission, etc...)

⁵ Règlement général sur la protection des données

CD_CONV_IDCC IDCC_LIBELLE	<i>Code de la convention collective sur laquelle dépend le salarié Et son libellé associé</i>
VARIABLES EN LIEN AVEC LE CONTRAT D'ASSURANCE	
RPP	<i>Numéro du contrat d'assurance auquel l'assuré est rattaché</i>
DATE_DEB_RPP	<i>Date de début du contrat d'assurance</i>
DATE_FIN_RPP	<i>Date de fin du contrat d'assurance</i>

Dans le cadre de l'étude, nous avons sélectionné l'ensemble des assurés présents dans les entreprises sur la période allant du 01/01/2017 au 31/12/2019 (3 ans d'observations) et ayant bénéficié d'une garantie en cas de décès que leur contrat de travail soit terminé ou non durant cette période.

Etant donné qu'un individu peut avoir plusieurs employeurs sur la période étudiée ou le même employeur et plusieurs contrats de travail, naturellement nous avons plusieurs doublons des variables CLE (identifiant de l'individu) et SALARIE (identifiant de l'individu pour une entreprise donnée) dont le traitement est détaillé dans le paragraphe qui suit.

2.1.3 Traitement des données de DSN

Avant traitement des données, notre base contient 5 083 301 lignes.

Nous disposons de plusieurs lignes en doublons dans notre base dû d'une part à des cas d'individus qui cumulent plusieurs contrats de travail courts dans une seule et même entreprise et d'autre part, par des individus qui ont eu plusieurs employeurs pendant la durée d'observation, soit en même temps, soit sur des périodes différentes. Il est donc important de supprimer ces doublons afin de prendre en compte les particularités et éviter de biaiser l'information.

En effet, la comptabilisation de ces doublons ne devrait pas avoir d'impact significatif sur l'estimation des niveaux de mortalité mais pourrait conduire à une sous-estimation des intervalles de confiance.

2.1.3.1 Etape 1 : Suppression des doublons de la variable « SALARIE » identifiant d'un individu dans la même entreprise

Pour cette étape, nous allons segmenter notre base en 2 tables disjointes dont l'une sans doublons de la variable « SALARIE » (2 514 289 lignes) et l'autre avec tous les doublons de la variable « SALARIE » (2 569 012 lignes).

Etape 1.1 : Dans un premier temps, nous supprimons les lignes qui concernent la même période d'embauche pour le même salarié. Ces doublons peuvent s'expliquer par des renumérotations de police d'assurance. Exemple : un assuré dont le contrat de travail concerne la période du 01/01/N au 30/06/N+1 et qui est rattaché à 2 contrats d'assurance P1 et P2 avec P2 la police d'assurance qui remplace P1 (exemple en cas de resouscription ou renouvellement). Il y a 561 020 lignes supprimées à cette étape et 438 131 salariés qui ne sont plus en doublons.

Etape 1.2 : il reste 1 569 861 lignes à ajuster et qui concernent d'autres cas de doublons de la variable « SALARIE ». Pour les traiter, nous allons diviser notre base en 3 selon l'année de fin du contrat de travail :

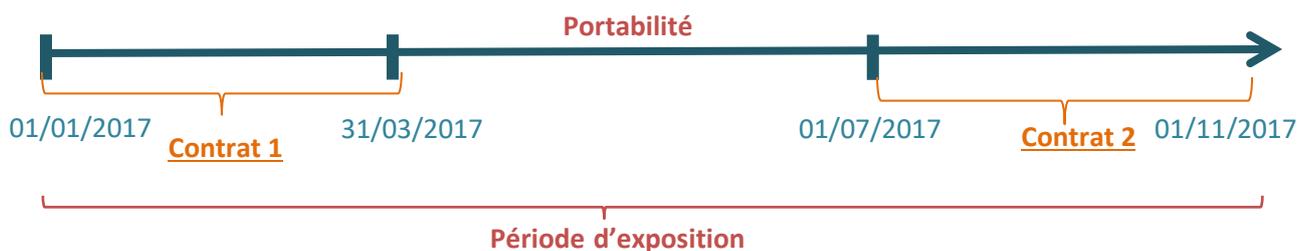
- Base1 : les salariés dont le contrat de travail se termine en 2017 (404 871 lignes)
- Base2 : les salariés dont le contrat de travail se termine en 2018 (510 926 lignes)
- Base3 : les salariés dont le contrat de travail se termine en 2019 (654 064 lignes)

A noter que tous les salariés dont les contrats de travail sont en vigueur au 31/12/2019 ont une date de fin forcée au 31/12/2019 correspondante à la date de fin d'observation.

Pour chacune de ces bases, nous regroupons pour chaque salarié, les périodes d'embauche qui se suivent ou qui se chevauchent avec une latitude de 3 jours pour prendre en compte les contrats qui se terminent la veille d'un week-end.

Dans ce traitement, nous prenons également en compte les périodes de portabilité pour lesquelles même si l'assuré n'est plus dans les effectifs, il reste couvert sur une période au moins égale à la durée de son précédent contrat et dans la limite de 12 mois (ANI).

Exemple : Pour un salarié avec 2 contrats avec contrat1 = du 01/01/2017 au 31/03/2017 et contrat2= du 01/07/2017 au 01/11/2017.



Une seule ligne conservée pour ce salarié avec une durée d'exposition au risque allant du **01/01/2017 au 01/11/2017**.

Figure 7: Traitement doublons1

Ce regroupement nous permet de passer de 404 871 à 132 732 lignes pour l'année 2017, de 510 926 à 160 823 lignes pour l'année 2018, de 654 064 à 204 914 lignes pour l'année 2019.

Etape 1.3 : fusion de ces 3 bases annuelles (489 469 lignes) et suppression des doublons concernant des contrats qui se suivent d'une année sur l'autre avec une latitude de 3 jours pour prendre en compte les contrats qui se terminent la veille d'un week-end.

Exemple : Pour un salarié avec 2 contrats avec contrat 1 = du 01/06/2017 au 29/12/2017 et contrat 2= du 01/01/2018 au 01/03/2018.



Une seule ligne conservée pour ce salarié avec une durée d'exposition au risque allant du **01/06/2017** au **01/03/2018**.

Figure 8: Traitement doublons2

Etape 1.4 : Après regroupement des périodes d'exposition dans les étapes 1.2 et 1.3, il nous reste 282 154 lignes en doublon (soit 11% de la base de doublons initiale). Pour cela, nous allons prendre pour chaque année d'observation (2017, 2018, 2019), la durée maximale sur laquelle le salarié est exposé et nous prenons l'hypothèse qu'il s'agit d'individus différents. Nous créons un nouvel identifiant SALARIE = ancien identifiant SALARIE + année d'observation (2017, 2018 ou 2019). Nous avons donc résolu le problème des salariés en doublon. Désormais nous avons une seule ligne pour un salarié donné et pour chaque entreprise (3 318 772 lignes).

2.1.3.2 Etape 2 : Suppression des doublons de la variable « CLE » identifiant d'un individu avec différents employeurs sur la période

Pour cette étape, nous allons segmenter notre base en 2 tables disjointes dont l'une sans doublons de la variable « CLE » (2 434 892 lignes) et l'autre avec tous les doublons de la variable « CLE » (883 880 lignes).

Etape 2.1 : Dans un premier temps, nous allons d'abord supprimer les lignes qui concernent des contrats de travail qui se chevauchent (même date de début de contrat de travail avec des dates de fin différentes ou même date de fin de contrat de travail avec des dates de début différentes pour des employeurs différents). Celui avec la période d'embauche la plus longue est conservée.

Exemple 1 : Pour un individu avec 2 contrats de travail pour 2 employeurs avec contrat 1 = du 01/01/2018 au 28/04/2018 et contrat 2= du 01/01/2018 au 29/12/2018



Une seule ligne conservée pour ce salarié avec une durée d'exposition au risque allant du **01/01/2018 au 29/12/2018**.

Figure 9: Traitement doublons3

Exemple 2 : Pour un individu avec 2 contrats de travail pour 2 employeurs avec contrat 1 = du 28/04/2018 au 29/12/2018 et contrat 2= du 01/07/2018 au 29/12/2018



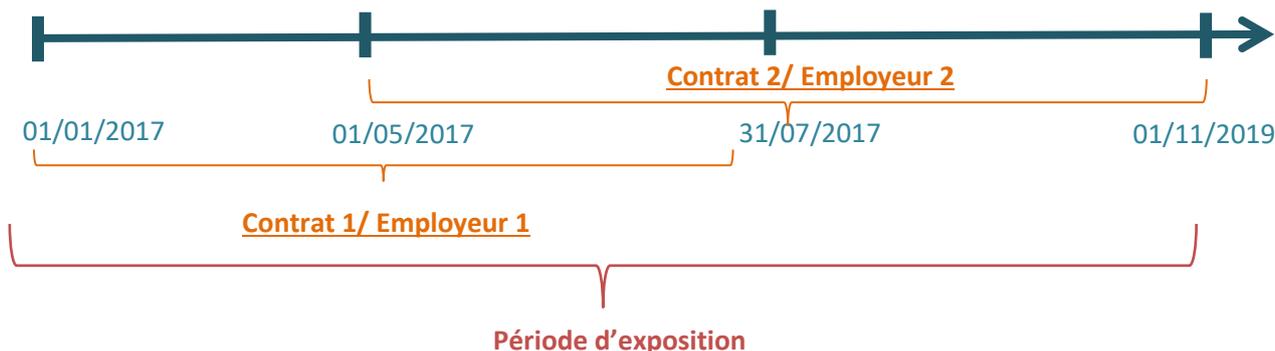
Une seule ligne conservée pour ce salarié avec une durée d'exposition au risque allant du **28/04/2018 au 29/12/2018**.

Figure 10: Traitement doublons

Nous passons de 883 880 lignes à 608 039 lignes.

Etape 2.2 : Nous traitons ensuite les cas de doublons concernant des contrats de travail qui se suivent ou qui se chevauchent pour différents employeurs avec une latitude de 3 jours pour prendre en compte les contrats qui se terminent la veille d'un week-end.

Exemple : Pour un individu avec 2 contrats avec contrat1 = du 01/01/2017 au 31/07/2017 et contrat2= du 01/05/2017 au 01/11/2019.



Une seule ligne conservée pour ce salarié avec une durée d'exposition au risque allant du 01/01/2017 au 01/11/2019.

Figure 11: Traitement doublons5

Etape 2.3 : A ce stade, il nous reste 268 931 lignes à ajuster.

Pour les traiter, nous allons diviser notre base en 3 selon l'année de fin du contrat de travail :

- les salariés dont le contrat se termine en 2017 (64 923 lignes),
- les salariés dont le contrat se termine en 2018 (96 775 lignes),
- les salariés dont le contrat se termine en 2019 (107 233 lignes).

Pour ces individus, nous retenons pour chaque année d'observation, la date de début du contrat de travail le plus ancien et la date de fin de contrat de travail la plus récente. Cette hypothèse peut légèrement surestimer la période d'exposition en fonction de la date d'anniversaire des individus mais ces cas concernent uniquement 8% des salariés.

Exemple pour l'année 2017 : Pour un salarié avec 2 contrats avec contrat 1 = du 01/04/2017 au 29/08/2017 et contrat 2= du 01/10/2017 au 31/12/2017.



Une seule ligne conservée pour ce salarié avec une durée d'exposition au risque allant du 01/04/2017 au 31/12/2017.

Figure 12: Traitement doublons6

Etape 2.4 : Après les traitements précédents, il nous reste 228 715 lignes en doublons. Pour cela, nous prenons l'hypothèse qu'il s'agit d'individus différents. Nous créons un nouvel identifiant CLE = ancien identifiant CLE + année d'observation (2017, 2018 ou 2019).

Après traitement de tous les doublons, notre base contient 2 954 957 lignes et autant d'individus.

Pour finir, nous allons retirer de notre base de données, les observations aberrantes (5 753 lignes : âges extrêmes, contrat d'assurance hors période, date de naissance manquante...).

Le tableau ci-après récapitule les différentes étapes du traitement des données :

Tableau 4: Récapitulatif des traitements

<i>Base de données avant retraitement</i>	5 083 301 lignes
<i>Traitement des doublons de la variable « SALARIE » identifiant d'un individu dans la même entreprise</i>	
<i>Suppression des lignes sur la même période d'embauche</i>	-561 020 lignes
<i>Regroupement des contrats de travail qui se suivent et prise en compte de la période de portabilité</i>	-1 133 894 lignes
<i>Autres cas</i>	-69 615 lignes
TOTAL	-1 764 529 lignes
Nombre de lignes après traitement des doublons de la variable « SALARIE »	3 318 772 lignes
<i>Traitement des doublons de la variable « CLE » identifiant d'un individu avec différents employeurs sur la période</i>	
<i>Conservation d'une ligne pour des individus avec la même date de début de contrat de travail et dates de fin de contrat de travail différentes ou inversement avec des employeurs différents (conservation de l'exposition la plus longue)</i>	-275 841 lignes
<i>Regroupement des contrats de travail qui se chevauchent ou se suivent</i>	-87 974 lignes
TOTAL	-363 815 lignes
Nombre de lignes après traitement des doublons de la variable « CLE »	2 954 957 lignes
<i>Suppression des valeurs aberrantes (âges extrêmes, contrat d'assurance hors période, date de naissance manquante...)</i>	-5 753 lignes
Base finale après retraitement	2 949 204 individus dont 3 962 décès

2.1.4 Fiabilisation des décès dans les données de DSN

Bien que les bases DSN permettent d'avoir à la fois l'exposition et les sinistres décès, nous les comparons à nos bases de sinistres internes, pour compléter éventuellement l'information sur le décès de l'assuré, dans certains cas. Pour ce faire, nous avons utilisé une clé unique construite à partir du contrat d'assurance, de la date de naissance de l'assuré et du sexe de l'assuré, permettant de faire le lien entre ces 2 bases.

- **Cas 1** : Fin du contrat avec un motif différent du décès et survenance du décès dans les 30 jours

Lorsqu'un assuré décède, l'employeur n'envoie plus la DSN associée. Selon la date de survenance du décès dans le mois, le motif de fin de contrat de travail du dernier envoi peut être différent du décès. Nous avons donc réajusté ces données en corrigeant l'information sur le décès directement dans la base DSN pour ceux dont le délai entre la date de fin de contrat et la date de survenance du sinistre ne dépasse pas 30 jours.

Exemple : Pour un assuré dont la fin du contrat est le 30/06/N avec un motif de fin de contrat différent du décès, si dans le système de gestion interne, nous avons une date de décès renseignée entre le 01/07/N et le 31/07/N, alors le motif de sortie dans la DSN sera corrigé comme un décès.

Cela concerne 197 assurés décédés.

- **Cas 2** : Salariés en arrêt de travail au moment du décès et dont le contrat d'assurance est en cours

Ces décès peuvent être topés dans les bases de sinistres internes mais non renseignés dans les bases DSN. Cela peut être le cas de salariés en état d'invalidité, bénéficiaires de la garantie maintien décès et dont les contrats de travail sont rompus. Pour cela, l'employeur n'envoie plus de DSN, nous allons donc les intégrer dans la base DSN en ajoutant comme motif de sortie le décès uniquement pour ceux dont les contrats d'assurance sont en cours. Cela concerne 1 059 assurés décédés.

Ces réajustements ont permis d'augmenter les décès de 32%, nous disposons désormais de 5 218 décès dans la base DSN.

2.2 Analyse descriptive des données

Notre base de modélisation contient 2 950 263 assurés et est composé de 54% d'hommes et 46% de femme. Elle compte 5 218 décès sur la période allant du 01/01/2017 au 31/12/2019. L'âge moyen à l'exposition est de 41 ans et l'âge moyen au décès de 51 ans.

Répartition des données par sexe

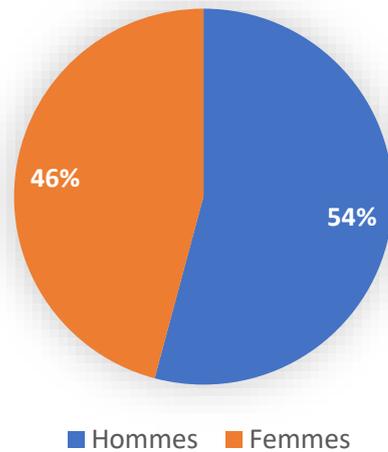


Figure 13: Répartition des données par sexe

Répartition des données par sexe et par catégorie socioprofessionnelle

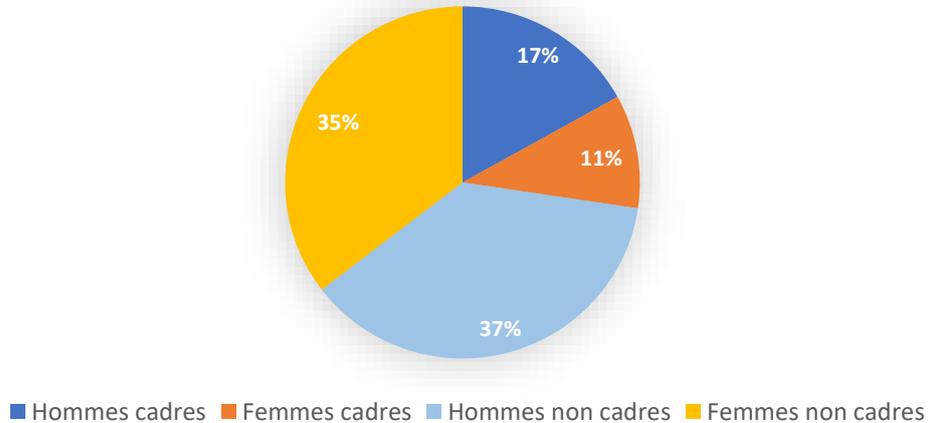


Figure 14: Répartition des données par sexe et par CSP

La répartition par catégorie socio-professionnelle nous indique que le portefeuille est composé essentiellement de non-cadres qui pèsent pour 72% (dont 37% d'hommes et 35% de femmes). Les cadres représentent 28% du portefeuille (dont 17% d'hommes et 11% de femmes).

Répartition des décès par sexe et par catégorie socioprofessionnelle

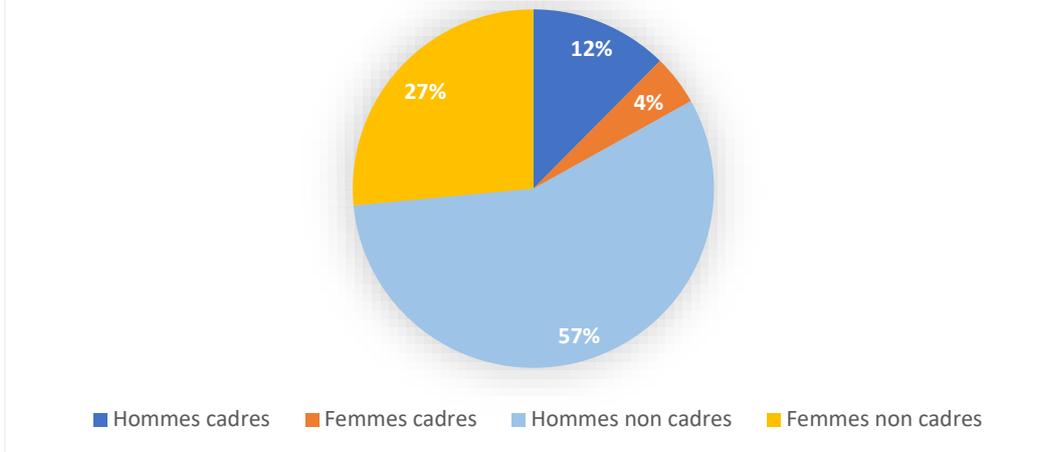


Figure 15: Répartition des décès par sexe et par CSP

La répartition des décès par sexe et catégorie socio-professionnelle montre une mortalité plus importante chez les hommes non-cadres qui représentent 57% des décès pour 37% des observations. Cela s'explique par la pénibilité des métiers exercés par ceux-ci notamment les ouvriers mais également leur mode de vie.

De plus, ils ont généralement un accès aux soins plus limité comparé aux cadres.

Compte tenu de ces statistiques, nous étudierons l'influence du sexe et de la catégorie socio-professionnelle dans la mortalité du portefeuille avec le modèle relationnel de Cox. Celui-ci prendra comme mortalité de référence celle des hommes non-cadres (voir paragraphe 2.3.2).

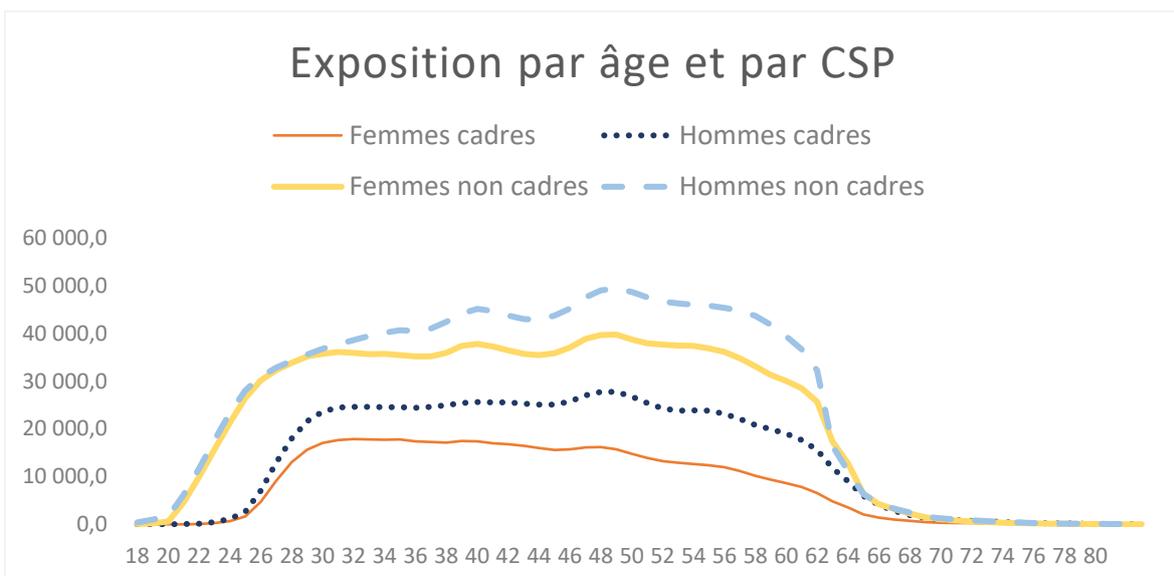


Figure 16: Exposition par âge, sexe et CSP

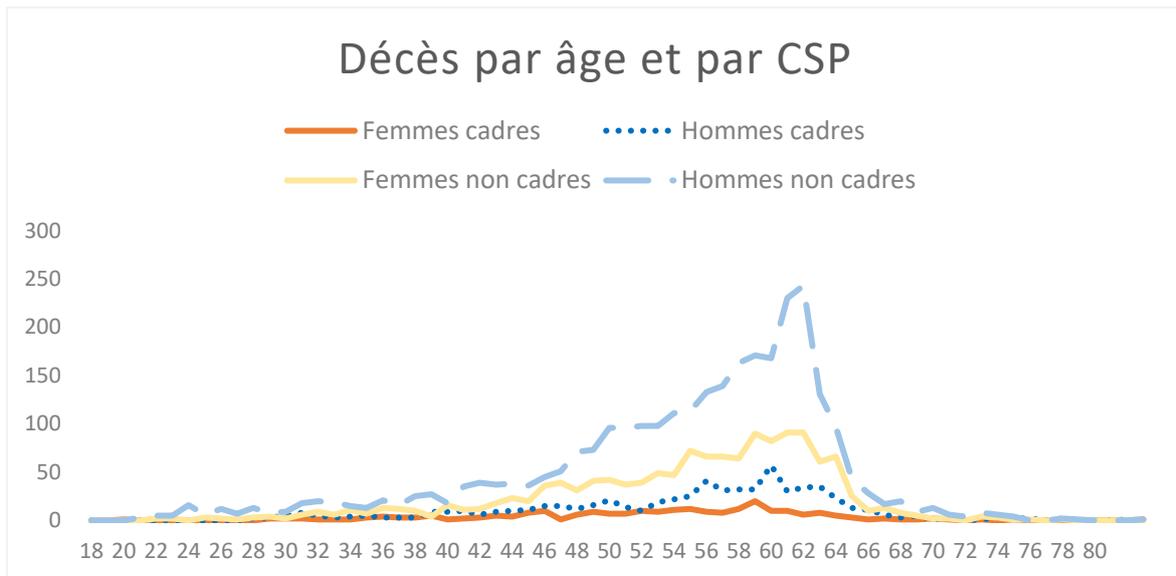


Figure 17: Décès par âge, sexe et CSP

Les deux graphiques ci-dessus représentent la répartition des décès et de l'exposition par âge, par sexe et par catégorie socio-professionnelle de la population étudiée (salariés actifs). Les données les plus représentatives sont celles des âges allant de 19 à 70 ans, correspondant généralement, aux âges d'entrée dans la vie active et de départ à la retraite.

Par ailleurs, le critère de validation de données de Cochran (détaillé en annexe 1) est respecté pour ces âges. Ce critère permet par la suite d'appliquer certaines hypothèses dans les algorithmes d'estimation de paramètres et de construction d'intervalles de confiance de ces derniers.

2.3 Construction des tables de mortalité d'expérience

L'objectif de cette section est de construire des tables de mortalité segmentées par sexe et par catégorie socio-professionnelle (hommes cadres, femmes cadres, hommes non-cadres, femmes non-cadres) avec des méthodes d'estimation paramétriques et non paramétriques.

Des tests statistiques sont ensuite effectués pour valider les modèles.

2.3.1 Introduction à l'analyse de survie

2.3.1.1 *Rappel de fonctions usuelles*

Avant d'entamer la construction de nos tables de mortalité, nous allons d'abord introduire certaines notions mathématiques importantes pour mieux assimiler l'analyse de survie. Nous cherchons à étudier la durée de vie d'un individu correspondant au temps écoulé de la naissance jusqu'à la

survenance du décès. Prenons l'hypothèse qu'il s'agit d'une variable aléatoire $T \geq 0$, alors sa distribution de probabilité peut être obtenue soit par sa fonction de répartition ou sa fonction de survie. Elle peut être également définie par sa densité, sa fonction de hasard instantanée ou celle cumulée.

Fonction de répartition F

Soit $t > 0$ fixé, la fonction de répartition constitue la probabilité pour un individu de mourir avant t . Elle s'écrit :

$$F(t) = P(T \leq t).$$

Fonction de survie S

Pour t fixé, la fonction de survie est la probabilité pour un individu de survivre après l'instant t , autrement dit :

$$S(t) = P(T \geq t) = 1 - F(t).$$

Fonction de survie conditionnelle

Soit t et x fixés avec $t \geq x$, la fonction de survie conditionnelle de T est la probabilité pour un individu de survivre après l'instant t sachant qu'il a survécu jusque l'instant x . Elle est définie par :

$$S(t/x) = P(T \geq t / T \geq x) = \frac{S(t)}{S(x)}.$$

Densité de probabilité f

Il s'agit de la fonction $f(t) \geq 0$ telle que pour tout $t \geq 0$:

$$F(t) = \int_0^t f(x) dx.$$

Lorsque la fonction de répartition F admet une dérivée au point t alors :

$$f(t) = \lim_{h \rightarrow 0} \frac{P(t \leq T < t + h)}{h} = F'(t) = -S'(t).$$

Pour t fixé, la densité de probabilité correspond à la probabilité pour un individu de mourir dans un petit intervalle de temps après l'instant t .

Risque instantané μ ou fonction de hasard

Pour t fixé, le risque instantané correspond à la probabilité pour un individu, de mourir dans un petit intervalle de temps après l'instant t , sachant qu'il a survécu jusqu'au temps t . Nous avons 2 définitions différentes du risque instantané selon le cas continu ou le cas discret.

- **Cas discret** : la variable aléatoire T prend des valeurs dans un ensemble discret (qui est dénombrable mais pas nécessairement fini) $\tau = \{t_1, \dots, t_n, \dots\}$,

$$\mu(t) = P(T = t / T \geq t) = \frac{P(T = t)}{P(T \geq t)}, \text{ avec la convention } \frac{0}{0} = 0.$$

Nous avons
$$S(t) = \prod_{t_i < t} (1 - \mu(t_i)).$$

Ce résultat est important pour comprendre la construction de l'estimateur de Kaplan Meier que nous détaillerons par la suite.

- **Cas continu** : la fonction de hasard est définie par

$$\mu(t) = \lim_{h \rightarrow 0} \frac{P(t \leq T < t + h / T \geq t)}{h}.$$

En appliquant la formule de Bayes, nous pouvons réécrire l'expression sous la formule suivante :

$$\mu(t) = \lim_{h \rightarrow 0} \frac{P(t \leq T < t + h \cap T \geq t)}{h P(T \geq t)} = \lim_{h \rightarrow 0} \frac{P(t \leq T < t + h)}{h P(T \geq t)}.$$

Avec les expressions de $f(t)$ et $S(t)$ définies dans les paragraphes précédents, on en déduit que :

$$\mu(t) = \frac{f(t)}{S(t)} = -\frac{S'(t)}{S(t)} = -\ln(S(t))'.$$

Fonction de hasard cumulée

La fonction de hasard cumulée est définie par :

1. $M(t) = \sum_{t_i \leq t} \mu(t_i)$ dans le cas discret.
2. $M(t) = \int_0^t \mu(u) du$ dans le cas continu.

Nous avons uniquement dans le cas où T est continue $S(t) = \exp(-M(t))$.

2.3.1.2 Tables réglementaires et notations actuarielles

Les tables de mortalité réglementaires sont construites sur des groupes fermés de la population française. Elles donnent pour chaque âge, le nombre de survivants du groupe et permettent ainsi, de calculer le taux annuel de mortalité et l'espérance de vie.

Notons :

l_x = nombre de survivants d'âge x ,

l_0 = 100 000 individus d'âge 0 constituant l'effectif initial du groupe fermé,

d_x = nombre de décès à l'âge x ,

p_x = probabilité de survivre à l'âge x ,

q_x = probabilité de décéder entre l'âge x et $x + 1$ d'un survivant à l'âge x .

Nous avons :

- La fonction de survie à l'âge x qui s'écrit :

$$S(x) = \frac{l_x}{l_0},$$

- La probabilité pour un individu de survivre jusque l'âge $x + t$ sachant qu'il a survécu jusque l'âge x s'écrit :

$${}_t p_x = \frac{l_{x+t}}{l_x} = P(T \geq x + t / T \geq x) .$$

- La probabilité pour un individu de mourir entre l'âge x et $x + t$ sachant qu'il a survécu jusque l'âge x s'écrit :

$${}_t q_x = 1 - {}_t p_x .$$

- On en déduit le taux annuel de mortalité d'un individu d'âge x :

$$q_x = P(x \leq T < x + 1 / T \geq x) = 1 - \frac{S(x+1)}{S(x)} .$$

Ce taux annuel de mortalité q_x est souvent l'élément modélisé dans la construction des tables de mortalité.

Exemple de tables de mortalité réglementaires

Les tables les plus utilisées pour le décès (hors rente viagère) sont la TH 00-02 et la TF 00-02 qui sont calibrées par âge (de 0 à 120 ans) sur les populations masculine et féminine française de 2 000 à 2 002 par l'INSEE.

Les tableaux et graphiques ci-dessous permettent d'illustrer ces 2 tables.

Tableau 5: Extrait des tables TH00-02 et TF00-02

Table de mortalité TH 00-02 (jusqu' à 30 ans)					Table de mortalité TF 00-02 (jusqu' à 30 ans)				
Age	L_x	d_x	p_x	q_x	Age	L_x	d_x	p_x	q_x
0	100 000	489	99,51%	0,49%	0	100 000	384	99,62%	0,38%
1	99 511	38	99,96%	0,04%	1	99 616	33	99,97%	0,03%
2	99 473	27	99,97%	0,03%	2	99 583	21	99,98%	0,02%
3	99 446	22	99,98%	0,02%	3	99 562	17	99,98%	0,02%
4	99 424	18	99,98%	0,02%	4	99 545	14	99,99%	0,01%
5	99 406	16	99,98%	0,02%	5	99 531	12	99,99%	0,01%
6	99 390	14	99,99%	0,01%	6	99 519	11	99,99%	0,01%
7	99 376	13	99,99%	0,01%	7	99 508	10	99,99%	0,01%
8	99 363	13	99,99%	0,01%	8	99 498	10	99,99%	0,01%
9	99 350	12	99,99%	0,01%	9	99 488	10	99,99%	0,01%
10	99 338	13	99,99%	0,01%	10	99 478	11	99,99%	0,01%
11	99 325	13	99,99%	0,01%	11	99 467	11	99,99%	0,01%
12	99 312	16	99,98%	0,02%	12	99 456	12	99,99%	0,01%
13	99 296	20	99,98%	0,02%	13	99 444	13	99,99%	0,01%
14	99 276	26	99,97%	0,03%	14	99 431	16	99,98%	0,02%
15	99 250	37	99,96%	0,04%	15	99 415	20	99,98%	0,02%
16	99 213	50	99,95%	0,05%	16	99 395	24	99,98%	0,02%
17	99 163	66	99,93%	0,07%	17	99 371	29	99,97%	0,03%
18	99 097	82	99,92%	0,08%	18	99 342	33	99,97%	0,03%
19	99 015	94	99,91%	0,09%	19	99 309	35	99,96%	0,04%
20	98 921	101	99,90%	0,10%	20	99 274	35	99,96%	0,04%
21	98 820	104	99,89%	0,11%	21	99 239	34	99,97%	0,03%
22	98 716	104	99,89%	0,11%	22	99 205	34	99,97%	0,03%
23	98 612	103	99,90%	0,10%	23	99 171	34	99,97%	0,03%
24	98 509	103	99,90%	0,10%	24	99 137	34	99,97%	0,03%
25	98 406	103	99,90%	0,10%	25	99 103	35	99,96%	0,04%
26	98 303	105	99,89%	0,11%	26	99 068	35	99,96%	0,04%
27	98 198	107	99,89%	0,11%	27	99 033	36	99,96%	0,04%
28	98 091	109	99,89%	0,11%	28	98 997	37	99,96%	0,04%
29	97 982	112	99,89%	0,11%	29	98 960	39	99,96%	0,04%
30	97 870	114	99,88%	0,12%	30	98 921	42	99,96%	0,04%

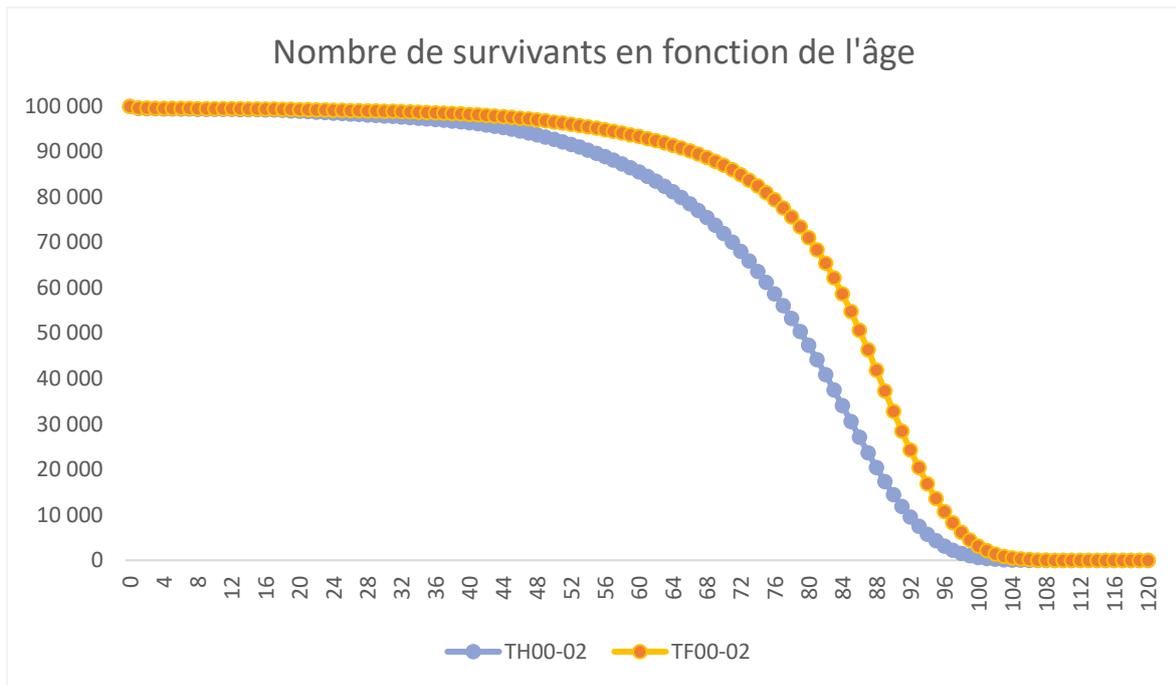


Figure 18: Nombre de survivants des tables TH00-02 et TF00-02

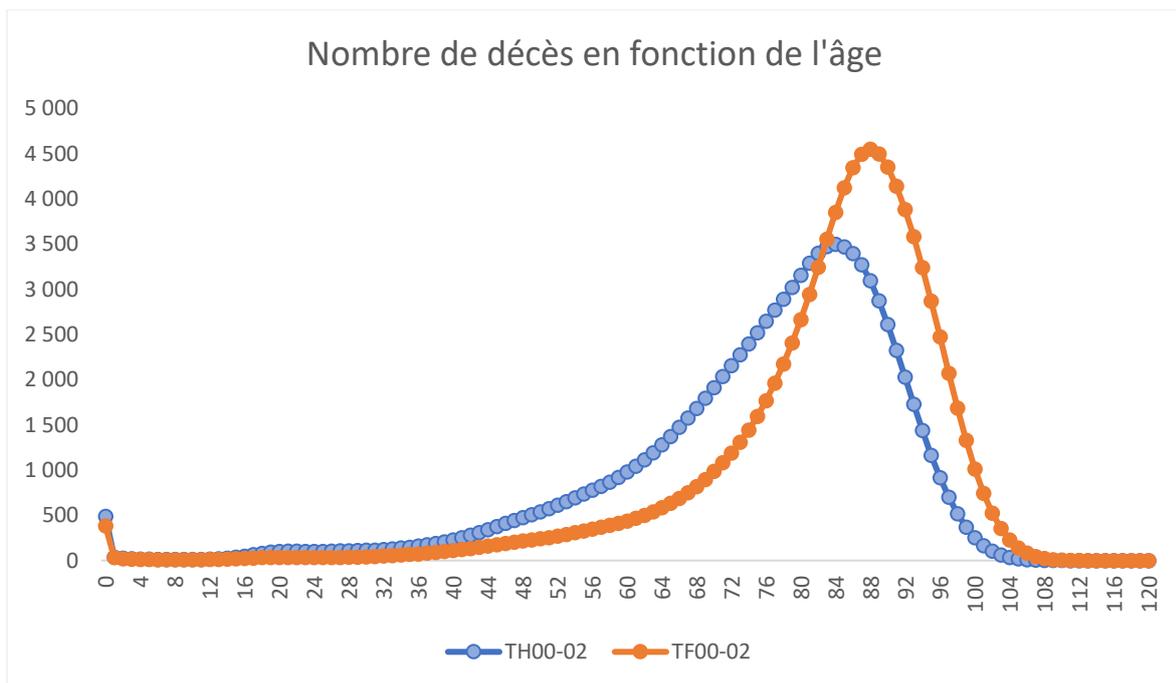


Figure 19: Nombre de décès des tables TH00-02 et TF00-02

L'analyse de ces graphiques montre une mortalité plus importante chez les hommes que chez les femmes.

2.3.1.3 Censures et troncatures

Compte tenu du temps « long » de collecte de données pour étudier la variable aléatoire T représentant la durée de vie d'un individu, des événements peuvent survenir pendant la période d'observation et rendent ainsi la construction d'inférence statistique très spécifique.

Ces phénomènes parasites correspondent à des censures et des troncatures. Dans le cas particulier de notre portefeuille d'assurance collective, les censures correspondent par exemple à des fins de contrats de travail ou d'assurance, des départs à la retraite, des démissions ou ruptures conventionnelles et les troncatures par exemple à des nouveaux embauchés, des nouveaux contrats d'assurance qui entrent en vigueur au cours de la période d'observation.

Il est important de prendre en compte ces événements dans l'étude de survie car leur exclusion conduirait à une perte d'information et un biais sur l'estimation de la durée de survie T (sous-estimations ou surestimations de valeurs atypiques de T).

Censure à droite

Soit (T_1, \dots, T_n) un n-échantillon de variables indépendantes et identiquement distribuées (i.i.d.) de la variable aléatoire T et (C_1, \dots, C_n) un n-échantillon i.i.d de la variable aléatoire censurée C .

En cas de censure à droite, (T_1, \dots, T_n) ne sont pas observées mais $(Y_1, \delta_1, \dots, Y_n, \delta_n)$ avec :

- $Y_i = \inf(T_i, C_i)$,
- $\delta_i = 1_{T_i \leq C_i}$.

Dans notre cas qui concerne la construction de table de mortalité, C_i correspond à l'âge auquel i quitte l'observation pour une cause différente du décès et Y_i à l'âge auquel i quitte l'observation quelle que soit la cause (décès ou autre raison).

Le graphique ci-dessous permet d'illustrer la censure à droite.

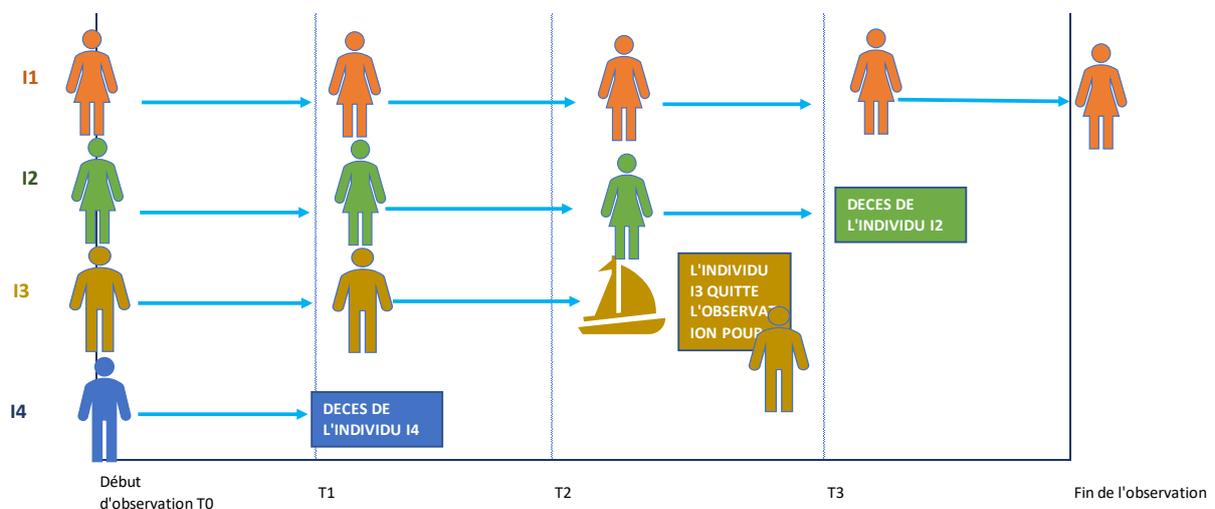


Figure 20: illustration de la censure à droite

Les individus I2 et I4 sont dits non censurés car nous avons observés leurs décès. Les individus I1 et I3 sont dits censurés à droite : l'individu I3 a quitté le groupe en cours d'observation sans que nous ayons observé le décès (exemple : départ à la retraite) et l'individu I1 est présent jusqu'à la fin de l'observation mais le décès n'a pas été observé.

Troncature

La troncature concerne les individus qui entrent dans le champ de l'étude au cours de la période d'observation ou des individus qui n'ont jamais été observés.

En règle générale, la troncature a lieu lorsqu'au lieu d'observer un n-échantillon (T_1, \dots, T_n) de variables i.i.d. ayant la même distribution que la variable aléatoire T notée $\varphi(T)$, on observe (T_1, \dots, T_n) avec la distribution $\varphi(T / T \in A)$, avec A un ensemble défini.

Dans ce cas, il est impossible de récupérer la distribution complète de la variable aléatoire T car une partie de celle-ci n'a pas été observée. Dans le cas de notre étude des données de survie, nous observons principalement les cas de troncature à gauche.

Pour un individu avec une durée de vie totale T , définissons C comme l'âge auquel il quitte l'observation pour une autre cause différente de la mort (moment de censure) et τ l'âge auquel l'individu commence à être observé. Cela implique que l'individu n'est observé que si T et C sont supérieurs à τ .

Les individus avec de petites valeurs de T sont plus susceptibles d'être tronqués, donc négliger la troncature entraînera une surestimation des valeurs atypiques prises par T .

Il existe des moyens de corriger ces cas particuliers de troncature, cependant la densité de τ doit être strictement positive sur un certain intervalle $[0, t_0]$.

Censure à droite et troncature à gauche

Soit (τ_1, \dots, τ_n) un n-échantillon de variables indépendantes et identiquement distribuées (i.i.d.) que la variable aléatoire τ :

En cas de censure à droite et troncature à gauche, sont observées : $(Y_1, \delta_1, \tau_1 \dots, Y_n, \delta_n, \tau_n)$ avec :

- $Y_i = \inf(T_i, C_i)$,
- $\delta_i = 1_{T_i \leq C_i}$,
- Pour tout $i \in \{1, \dots, n\}, \tau_i \leq Y_i$.

Avec τ_i correspondant à l'âge auquel l'individu i commence à être observé.

Le graphique ci-dessous permet d'illustrer ces situations.

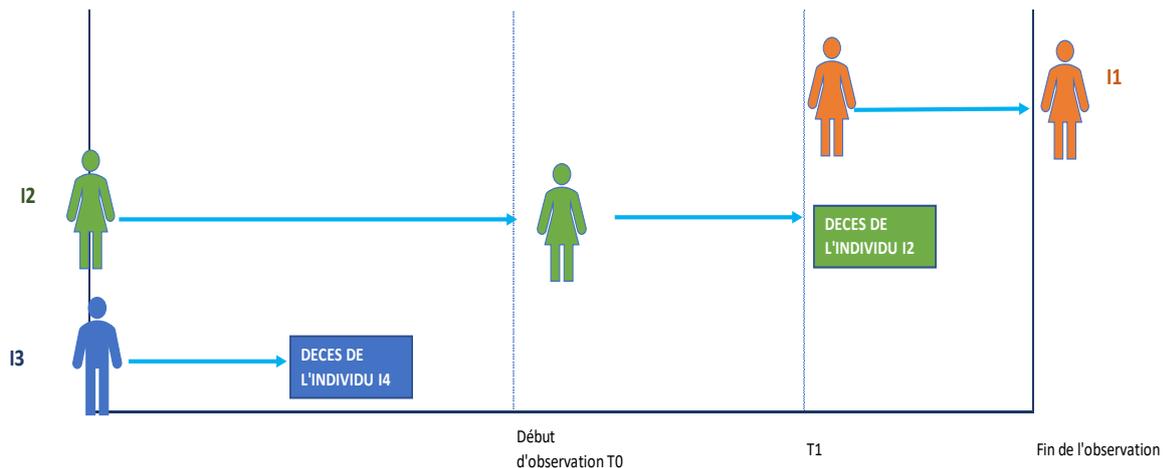


Figure 21: Illustration de la censure et de la troncation

Dans le graphique ci-dessous, l'individu I1 est tronquée à gauche et censurée à droite. L'individu I3 est quant à lui tronqué et n'a jamais été observé.

D'autres cas de censures (ex : censures par intervalle) et de troncation existent mais ne seront pas présentés dans le cadre de notre étude car ne concernent pas directement notre problématique.

Par la suite, toute la modélisation de la mortalité se fera avec le logiciel R et à l'aide des fonctions disponibles sur le site ressources actuarielles.net.

2.3.2 Modélisation des taux de mortalité

2.3.2.1 Estimation des taux bruts par la méthode des moments de Hoem

L'estimateur des moments de Hoem est un estimateur paramétrique prenant en compte les censures et les troncatures. Contrairement à l'estimateur de Kaplan Meier, il ne permet pas l'estimation de la fonction de survie mais détermine directement les taux bruts.

Cet estimateur repose sur plusieurs hypothèses fortes dont la linéarité de la distribution des décès entre les âges x et $x + 1$.

Cadre théorique et hypothèses

Soit :

- n_x , le nombre d'individus en vie entre les âges x et $x + 1$.
- d_x , le nombre de décès entre les âges x et $x + 1$.

Pour chaque individu i , on note : $[de_i; fe_i] \subset]x; x + 1]$, son exposition

avec de_i et fe_i correspondant respectivement à la date de début et de fin d'observation pour l'individu i d'âge x .

- X_i : la variable aléatoire représentant le décès de l'individu i (variable indicatrice qui vaut 1 si l'individu décède et 0 sinon). Les X_i sont indépendants et identiquement distribués.

Etant donné que les décès sont supposés linéaires sur $]x ; x + 1]$, elles suivent une loi de Bernoulli de paramètre $(fe_i - de_i) \times q_x$.

On en conclut que :

le nombre de décès à l'âge $x = \sum_{i=1}^{n_x} X_i$ suit une loi **Binomiale de paramètres (n_x, \hat{q}_x)** ,

avec $\hat{q}_x = E(X_i) = (fe_i - de_i) \times q_x$.

La vraisemblance du modèle est donnée par :

$$L(q_x) = \prod_{i=1}^{n_x} C_{n_x}^k \times [(fe_i - de_i) \times q_x]^{d_x} \times [1 - (fe_i - de_i) \times q_x]^{n_x - d_x} .$$

L'estimateur \hat{q}_x du maximum de vraisemblance est donnée par :

$$\hat{q}_x = \frac{d_x}{\sum_i (fe_i - de_i)} .$$

Application à nos données

La méthode de Hoem appliquée à nos données donne les taux bruts estimés ci-dessous pour les segments hommes cadres, femmes cadres, hommes non-cadres et femmes non-cadres :

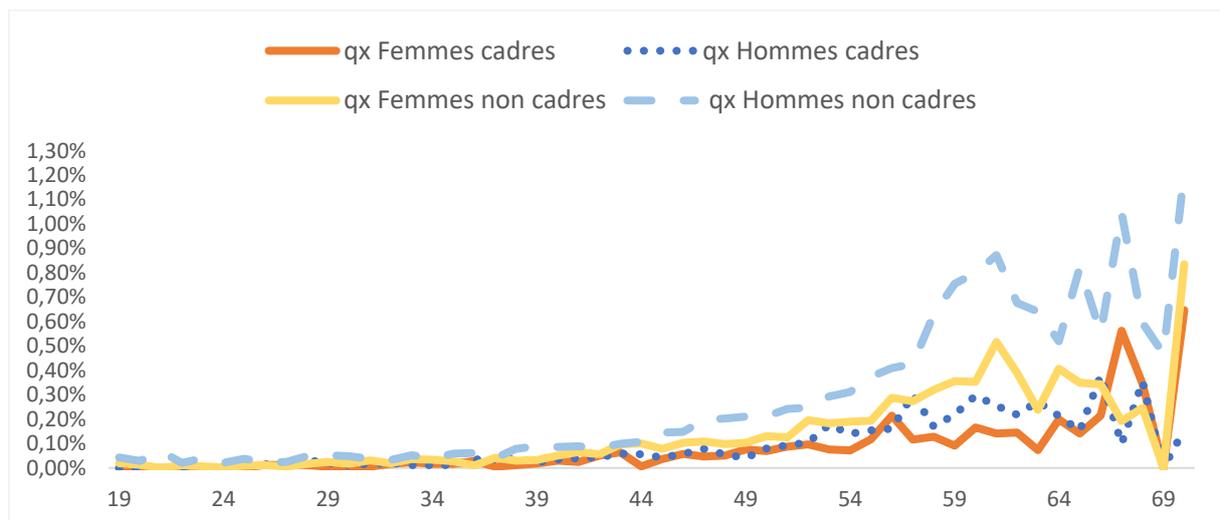


Figure 22: Estimation des taux bruts par la méthode de Hoem

Nous observons que les taux de mortalité des hommes non-cadres sont largement supérieurs à ceux des autres groupes notamment sur les âges de 45 à 70 ans. Ils sont en moyenne 2 fois plus élevés que ceux des hommes cadres et des femmes non-cadres. Les taux de mortalité observés sont très volatiles de 60 à 70 ans dû aux faibles expositions sur ces âges.

De plus, nous notons un faible nombre de décès voire une absence de décès pour les âges jeunes sur les autres groupes.

2.3.2.2 Influence du sexe et de la CSP sur la mortalité : Modèle de Cox

Comme évoqué dans les paragraphes précédents, l'analyse de la distribution de nos données montre une certaine disparité en termes de représentativité des décès et expositions selon le sexe et la CSP (ex : les hommes non-cadres représentent 57% des décès pour 37% des observations).

Dans cette partie, nous allons chercher à évaluer l'influence de la catégorie socio-professionnelle et du sexe sur la mortalité de notre portefeuille en utilisant le modèle relationnel de Cox.

Notre groupe de référence sera celui des hommes non-cadres pour lequel nous estimons que les données sont les plus robustes. En cas de validation du modèle de Cox par des tests statistiques, nous allons déduire la mortalité des autres groupes (hommes cadres, femmes cadres et femmes non - cadres) à partir de celle de référence avec une relation mathématique simple que nous présenterons dans le paragraphe qui suit.

Cadre théorique et hypothèses

Le modèle de Cox est un modèle semi paramétrique et est un cas particulier des modèles dits à hasard proportionnel (voir annexe 2). Il cherche à évaluer l'effet de variables explicatives sur la fonction de survie et suppose que pour 2 individus i et j , le rapport des fonctions de hasard définies ci-après ne varie pas au cours du temps.

Soit $\mu(t/z)$ la fonction de hasard de la variable aléatoire $T/Z = z$, le modèle de Cox prend l'hypothèse que :

$$\mu(t/Z) = \mu_0(t) \exp(\theta_0^T \times Z).$$

Avec $\theta_0 \in R^d$ un paramètre inconnu (coefficient de régression), μ_0 la fonction de hasard de référence qui est inconnue également (lorsqu'elle est connue le modèle devient complètement paramétrique) et Z un vecteur de covariables de dimension $d \times 1$.

La théorie mathématique sous-jacente à l'estimation du paramètre θ_0 est fondée sur une maximisation de vraisemblance partielle (voir annexe 2).

Si nous revenons dans notre problématique, nous avons 2 variables explicatives ($d=2$) avec

$Z_1 =$ la variable indicatrice qui vaut 1 si sexe = femme et 0 sinon ,

et $Z_2 =$ la variable indicatrice qui vaut 1 si CSP = cadre et 0 sinon .

A partir des relations suivantes :

$$\mu(x) = -\ln(S(x))' \text{ et,}$$

$$q_x = 1 - \frac{S(x+1)}{S(x)},$$

on peut déduire aisément l'expression du q_x pour chaque groupe à partir du modèle de Cox.

Nous aurons donc pour chaque groupe i , avec $i = 2, 3, 4$ et chaque âge x :

$$q_x^i = 1 - \frac{S_i(x+1)}{S_i(x)} = 1 - \left(\frac{S_1(x+1)}{S_1(x)} \right)^{\exp(\theta_0^T Z^i)} = 1 - (1 - q_x^1)^{\exp(\theta_0^T Z^i)}.$$

Avec q_x^1 la mortalité de notre groupe de référence qui est celle des hommes non-cadres et Z^i les valeurs des indicatrices associées au groupe i .

Application à nos données

L'application du modèle de Cox à nos données nous donne les estimations de coefficients suivantes :

Tableau 6: Estimation des paramètres du modèle de Cox

Variables explicatives	θ_0 estimé	$\exp(\theta_0)$ estimé	Ecart-type	Probabilité critique p-value
<i>femme</i>	-0.53	0.59	0.030	<2e-16
<i>cadre</i>	-0.87	0.42	0.037	<2e-16

L'interprétation des coefficients estimés nous indique que la fonction de survie des hommes non-cadres décroît plus rapidement que celle des autres groupes.

Elle nous montre également que les femmes ont des probabilités conditionnelles de décès inférieures à celles des hommes d'environ 41 % (1-0.59) et les cadres ont des probabilités conditionnelles de décès inférieures à celles des non-cadres d'environ 58% (1-0.42), ce qui constitue un ordre de grandeur usuel sur la plage d'âge étudiée.

Validation du modèle : Test de significativité des coefficients

Pour valider la significativité du coefficient global θ_0 , nous allons utiliser le test statistique de Wald qui mesure l'écart entre θ_0 et $\theta_0' = 0$.

Le test se présente sous la forme suivante :

$$\begin{cases} H_0: \theta_0 = 0 \text{ contre,} \\ H_1: \text{au moins une des composante de } \theta_0 \text{ est non nulle.} \end{cases}$$

Sous l'hypothèse H_0 , la statistique de test de Wald suit une loi du Chi-2 à 2 degrés de liberté.

Le logiciel R nous donne le résultat de statistiques supplémentaires que sont celle du rapport de vraisemblance et celle du score. La théorie statistique sur ces trois tests est développée en annexe 2. Ces trois tests rejettent l'hypothèse H_0 avec un risque d'erreur de 1%.

Tableau 7: Statistiques du test de significativité des coefficients (COX)

Test	Chi-2	Probabilité critique p-value
Rapport de Vraisemblance	894.4	<.0001
Score	828.8	<.0001
Wald	797	<.0001

Nous en concluons que le vecteur de coefficient θ_0 estimé par le modèle de Cox est non nul et significatif.

Validation du modèle : hypothèse de hasard proportionnel

L'hypothèse de hasard proportionnel est vérifiée lorsque pour 2 individus i et j qui ont pour covariables Z_i et Z_j , le rapport des fonctions de hasard ne varie pas au cours du temps. En d'autres termes, sous l'hypothèse H_0 :

$$\frac{\mu(t/Z_i)}{\mu(t/Z_j)} = \frac{\exp(\theta_0^T Z_i)}{\exp(\theta_0^T Z_j)}.$$

Il est important de valider cette hypothèse des risques proportionnels pour justifier le calibrage du modèle de Cox à nos données. Pour cela, nous nous basons sur la méthode des résidus de Schoenfeld. A chaque date de décès k , le résidu de Schoenfeld correspond à la différence entre les caractéristiques de l'individu décédé et une moyenne pondérée des caractéristiques des individus à risque au temps k . Si plusieurs décès surviennent au même instant, on somme les résidus de chaque individu.

Pour valider l'hypothèse de hasard proportionnel, les résidus doivent être répartis aléatoirement à l'intérieur d'une bande horizontale centrée en zéro (indépendance).

Le logiciel R avec le « package survminer » nous permet de réaliser ce test et nous fournit les résultats ci-après.

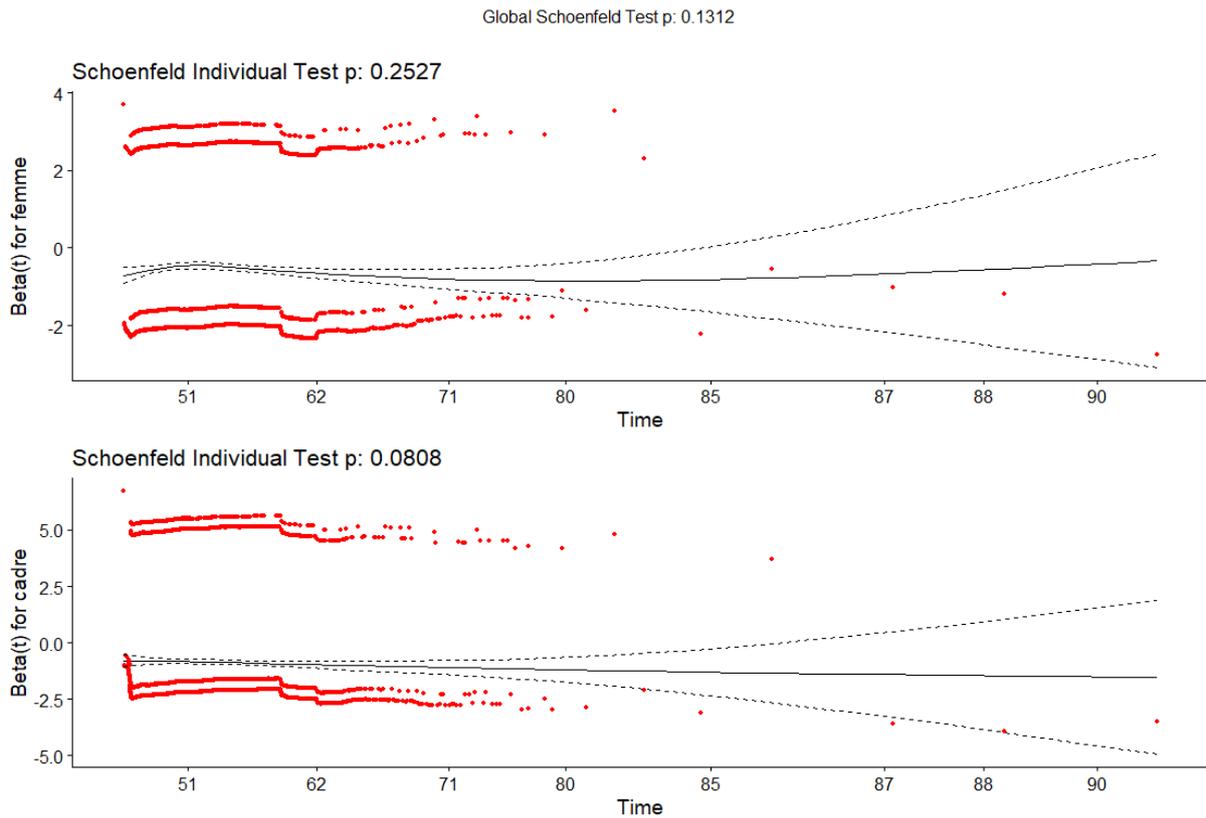


Figure 23: Test des résidus de Schoenfeld

L'interprétation des graphiques nous indique que les résidus ne présentent pas de tendance particulière ; nous en déduisons donc que le modèle de hasard proportionnel est validé.

De plus, le test global des résidus de Schönfeld nous donne une p-value supérieure à 5%, ce qui permet d'accepter l'hypothèse H_0 avec un risque d'erreur de 5%.

Compte tenu de la validation du modèle de Cox, nous allons par la suite procéder à l'estimation de la mortalité de notre groupe de référence qui est celle des hommes non-cadres par la méthode de Kaplan Meier. Un modèle adéquat sera ensuite choisi pour ce groupe de référence après lissage des taux bruts par Whittaker-Henderson et ajustement à une loi de Makeham. Les taux de mortalité des autres groupes en seront déduits à partir des coefficients obtenus par le modèle de Cox.

2.3.2.3 Estimation des taux bruts de mortalité des hommes non-cadres : méthode de Kaplan Meier

L'estimateur de Kaplan-Meier est un estimateur empirique non paramétrique de la fonction de survie permettant de tenir compte des censures et des troncatures. C'est un estimateur de maximum de vraisemblance généralisé et convergent. Il est non paramétrique car il permet d'approcher la forme empirique prise par le risque de décès, sans aucune hypothèse spécifique de loi de probabilité.

L'objectif des prochains paragraphes est de montrer comment estimer la fonction de survie en présence de troncature et de censure et comment est construit l'estimateur de Kaplan Meier.

Cadre théorique et hypothèses

Comment approcher la fonction de distribution empirique ?

Dans un cas classique d'inférence statistique, nous observons les variables (T_1, \dots, T_n) ayant la même distribution que la variable aléatoire T . Il est donc possible d'estimer la distribution de la variable aléatoire de façon non paramétrique avec des hypothèses faibles (à noter que cette estimation non paramétrique converge vers la vraie distribution) :

- La fonction de répartition $F(t) = P(T \leq t)$ est estimée par la fonction suivante :

$$\hat{F}_{emp}(t) = \frac{1}{n} \sum_{i=1}^n 1_{T_i \leq t} .$$

- La moyenne empirique estimée par : $\frac{1}{n} \sum_{i=1}^n T_i = \int t d\hat{F}_{emp}(t)$.
- La variance empirique estimée par :

$$\frac{1}{n} \sum_{i=1}^n T_i^2 - \left(\frac{1}{n} \sum_{i=1}^n T_i \right)^2 = \int t^2 d\hat{F}_{emp}(t) - \left(\int t d\hat{F}_{emp}(t) \right)^2 .$$

Dans le cas d'analyse de survie, avec présence de censure et de troncature, nous n'observons pas les variables (T_1, \dots, T_n) mais les variables $(Y_1, \delta_1, \tau_1, \dots, Y_n, \delta_n, \tau_n)$.

Rappelons que :

- $Y_i = \inf(T_i, C_i)$,
- $\delta_i = 1_{T_i \leq C_i}$,
- Pour tout $i \in \{1, \dots, n\}$, $\tau_i \leq Y_i$.

τ_i correspondant à l'âge auquel l'individu i commence à être observé,

C_i l'âge auquel i quitte l'observation pour une cause différente du décès et

Y_i à l'âge auquel i quitte l'observation quelle que soit la cause (décès ou pas décès).

Nous cherchons donc à approcher la fonction de distribution empirique dans le cas particulier de présence de censure et de troncature par d'autres moyens. Pour cela, nous allons introduire les fonctions de distributions suivantes :

$$\hat{H}_1(t) = \frac{1}{n} \sum_{i=1}^n \delta_i 1_{Y_i \leq t} ,$$

$$\hat{H}_0(t) = \frac{1}{n} \sum_{i=1}^n (1 - \delta_i) 1_{Y_i \leq t} ,$$

$$\hat{H}(t) = \frac{1}{n} \sum_{i=1}^n 1_{Y_i \leq t} = \hat{H}_0(t) + \hat{H}_1(t).$$

Leurs fonctions limites sont données par :

$$H_1(t) = P(Y \leq t, \delta = 1 / Y \geq \tau),$$

$$H_0(t) = P(Y \leq t, \delta = 0 / Y \geq \tau),$$

$$H(t) = P(Y \leq t / Y \geq \tau).$$

Construction de l'estimateur de Kaplan Meier

Le but est de trouver un estimateur \hat{F} de la fonction F , extension de la fonction \hat{F}_{emp} en cas présence de troncature et de censure.

\hat{F}_{emp} définit une fonction de distribution discrète de $\{T_1, \dots, T_n\}$ et nous cherchons la distribution discrète de $\{Y_1, \dots, Y_n\}$.

Soit $\hat{\mu}$ la fonction de hasard de \hat{F} , rappelons que dans le cas discret :

$$\hat{F}(t) = 1 - \prod_{Y_i \leq t} (1 - \hat{\mu}(Y_i)).$$

Le principe de l'estimateur de Kaplan Meier est de d'exprimer la fonction de hasard μ en fonction de H_0, H_1, H et de les remplacer par leurs versions empiriques pour obtenir $\hat{\mu}$ et en déduire \hat{F} .

Nous supposons dans la suite que les variables aléatoires T et C sont indépendantes. C'est une hypothèse très importante pour construire l'estimateur de Kaplan Meier.

Nous rappelons que dans le cas discret :

$$\mu(t) = \frac{P(T=t)}{1-F(t^-)} = \frac{dF(t)}{1-F(t^-)},$$

En cas de censure (absence de troncature) : $\frac{dF(t)}{1-F(t^-)} = \frac{dH_1(t)}{1-H(t^-)},$

nous pouvons donc en déduire :

$$\hat{\mu}(t) = \frac{d\hat{H}_1(t)}{1 - \hat{H}(t)} = \frac{\#\{i: Y_i = t, \delta_i = 1\}}{\#\{i: Y_i \geq t\}},$$

Nous en déduisons, l'estimateur de Kaplan Meier qui est donné par :

$$\hat{F}(t) = 1 - \prod_{i: Y_i \leq t} \left(1 - \frac{\delta_i}{\sum_{j=1}^n 1_{Y_i \leq Y_j}} \right).$$

Cas général de l'estimateur de Kaplan Meier en cas de censure et de troncature :

$$\hat{F}(t) = 1 - \prod_{i: Y_i \leq t} \left(1 - \frac{\delta_i}{\sum_{j=1}^n 1_{\tau_i \leq Y_i \leq Y_j}} \right).$$

Nous remarquons qu'avec cette expression de la fonction de distribution empirique, l'estimateur de Kaplan Meier n'affecte du poids que sur les données non censurées mais garde en mémoire les données censurées avec des sauts de la fonction.

Application à nos données

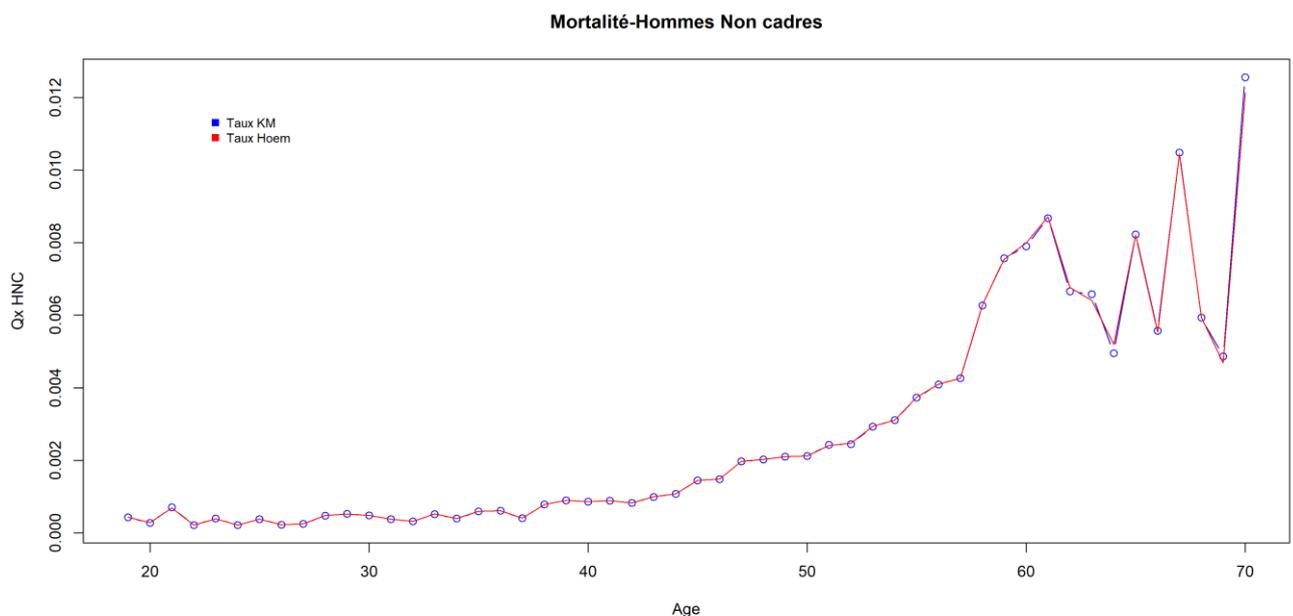


Figure 24: Estimation des taux bruts par Kaplan Meier

L'estimation de la mortalité des hommes non-cadres par la méthode de Kaplan Meier nous donne les taux présentés dans le graphique ci-dessus en bleu.

La comparaison de ces derniers avec ceux obtenus par la méthode de Hoem montre des résultats presque identiques à l'exception de quelques âges, sur lesquels les volumes de données sont faibles. Par la suite, nous ferons le choix de retenir les taux bruts estimés par la méthode de Kaplan Meier car elle est non paramétrique et moins coûteuse en termes d'hypothèses.

Cependant, afin d'obtenir une courbe de mortalité régulière et croissante, ces taux bruts sont lissés par la méthode de Whittaker-Henderson (méthode non paramétrique) et ajustés à une loi de

Makeham (méthode paramétrique). Un modèle adéquat sera ensuite choisi pour les hommes non-cadres après validation par des tests statistiques.

2.3.2.4 Lissage non paramétrique des taux bruts obtenus par Kaplan Meier : méthode de Whittaker-Henderson

Le principe de Whittaker-Henderson consiste à trouver une fonction $x \rightarrow \hat{q}_x$ telle que \hat{q}_x soit assez proche de la vraie courbe q_x tout en supposant qu'à priori cette courbe est lisse. q_x correspondant au taux obtenus par Kaplan Meier.

Cadre théorique et hypothèses

L'objectif étant de trouver le bon compromis entre un critère de régularité et un critère de fidélité pour avoir une courbe $q'_x(h)$ proche de celle d'origine en minimisant l'expression suivante :

$$WH_h(c) = F(c) + h S(c) .$$

- Le critère de fidélité est défini par :

$$F(c) = \sum_{i=1}^{x_{max}} w_x (c_x - \hat{q}_x)^2,$$

où w_x une séquence de poids positive et

- Le critère de régularité est défini par :

$$S(c) = \sum_{x=1}^{x_{max}-z} ((\Delta^z c)_x)^2,$$

avec $(\Delta^z c) = \Delta(\Delta^{z-1} c)$ et $(\Delta c)_x = c_{x+1} - c_x$ pour $x \leq x_{max} - 1$.

- h est un paramètre de lissage qui permet de modifier l'impact du lissage sur le critère global.

La solution au problème d'optimisation vérifie les conditions suivantes :

$$\frac{\partial WH_h}{\partial q_x} = 0 \text{ pour } 1 \leq x \leq x_{max} .$$

Ecriture Matricielle : Définissons les vecteurs suivants :

$$c = \begin{pmatrix} C_1 \\ \vdots \\ c_{max} \end{pmatrix}, \quad \hat{q} = \begin{pmatrix} \hat{q}_1 \\ \vdots \\ \hat{q}_{max} \end{pmatrix}, \quad W = \begin{pmatrix} w_1 & 0 & \dots & 0 \\ 0 & w_2 & \dots & 0 \\ \dots & 0 & \ddots & 0 \\ 0 & 0 & \dots & w_{x_{max}} \end{pmatrix}.$$

Nous pouvons écrire la fonction $F(c)$ sous la forme matricielle suivante :

$$F(c) = (c - \hat{q})^T W (c - \hat{q}).$$

L'opération $c \rightarrow \Delta^z c$ étant linéaire, nous pouvons le réécrire sous la forme d'une matrice K_z avec x_{max} lignes et $x_{max} - z$ colonnes telle que $\Delta^z c = K_z c$.

Ainsi : $S(c) = c^T K_z^T K_z c.$

On en déduit l'expression de la solution du problème d'optimisation de Whittaker-Henderson :

$$q'(h) = (W + hK_z^T)^{-1} W \hat{q}.$$

On remarque que pour une séquence de poids $w_i > 0$, $W + hK_z^T$ est inversible pour h petit.

Application à nos données

Nous allons faire plusieurs simulations de h et z et choisir celle qui représente le mieux la courbe obtenue par Kaplan Meier de façon visuelle mais aussi par des tests statistiques.

- Pour h=1 et z=1, nous obtenons la courbe suivante

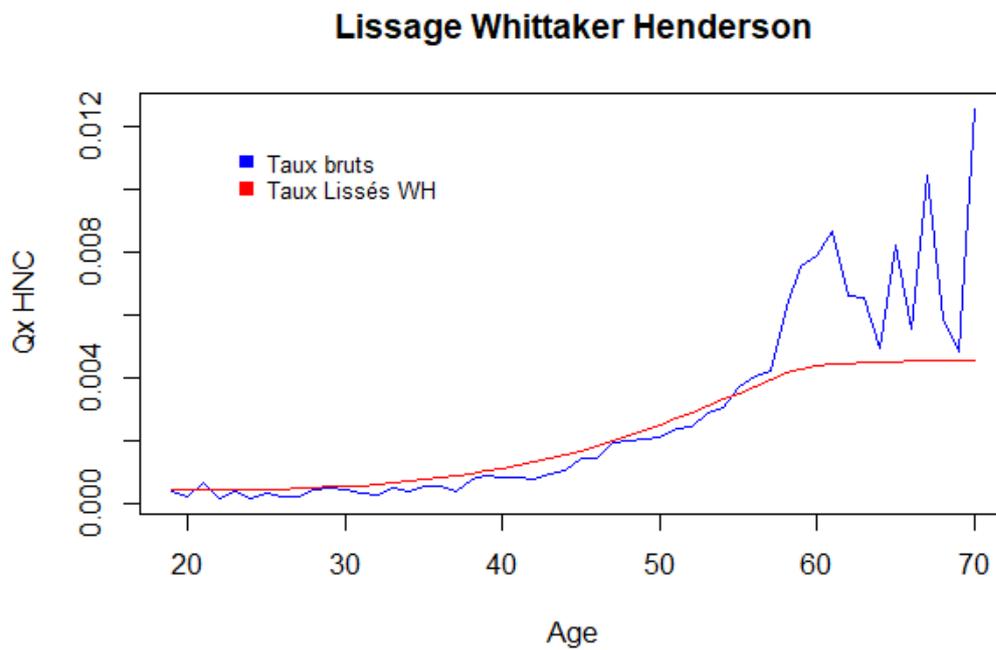


Figure 25: Lissage par Whittaker-Henderson (h=1, z=1)

La lecture de ce graphique nous permet de conclure que ces paramètres choisis pour h et z ne suffisent pas à obtenir un lissage adéquat à notre courbe.

- Pour $h=2$ et $z=1$, nous obtenons la courbe suivante

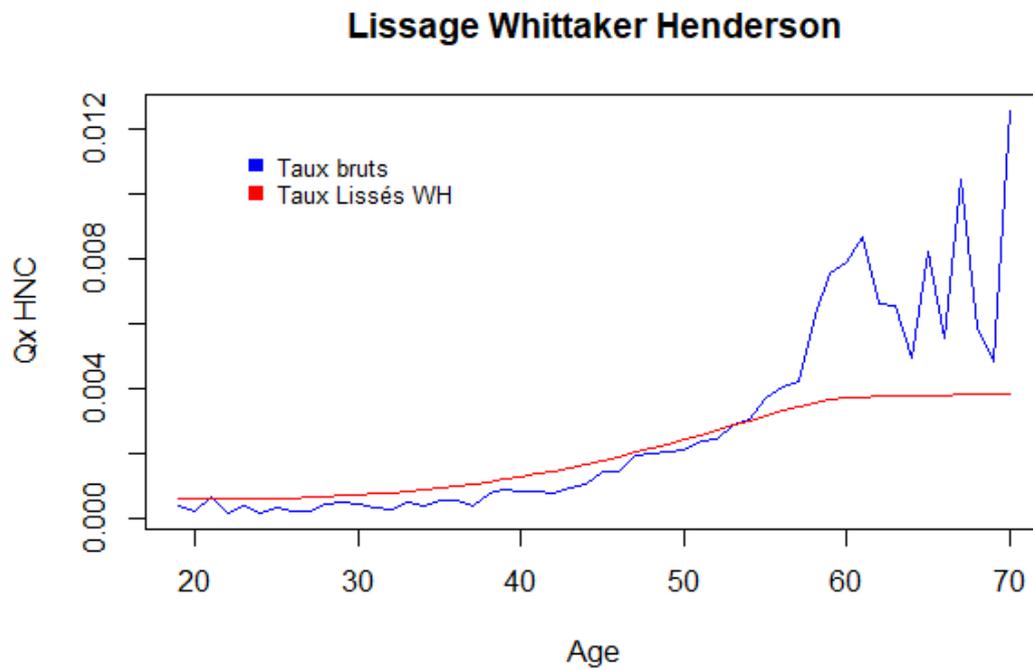


Figure 26: Lissage par Whittaker-Henderson ($h=2, z=1$)

Nous avons les mêmes conclusions que pour les paramètres précédemment choisis ($h=1$ et $z=1$). Les paramètres choisis pour $h=2$ et $z=1$ ne suffisent pas à obtenir un lissage adéquat à notre courbe.

- Pour $h=1$ et $z=2$, nous obtenons la courbe suivante

Lissage Whittaker Henderson

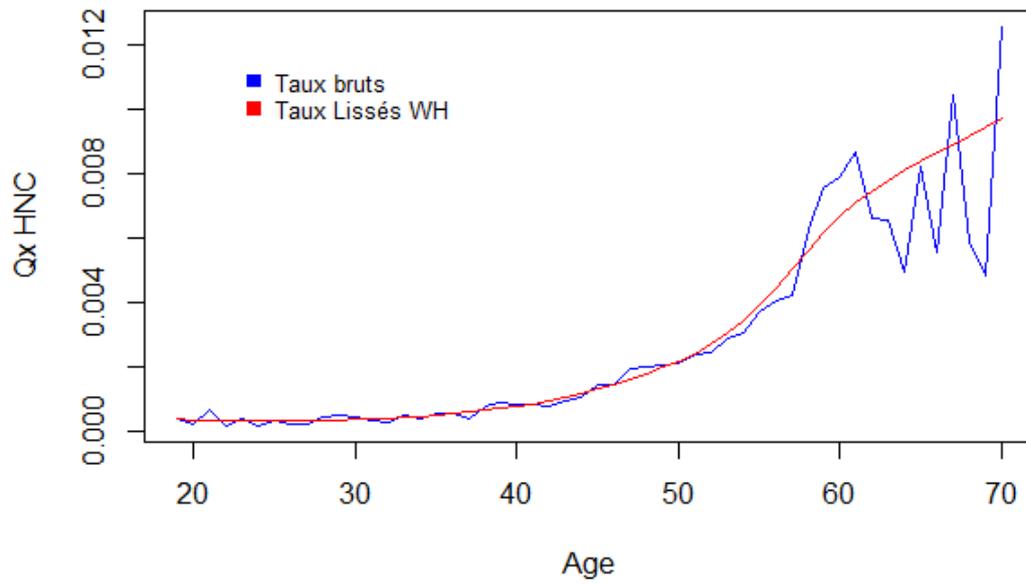


Figure 27: Lissage par Whittaker-Henderson ($h=1, z=2$)

- Pour $h=2$ et $z=2$, nous obtenons la courbe suivante

Lissage Whittaker Henderson

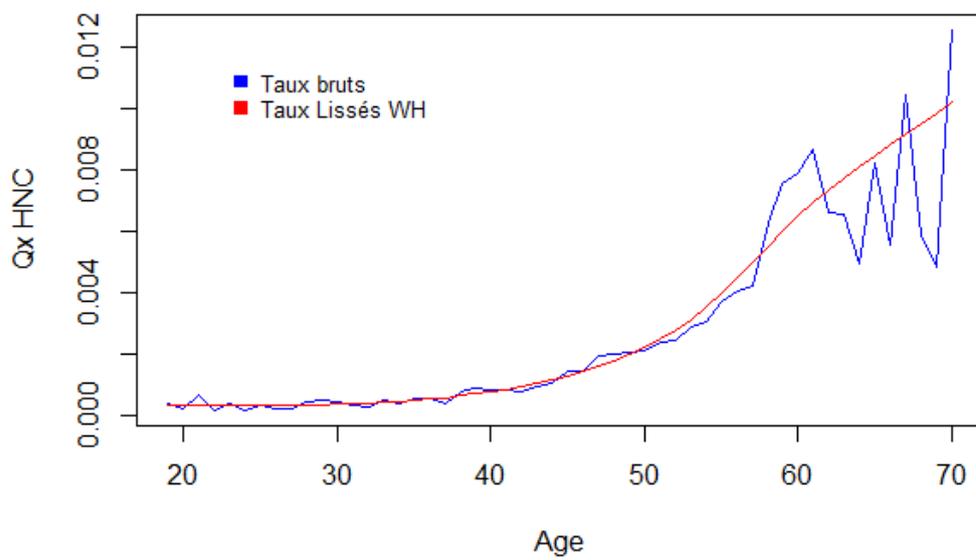


Figure 28: Lissage par Whittaker-Henderson ($h=2, z=2$)

- Pour $h=4$ et $z=2$, nous obtenons la courbe suivante

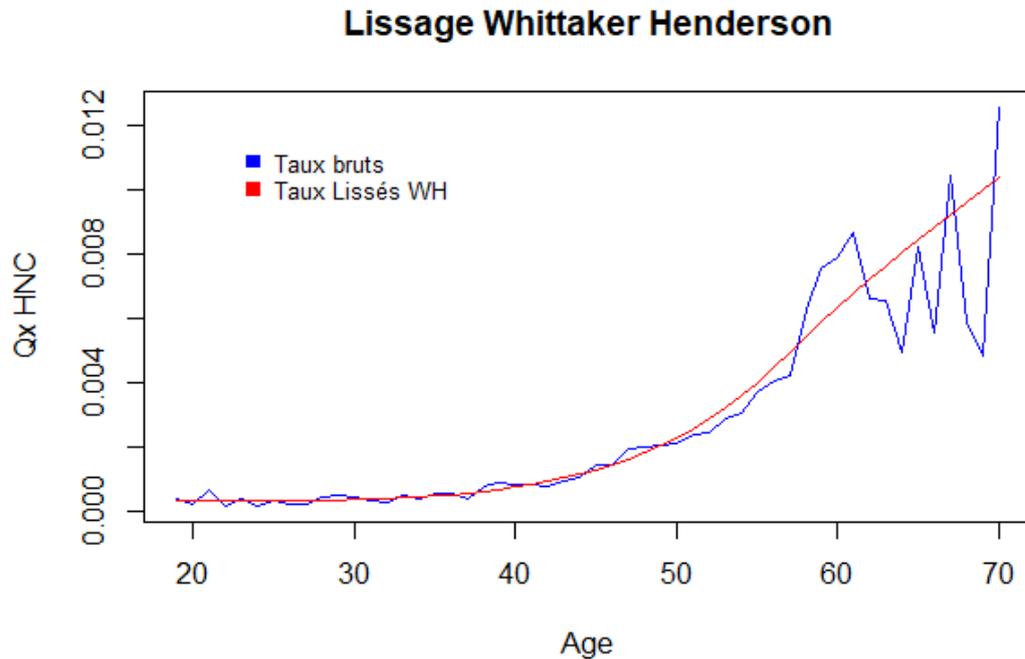


Figure 29: Lissage par Whittaker-Henderson ($h=4, z=2$)

Après plusieurs simulations des valeurs des paramètres h et z , nous décidons de retenir le lissage pour $h=1$ et $z=2$ car les taux lissés sont plus proches des taux bruts à tout âge et montrent à priori un bon lissage et un respect du critère de fidélité.

Cependant, nous allons valider ce lissage avec des tests statistiques tel que détaillé dans la section 2.3.2.6.

2.3.2.5 Lissage paramétrique des taux bruts obtenus par Kaplan Meier : Ajustement à une loi de Makeham

Le modèle de Makeham est basé sur l'hypothèse que le taux instantané de mortalité μ_x s'exprime sous la forme :

$$\mu_x = a + b \times c^x,$$

Où a, b et c sont des constantes vérifiant : $a > 0, b > 0, c > 1$.

Le terme $b \times c^x$ représente un effet associé au vieillissement faisant croître la mortalité de façon exponentielle et la constante a est interprétée comme une incidence accidentelle et doit être indépendante de l'âge.

Pour rappel, nous avons $\mu(x) = -\ln(S(x))'$,

ce qui implique que :

$$p_x = \exp \left[- \int_x^{x+1} \mu_y d_y \right],$$

$$p_x = \exp \left[- \int_x^{x+1} (a + b \times c^y) d_y \right],$$

$$p_x = \exp \left[-a - \frac{b}{\ln(c)} \times c^x \times (c - 1) \right],$$

$$q_x = 1 - \exp \left[-a - \frac{b}{\ln(c)} \times c^x \times (c - 1) \right].$$

Cadre théorique et hypothèses

Lorsqu'une courbe de mortalité suit une loi de Makeham, alors les points $(x, \ln|q_{x+1} - q_x|)$ sont alignés sur une droite de pente $\ln(c)$ avec q_x correspondant au taux obtenus par Kaplan Meier.

Il s'agit d'une propriété essentielle et doit être vérifiée pour justifier l'ajustement à cette loi.

En effet, d'après la définition du q_x énoncée précédemment, nous pouvons en déduire les expressions suivantes :

$$\ln(1 - q_x) = \ln(p_x) = -a - \frac{b}{\ln(c)} \times c^x \times (c - 1).$$

Pour q_x suffisamment petit, nous avons l'approximation suivante par un développement limité :

$$\ln(1 - q_x) \approx -q_x \rightarrow q_x \approx a + \frac{b}{\ln(c)} \times c^x \times (c - 1),$$

$$q_{x+1} - q_x = \frac{b}{\ln(c)} \times (c^{x+1} - c^x) \times (c - 1) = \frac{b}{\ln(c)} \times (c - 1)^2 \times c^x,$$

$$\ln|q_{x+1} - q_x| = \ln \left| \frac{b}{\ln(c)} \times (c - 1)^2 \right| + x \times \ln|c|,$$

En posant :

$$s = \exp(-a) \text{ et } g = \exp\left(-\frac{b}{\ln(c)}\right),$$

l'expression est simplifiée par la formule suivante :

$$\ln|q_{x+1} - q_x| = \ln \left| (c - 1)^2 \times \ln \left(\frac{1}{g} \right) \right| + x \times \ln|c|.$$

La démarche est d'abord de faire une régression linéaire sur la base de la relation établie ci-dessus.

Le modèle s'écrit alors :

$$Y = \beta_0 + \beta_1 \times X + \varepsilon ,$$

Avec :

$$Y = \ln|q_{x+1} - q_x|,$$

$$\beta_0 = \ln \left| (c - 1)^2 \times \ln \left(\frac{1}{g} \right) \right| ,$$

$$\beta_1 = \ln|c| ,$$

ε = les résidus du modèle de régression linéaire.

Après estimation des coefficients du modèle de régression linéaire simple et validation du modèle par des tests statistiques (test de significativité des coefficients de Fisher, analyse de la normalité des résidus...), celles des paramètres a , b et c en seront déduits.

Ces derniers serviront ensuite de jet initial dans un problème d'optimisation permettant d'obtenir l'estimation par maximum de vraisemblance des paramètres c , g et s (voir annexe 3).

Nous utiliserons l'algorithme de Nedler-Mead sur R pour résoudre ce problème d'optimisation (*Package Stats*).

Application à nos données

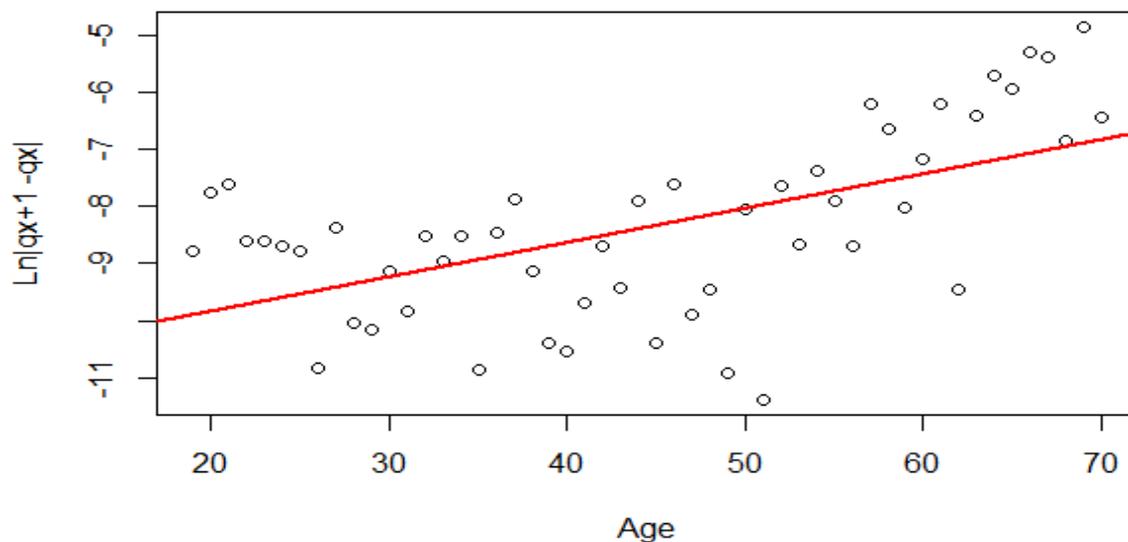


Figure 30: $\ln|q_{x+1} - q_x|$ en fonction de l'âge x

L'analyse du graphique nous montre une tendance linéaire et nous permet d'envisager l'utilisation du modèle de Makeham pour ajuster la courbe de mortalité à cette loi.

Par ailleurs, la modélisation de la régression à l'aide de la fonction « lm » du logiciel R (*Package Stats*) nous donne les estimations de coefficients ci-dessous :

Tableau 8: Régression linéaire Makeham

Coefficients	Estimation	Std Error	Valeur de la statistique de test	Probabilité critique p-value
β_0	-11.04698	0.57220	-19.306	<2e-16
β_1	0.06036	0.01218	4.954	8.7e-06

L'application du test de significativité des coefficients de régression de Fisher suivant :

$$H_0: \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \text{ contre } H_1: \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

donne des valeurs de p-value très faible permettant de rejeter l'hypothèse H_0 avec un risque d'erreur de 1% (p-value <0,01).

De plus, l'application du test de Shapiro-Wilk pour tester la normalité des résidus :

H_0 : les résidus suivent une loi normale contre ,

H_1 : les résidus ne suivent pas une loi normale .

donne une valeur de p-value permettant d'accepter l'hypothèse H_0 avec un risque d'erreur de 1% (p-value >0,01) .

Tableau 9: Test de normalité des résidus - régression Makeham

Données	Valeur de la statistique de test	Probabilité critique p-value
Résidus	0.97	0.16

Après estimation par maximum de vraisemblance, nous obtenons les paramètres optimaux suivants pour le modèle de Makeham :

Tableau 10 : Paramètres optimaux estimés - Modèle de Makeham

Paramètre	Valeur
c	1.104
g (pour rappel g doit être inférieur à 1 pour que l'algorithme converge)	0.999
s	1

Nous obtenons également la courbe d'ajustement de Makeham ci-après.

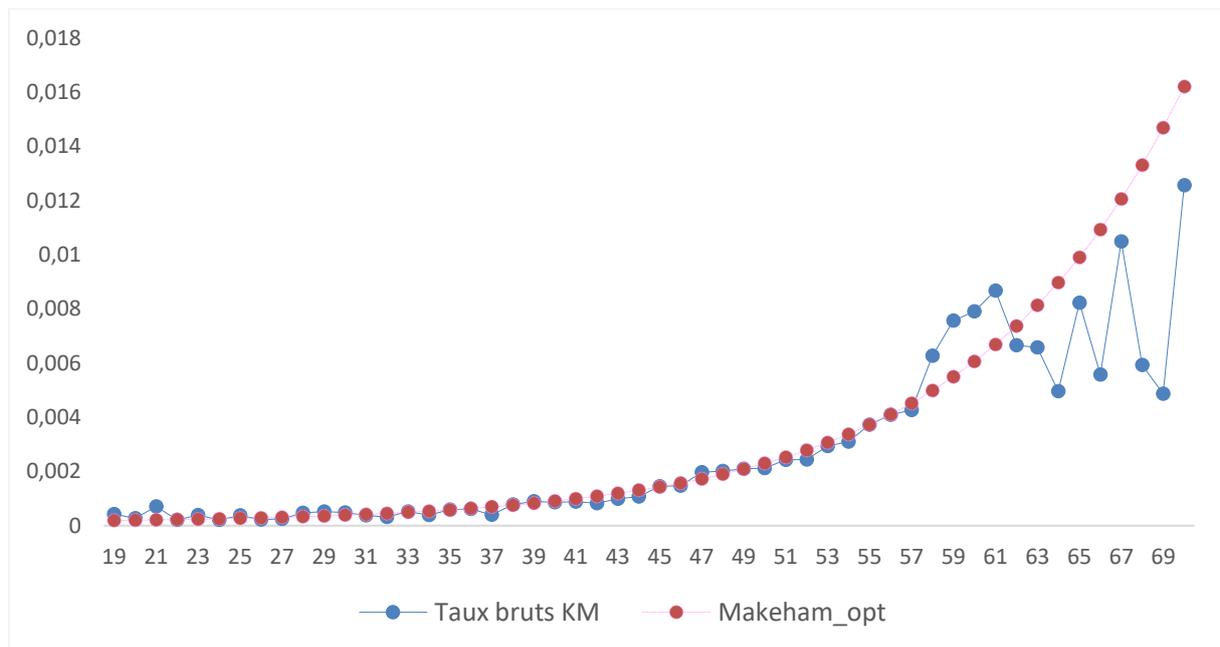


Figure 31: Ajustement des taux bruts au modèle de Makeham pour les hommes non-cadres

Nous constatons que la courbe est assez régulière, croissante et très proche de la courbe de Kaplan Meier. Cependant, nous observons une sous-estimation de la mortalité aux âges jeunes de 19 à 30 ans et une sur-estimation de la mortalité au-delà de 62 ans dû à la structure du modèle qui prend en compte une augmentation exponentielle de la mortalité en fonction de l'âge.

La validation statistique de cet ajustement à la section 2.3.2.6, nous permettra de confirmer son adéquation avec la courbe des taux bruts de Kaplan Meier.

2.3.2.6 Validation des lissages paramétriques et non paramétriques

2.3.2.6.1 Test de changement des signes

Cadre théorique et hypothèses

Le test du changement de signes est un test non paramétrique qui observe la fréquence des changements de signe de la différence entre les taux bruts de décès et les taux lissés.

Posons $d_x = q_x - \hat{q}_x$,

Le test s'écrit sous la forme suivante :

$$H_0 : P(d_x > 0) = P(d_x < 0) = \frac{1}{2} \text{ contre } H_1 : P(d_x > 0) \neq \frac{1}{2}.$$

Le non-rejet de l'hypothèse H_0 conduit à la validation du lissage des taux bruts.

Lorsque les âges sont indépendants, nous avons sous l'hypothèse H_0 :

$$P(d_x \text{ et } d_{x+1} \text{ de signes opposés}) = P(d_x > 0)P(d_{x+1} < 0) + P(d_x < 0)P(d_{x+1} > 0),$$

Ce qui implique que $P(d_x \text{ et } d_{x+1} \text{ de signes opposés}) = \frac{1}{2}$ et,

la variable aléatoire $1_{\{d_x \text{ et } d_{x+1} \text{ de signes opposés}\}}$ suit une loi de Bernoulli de paramètre $\frac{1}{2}$ sous l'hypothèse H_0 .

Par conséquent, la statistique suivante :

$$S_+^- = \sum_{i=1}^{N-1} 1_{\{(q_{x_i} - \hat{q}_{x_i}) \text{ et } (q_{x_{i+1}} - \hat{q}_{x_{i+1}}) \text{ de signes opposés}\}}$$
 suit une loi binomiale $\mathcal{B}(N-1; \frac{1}{2})$ sous H_0 .

En fonction de la taille N de notre plage de donnée, nous pouvons soit utiliser le test à distance fini soit faire une approximation gaussienne.

Lorsque N est petit, un test à distance finie est fait.

Nous cherchons dans ce cas à trouver pour un seuil α , le plus grand nombre k_α tel que pour une variable $X \sim \mathcal{B}(N-1; \frac{1}{2})$, $P_{H_0}(X < k_\alpha) < \frac{\alpha}{2}$.

{ Si $S_+^- \in]k_\alpha; N-1-k_\alpha[$, alors nous acceptons H_0 et validons le lissage ,
 { Si $S_+^- \notin]k_\alpha; N-1-k_\alpha[$, alors nous rejetons H_0 et nous cherchons un meilleur lissage .

Lorsque N est suffisamment important, alors nous pouvons faire l'approximation que S_+^- suit une loi normale $\mathcal{N}(\frac{(N-1)}{2}; \frac{(N-1)}{4})$ sous H_0 .

Nous en déduisons que la statistique

$$Z = \frac{(S_+^- - \frac{N-1}{2})}{\sqrt{\frac{N-1}{4}}} = \frac{2S_+^- - N + 1}{\sqrt{N-1}}, \text{ suit une loi normale centrée réduite sous } H_0.$$

La comparaison de la valeur absolue de cette statistique avec le quantile d'ordre $\frac{\alpha}{2}$ de la loi normale centrée réduite nous permet de rejeter ou pas l'hypothèse nulle H_0 avec α le risque d'erreur de première espèce du test (probabilité de rejeter H_0 à tort).

Application à nos données

Compte tenu de la taille de notre échantillon (52 lignes), nous allons choisir d'utiliser l'approximation gaussienne avec la statistique de test $Z = \frac{2S_+^- - N + 1}{\sqrt{N-1}}$ pour valider nos différents lissages.

Tableau 11: Résultats du test de changement des signes

Test de changement des signes		
	<i>Lissage par Whittaker-Henderson (h=1, z=2)</i>	<i>Ajustement à une loi de Makeham</i>
<i>Hypothèse H_0</i>	<i>$H_0 : P(d_x > 0) = P(d_x < 0) = \frac{1}{2} \rightarrow$ Validation du lissage si H_0 non rejetée</i>	
<i>Valeur absolue de la statistique de test Z</i>	<i>0,420</i>	<i>1,26</i>
<i>Loi de la statistique de test sous H_0</i>	<i>Loi normale centrée réduite</i>	
<i>Risque d'erreur de première espèce α</i>	<i>5%</i>	
<i>Quantile de la loi normale centrée réduite à $1 - \frac{\alpha}{2} = 97,5\%$</i>	<i>1,96</i>	
<i>H_0 est rejetée si $Z >$</i>	<i>1,96</i>	
<i>p-value</i>	<i>0,66</i>	<i>0,90</i>
<i>Conclusion :</i>	<i>La valeur de la statistique de test obtenue et celle de la p-value nous permettent d'accepter l'hypothèse nulle et donc de valider le lissage par Whittaker-Henderson et l'ajustement à une loi de Makeham.</i>	

2.3.2.6.2 Test SMR (Standard Mortality Ratio)

Cadre théorique et hypothèses

Le SMR est un ratio qui mesure l'écart de mortalité entre deux populations données en comparant leurs taux de mortalité. L'idée du test SMR est de comparer le nombre de décès observé et celui prédit par le modèle. Un $SMR > 1$ signifie qu'il existe un excédent de mortalité de la sous population par rapport à la mortalité de la population de référence, on parle de surmortalité. Lorsque le $SMR < 1$, on parle de sous-mortalité.

Soit :

- q_x le taux de mortalité à l'âge x ,
- N_x l'exposition à l'âge x ,
- O_x le nombre de décès observé à l'âge x et $O = \sum_{x=x_0}^{x_{max}} O_x$ le nombre total de décès observés,
- $E_x = N_x q_x$ le nombre de décès attendus à l'âge x et $E = \sum_{x=x_0}^{x_{max}} E_x$ le nombre total de décès attendus total.

- Le SMR est défini par le rapport suivant : $SMR = \frac{O}{E}$,

Le test SMR repose sur l'hypothèse nulle suivante :

H_0 : La distribution des décès observés (O) est identique dans la population de référence et dans la population étudiée (SMR=1). En d'autres termes SMR n'est pas significativement différente de 1

Sous l'hypothèse H_0 , le nombre de décès observés O suit une loi de poisson de paramètre E (nombre de décès estimés).

Lorsque le nombre de décès est faible, les bornes de l'intervalle de confiance à 95% du nombre de décès peuvent être observées avec les tables de la loi de Poisson.

Pour avoir une approximation précise du test de poisson, (Boice & Rothman, 1979) suggèrent d'utiliser la statistique suivante pour tester l'hypothèse nulle lorsque le nombre de décès observé est petit :

$$X = \sqrt{9 O^*} \left(1 - \frac{1}{9 O^*} - \left(\frac{E}{O^*} \right)^{\frac{1}{3}} \right),$$

Où $O^* = O$ si $O \geq E$ et $O^* = O + 1$ si $O < E$.

Sous l'hypothèse nulle, $X \sim \mathcal{N}(0 ; 1)$.

La comparaison de la valeur absolue de cette statistique avec le quantile d'ordre α de la loi normale centrée réduite nous permet de rejeter ou pas l'hypothèse nulle H_0 avec α le risque d'erreur de première espèce du test (probabilité de rejeter H_0 à tort).

Lorsque le nombre de décès est significatif, la loi de Poisson $\mathcal{P}(E)$ peut-être approximé par une loi normale $\mathcal{N}(E ; E)$ et par conséquent le SMR suit une loi normale $\mathcal{N}\left(1 ; \frac{1}{E}\right)$ sous H_0 .

Nous avons par déduction $(SMR - 1) \times \sqrt{E} \sim \mathcal{N}(0 ; 1)$.

Il est possible d'améliorer cette statistique en utilisant une approche conventionnelle par la loi du khi-deux corrigée par simple continuité (Breslow & Day, 1987), (Monson,1980) :

$$\left(|SMR - 1| - \frac{0,5}{E} \right)^2 \times E \sim \chi^2.$$

Compte tenu du nombre significatif de décès observés dans notre échantillon, nous faisons le choix d'utiliser l'approximation par la loi normale pour effectuer notre test d'hypothèses.

Application à nos données

Le test SMR appliqué à nos données nous donne les résultats suivants.

Tableau 12: Résultats du test SMR

Test SMR		
	<i>Lissage par Whittaker-Henderson</i>	<i>Ajustement par une loi de Makeham</i>
<i>Hypothèse H_0</i>	<i>SMR n'est pas significativement différent de 1</i>	
<i>Valeur du SMR</i>	<i>1,00</i>	<i>1,02</i>

Loi de la statistique de test sous H_0	Loi normale $\mathcal{N}\left(1; \frac{1}{E}\right)$	
Risque d'erreur de première espèce α	5%	
Quantile de la loi normale sous H_0 à $\frac{\alpha}{2}$ et $1 - \frac{\alpha}{2}$	0,964 et 1,036	0,964 et 1,036
H_0 est rejetée si SMR n'appartient pas à cet intervalle de confiance	[0,9638 ; 1,0362]	[0,9635 ; 1,0365]
Décision	H_0 non rejetée	H_0 non rejetée
Conclusion :	Le SMR n'est pas significativement différent de 1. Le lissage est bien ajusté aux données observées.	

2.3.2.6.3 Intervalles de confiance pour les taux bruts

Cadre théorique et hypothèses

En complément des tests SMR et de changement de signe, nous allons apprécier la précision du lissage par la détermination d'intervalle de confiance. La largeur de cet intervalle de confiance permet de mesurer le degré de précision de l'estimation.

Pour cela, nous allons utiliser une approche par approximation gaussienne car nous disposons d'assez de données par âge pour le faire (voir annexe 4).

En effet, la statistique suivante permet de faire le lien entre l'incertitude de l'estimation, le nombre d'observations et le niveau de confiance de l'intervalle souhaité :

$$\Delta_p = u_{\alpha/2} \sqrt{\frac{f \times (1 - f)}{N}}$$

Où f la valeur autour de laquelle est construit l'intervalle de confiance et u_p le quantile d'ordre p de la loi normale centrée réduite.

Ici $f = \hat{q}_x$: taux de mortalité estimé et lissé.

Ainsi, pour un degré de précision de l'ordre de 0,01 et un intervalle de confiance à 95%, nous devons

disposer pour chaque âge x de : $N_x = \left(\frac{1,96^2}{0,01^2}\right) \times \hat{q}_x \times (1 - \hat{q}_x)$ effectif sous risque. Ce qui est le cas.

D'après le théorème central limite, nous avons :

$$\sqrt{N_x} \times \frac{q_x - \hat{q}_x}{\sqrt{\hat{q}_x \times (1 - \hat{q}_x)}} \xrightarrow{N \rightarrow \infty} N(0,1).$$

On en déduit l'intervalle de confiance asymptotique de niveau α pour q_x donné par :

$$I_\alpha = \left[\hat{q}_x - u_{\frac{\alpha}{2}} \sqrt{\frac{\hat{q}_x \times (1 - \hat{q}_x)}{N_x}}, \hat{q}_x + u_{\alpha/2} \sqrt{\frac{\hat{q}_x \times (1 - \hat{q}_x)}{N_x}} \right].$$

Applications à nos données

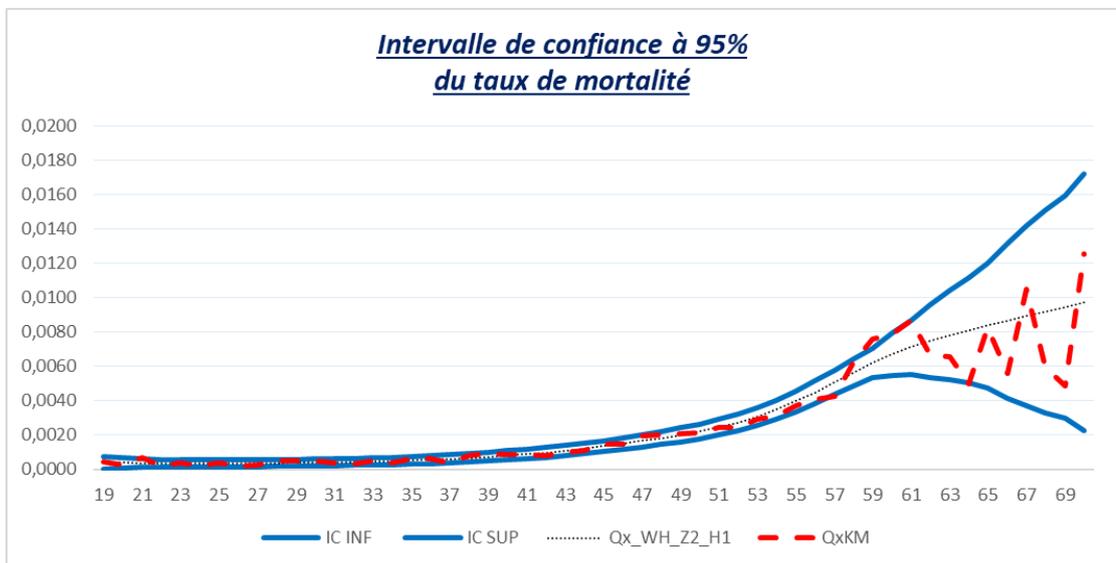


Figure 32: Intervalle de confiance pour Whittaker-Henderson

Nous remarquons que globalement tous les points de la courbe des taux bruts se situent à l'intérieur de l'intervalle de confiance sauf celui pour l'âge 59 ans mais il reste très proche de la borne supérieure de l'intervalle de confiance.

Nous pouvons donc conclure qu'après lissage des taux de mortalité par la loi Whittaker-Henderson, les taux bruts et ceux lissés sont assez proches.

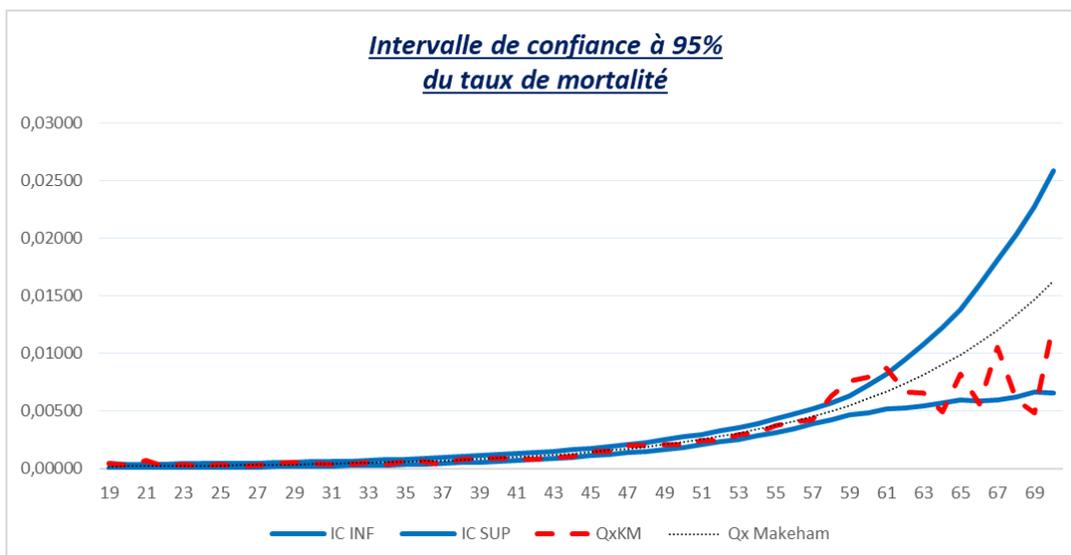


Figure 33: Intervalle de confiance pour Makeham

Pour l'ajustement à une loi de Makeham, nous constatons plus de points de la courbe des taux bruts en dehors de l'intervalle de confiance notamment sur les âges très jeunes et âges très élevés.

2.3.2.7 Choix de modèle pour la mortalité des hommes non-cadres

Les différents tests statistiques appliqués permettent de valider l'estimation des taux de mortalité par la méthode non paramétrique de Kaplan Meier qu'ils soient lissés par Whittaker-Henderson ou ajustés à la loi de Makeham.

Cependant, nous ferons le choix du modèle totalement non paramétrique (Kaplan Meier lissé par Whittaker-Henderson).

En effet, même si globalement le modèle de Makeham s'ajuste bien aux données, nous remarquons toutefois une sous-estimation de la mortalité sur les âges très jeunes et une forte sur estimation sur les âges au-delà de 50 ans du fait de la structure du modèle qui contient un paramètre exponentiel dépendant de l'âge.

Par ailleurs nous n'allons pas prolonger notre table au-delà de 70 ans car l'âge maximal d'adhésion au contrat collectif de décès correspond en général à l'âge de la retraite qui ne peut excéder 70 ans pour les salariés du secteur privé.

Pour les âges de 16 à 18 ans, nous préconisons d'appliquer le taux de mortalité estimé à l'âge de 19 ans, compte tenu du faible volume de données disponibles sur ces âges et ne permettant pas de faire une estimation fiable. De plus, il est très rare d'avoir des entreprises avec des âges moyens d'effectifs aussi jeunes.

2.3.2.8 Dédution de la mortalité des autres groupes : application des coefficients du modèle de Cox

Après l'estimation de la mortalité de notre groupe de référence (hommes non-cadres), nous en déduisons celle des autres groupes en utilisant la formule de passage du modèle de Cox et les paramètres estimés par celui-ci.

Rappel :

Nous avons pour chaque groupe i , avec $i = 2, 3, 4$:

$$q_x^i = 1 - \frac{S_i(x+1)}{S_i(x)},$$

$$q_x^i = 1 - \left(\frac{S_1(x+1)}{S_1(x)} \right)^{\exp(\theta_0^T Z^i)},$$

$$= 1 - (1 - q_x^1)^{\exp(\theta_0^T Z^i)}.$$

Avec q_x^1 la mortalité de notre groupe de référence qui est celle des hommes non-cadres et

$$Z_1^i = 1 \text{ si sexe= femme et } 0 \text{ sinon ,}$$

$$Z_2^i = 1 \text{ si CSP=cadre et } 0 \text{ sinon.}$$

Nous obtenons ainsi la formule suivante pour chaque âge et chaque groupe de risque :

Tableau 13: SMR par segment

Groupe de risque	Qx estimé par âge	SMR
Femmes non-cadres	$q_x^{\text{Femmes non cadres}} = 1 - (1 - q_x^1)^{\exp(-0,53462)}$	0,967
Femmes cadres	$q_x^{\text{Femmes cadres}} = 1 - (1 - q_x^1)^{\exp(-0,53462-0,86557)}$	1,15
Hommes cadres	$q_x^{\text{Hommes cadres}} = 1 - (1 - q_x^1)^{\exp(-0,86557)}$	0,935

Nous remarquons un SMR supérieur à 1 pour le groupe des femmes cadres.

Cependant, le faible volume de données (exposition et décès) sur ce périmètre semble ajouter beaucoup de volatilité à l'estimation des q_x .

De plus, la marge constatée sur les autres groupes de risques permet d'absorber cette éventuelle sous-estimation de la mortalité, compte tenu des faibles volumes d'expositions.

Nous décidons donc de les conserver.

Le graphique ci-après nous montre les courbes de mortalité estimés pour chaque âge.

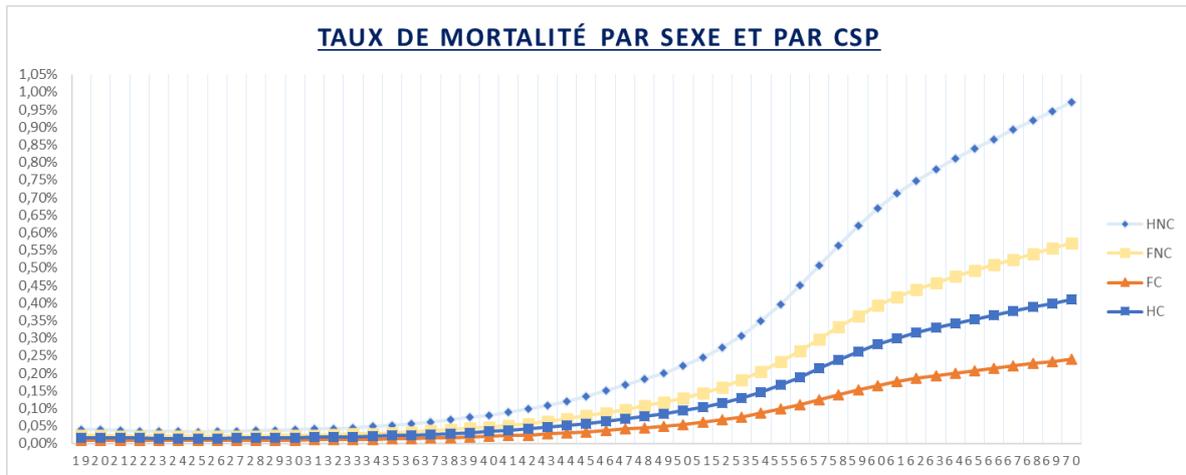


Figure 34: Taux de mortalité par sexe et par CSP retenus

Nous remarquons que le critère catégorie socio-professionnelle semble jouer plus sur la mortalité que le critère sexe. Le détail des q_x par âge est donné en annexe 5.

3 Impact de l'utilisation des tables de mortalité d'expérience

L'objectif de cette section est de challenger les tables de tarification actuelles en les comparant avec celles construites dans la section précédente mais aussi d'évaluer l'impact de l'utilisation de ces tables d'expérience segmentées par CSP dans le calcul du capital économique réglementaire requis pour le risque de catastrophe en vie dans le cadre de solvabilité 2.

3.1 Impact dans la tarification

3.1.1 Comparaison des tables d'expérience et des tables de tarification

La comparaison des tables de tarification actuelles (tables réglementaires ajustées) avec les tables d'expérience nous montre que les taux de mortalité estimés par celles-ci sont largement en dessous de ceux appliqués pour le tarif. Ils sont en moyenne inférieurs de 60% chez les non-cadres et 70% chez les cadres.

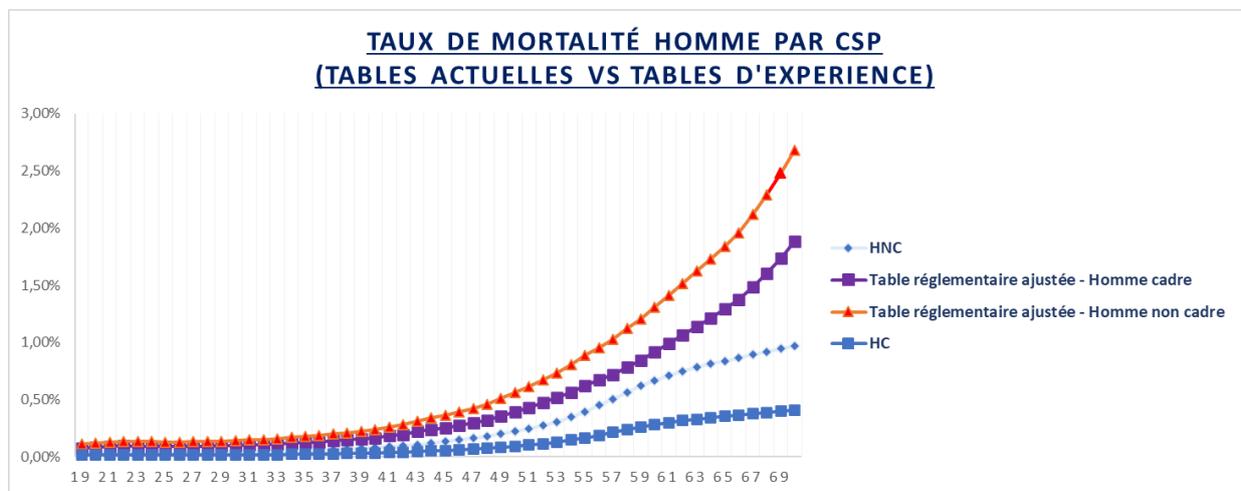


Figure 35: Comparaison tables d'expérience vs tables réglementaires (Hommes)

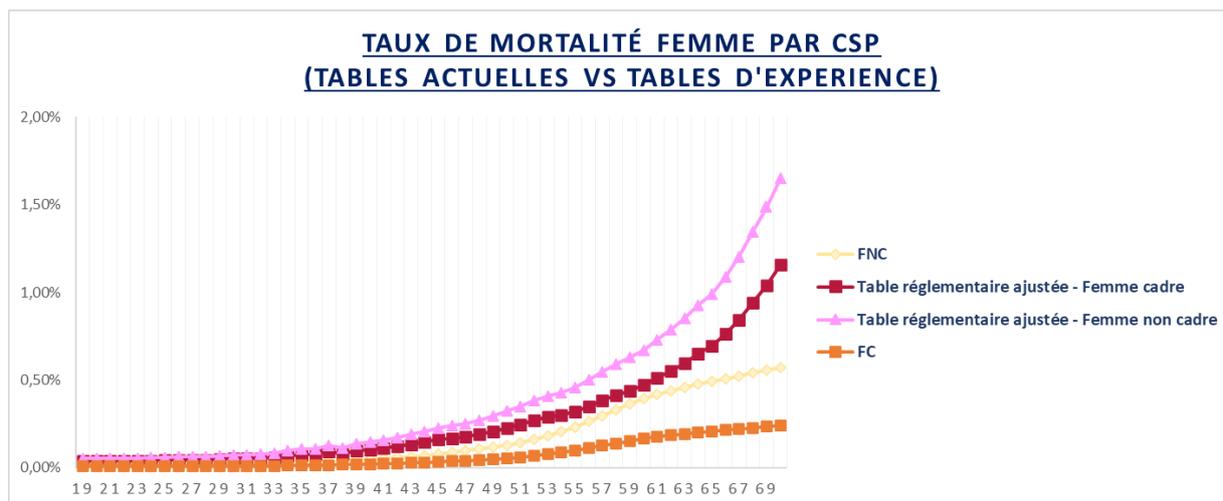


Figure 36: Comparaison tables d'expérience vs tables réglementaires (Femmes)

De plus, nous remarquons que la mortalité des tables réglementaires ajustées pour les cadres est au-dessus de celle des non-cadres obtenue avec les tables d'expérience à la fois chez les hommes et les femmes. Ce qui indique une forte sur estimation de la mortalité avec les tables actuelles.

Cependant, d'un point de vue opérationnel, des correctifs commerciaux sont appliqués aujourd'hui au tarif décès final pour tenir compte de la réalité du risque et de l'environnement concurrentiel.

L'utilisation de ces tables d'expérience devrait générer un gain opérationnel avec la suppression de ces correctifs (simplification des process) et permettra d'améliorer notre positionnement concurrentiel avec un tarif au plus proche du risque.

3.1.2 Optimisation du tarif décès à la maille groupe

Pour un individu d'âge x , la prime pure du décès en capital est donnée par la formule suivante :

$$prime = q_x * Capital\ garanti .$$

Le capital garanti correspond au montant des engagements à la charge de l'assureur en cas de décès et dépend de plusieurs paramètres dont le salaire, la situation familiale, le nombre d'enfants à charge, la catégorie socio-professionnelle, etc.

Comme évoqué au début de ce mémoire, pour les contrats collectifs, l'assureur ne dispose généralement pas des données tête par tête au moment de la souscription.

La tarification est réalisée sur des indicateurs moyens dont l'âge moyen du groupe, le salaire moyen, le nombre d'enfants moyen etc. Cependant, pour éviter de biaiser la tarification, il est important de prendre en compte la répartition de ces paramètres notamment les âges des effectifs et les capitaux garantis associés. Pour cela, nous allons appliquer le correctif âge actuariel.

L'âge actuariel est l'âge moyen x pour lequel la prime moyenne du groupe assuré est la plus proche du « juste prix » des risques engagés. Il est différent de l'âge moyen arithmétique.

Prenons l'exemple du tableau ci-dessous :

Tableau 14: Exemple de tarification avec prise en compte de l'âge actuariel

	<i>Groupe 1</i>			<i>Groupe 2</i>			<i>Groupe 3</i>		
	<i>Ind1</i>	<i>Ind2</i>	<i>Ind3</i>	<i>Ind1</i>	<i>Ind2</i>	<i>Ind3</i>	<i>Ind1</i>	<i>Ind2</i>	<i>Ind3</i>
Age	25	25	25	20	20	35	19	26	31
q_x	0,031%	0,031%	0,031%	0,026%	0,026%	0,057%	0,025%	0,033%	0,043%
Capital garanti en K€	63	63	63	32	32	144	28	72	114
Prime individuelle	19,73	19,73	19,73	8,21	8,21	81,67	7,19	23,31	48,60
Age moyen du groupe	25			25			25		
Prime moyenne du groupe sans prise en compte des spécificités individuelles	19,73			19,73			19,73		
Prime moyenne du groupe avec prise en compte des spécificités individuelles	19,73			32,70			26,37		
Age pour lequel la prime est la plus proche de la prime moyenne du groupe avec prise en compte des spécificités individuelles (ou Age actuariel)	25			29			28		

Nous avons 3 groupes d'assurés avec un âge moyen arithmétique de 25 ans. Cependant, nous constatons que la prime n'est pas la même en prenant en compte les spécificités de chaque âge individuel (q_x , capitaux garantis) dans chaque groupe.

En effet, la prime moyenne du groupe ajustée des effets individuels est plus proche de celle d'un groupe d'âge moyen 29 ans pour le groupe 2 et 28 ans pour le groupe 3 au lieu de 25 ans. L'âge actuariel pour le groupe 2 est alors de 29 ans et celui du groupe 3 de 28 ans.

Pour déterminer les correctifs « âge actuariel » à appliquer dans la tarification, nous allons nous appuyer sur l'expérience du portefeuille.

La méthode consiste, dans un premier temps, à construire un tableau de répartition des effectifs par âge, pour chaque âge moyen x , à partir des données des entreprises assurées. Celui-ci est complété par les capitaux moyens garantis par âge, sexe et catégorie socio-professionnelle afin d'avoir un profil moyen d'entreprise pour chaque âge moyen x .

Le tableau ci-après illustre un exemple de tableau de répartition des effectifs pour des âges moyens allant de 40 à 45 ans :

Tableau 15: Exemple de répartition des effectifs autour de la moyenne

Age des effectifs	Age moyen des groupes						Capital garanti en K€
	40	41	42	43	44	45	
20	4%	0%	0%	9%	5%	1%	32
21	3%	5%	0%	0%	4%	1%	39
22	2%	0%	2%	0%	3%	1%	48
23	0%	0%	2%	1%	2%	2%	60
24	0%	0%	1%	2%	0%	2%	73
25	0%	0%	0%	1%	0%	2%	83
26	0%	0%	0%	0%	0%	5%	89
27	0%	0%	0%	0%	0%	0%	98
28	0%	0%	0%	0%	0%	0%	107
29	0%	0%	0%	0%	0%	0%	116
30	2%	0%	4%	1%	5%	0%	125
31	0%	0%	1%	0%	0%	0%	131
32	0%	0%	1%	0%	0%	0%	143
33	0%	5%	0%	0%	0%	0%	154
34	0%	0%	1%	0%	0%	0%	156
35	2%	6%	3%	0%	0%	10%	168
36	2%	1%	8%	2%	3%	0%	177
37	1%	1%	0%	0%	1%	0%	193
38	0%	1%	0%	0%	5%	0%	206
39	3%	3%	1%	1%	0%	0%	209
40	10%	5%	5%	5%	0%	0%	212
41	13%	8%	7%	5%	0%	0%	220
42	10%	12%	10%	5%	0%	0%	224
43	12%	15%	4%	12%	0%	0%	228
44	20%	12%	4%	3%	15%	9%	242
45	9%	12%	10%	5%	5%	12%	245
46	7%	8%	6%	6%	5%	0%	247
47	0%	6%	10%	10%	0%	0%	255
48	0%	0%	10%	10%	0%	0%	263
49	0%	0%	5%	5%	0%	0%	270
50	0%	0%	2%	2%	20%	9%	278
51	0%	0%	4%	4%	0%	9%	286
52	0%	0%	0%	5%	0%	9%	294
53	0%	0%	0%	5%	8%	6%	302
54	0%	0%	0%	2%	8%	4%	309
55	0%	0%	0%	0%	8%	7%	317
56	0%	0%	0%	0%	0%	7%	325
57	0%	0%	0%	0%	0%	0%	333
58	0%	0%	0%	0%	3%	2%	341
59	0%	0%	1%	0%	0%	1%	348
60	0%	0%	0%	0%	0%	1%	356

61	0%	0%	0%	1%	0%	0%	364
Total	100%	100%	100%	100%	100%	100%	

Dans un second temps, pour chaque âge moyen x , la prime est calculée en faisant la moyenne pondérée des coûts individuels considérant les particularités (capitaux garantis par sexe et CSP).

La prise en compte de ces individualités conduit à augmenter le tarif de façon plus ou moins importante selon les âges moyens x (moins sur les âges jeunes et élevés et en moyenne + 46% chez les hommes non-cadres, +32% chez les femmes non-cadres, + 62% chez les hommes cadres, +46% chez les femmes cadres).

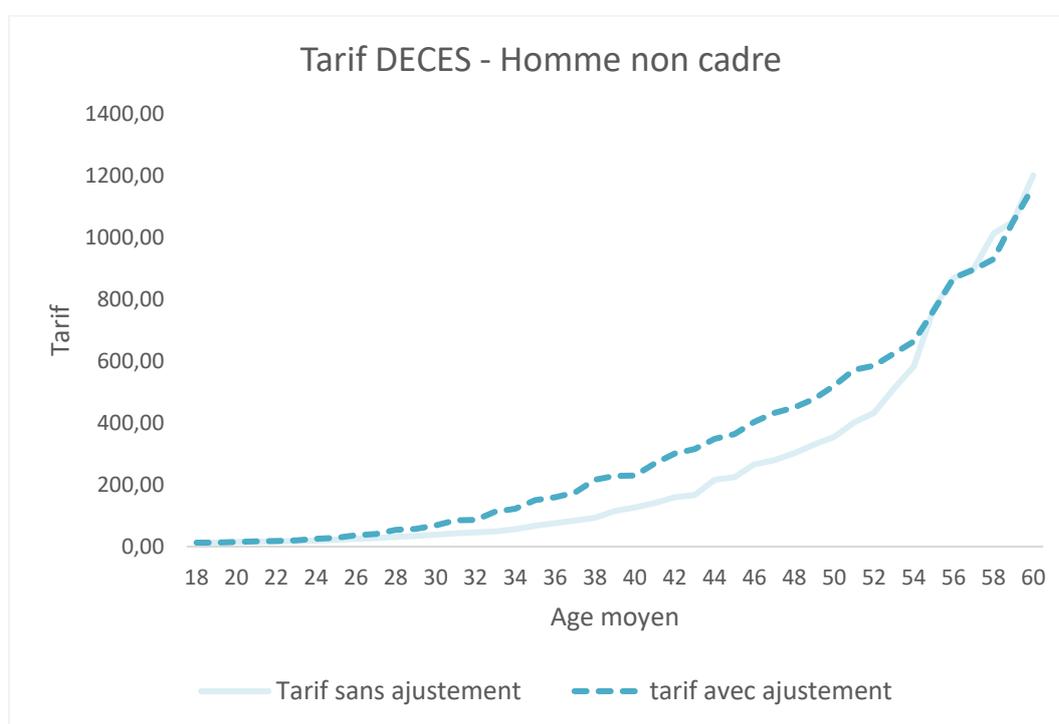


Figure 37: Tarif du décès avec et sans ajustement « âge actuariel » pour les hommes non-cadres

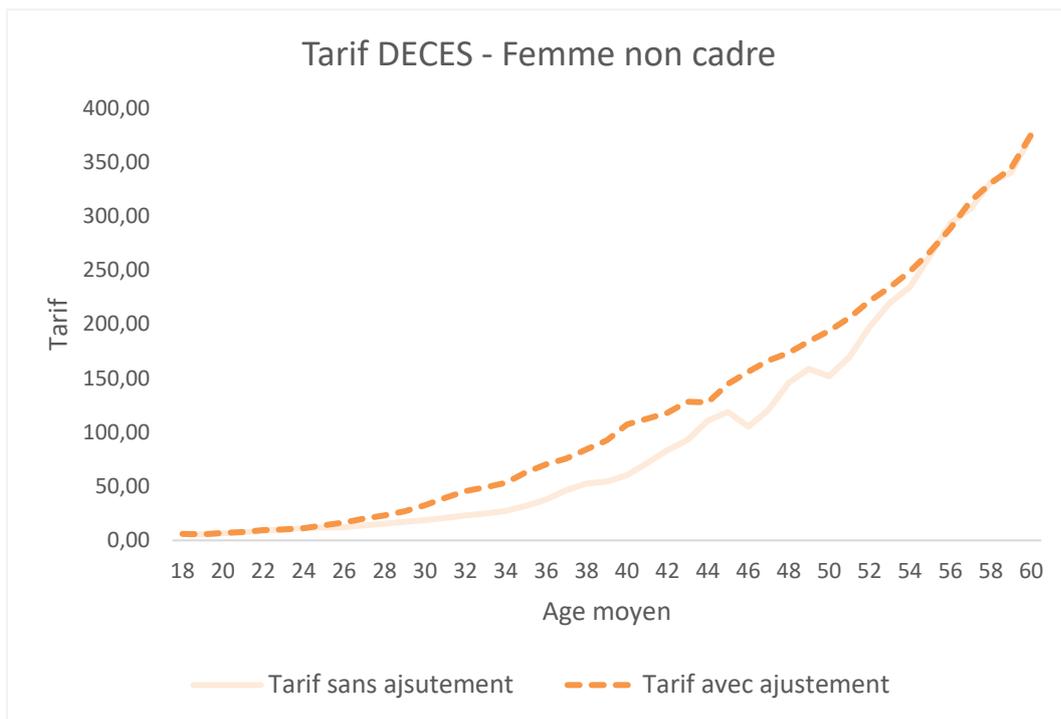


Figure 38: Tarif du décès avec et sans ajustement « âge actuariel » pour les femmes non-cadres

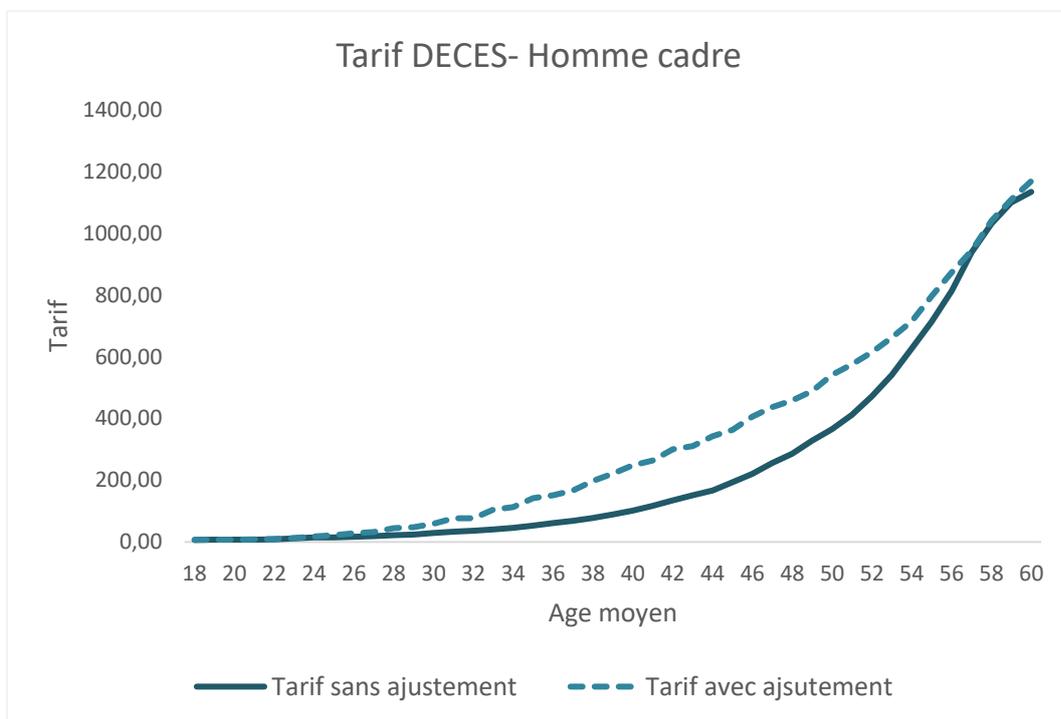


Figure 39: Tarif du décès avec et sans ajustement « âge actuariel » pour les hommes cadres

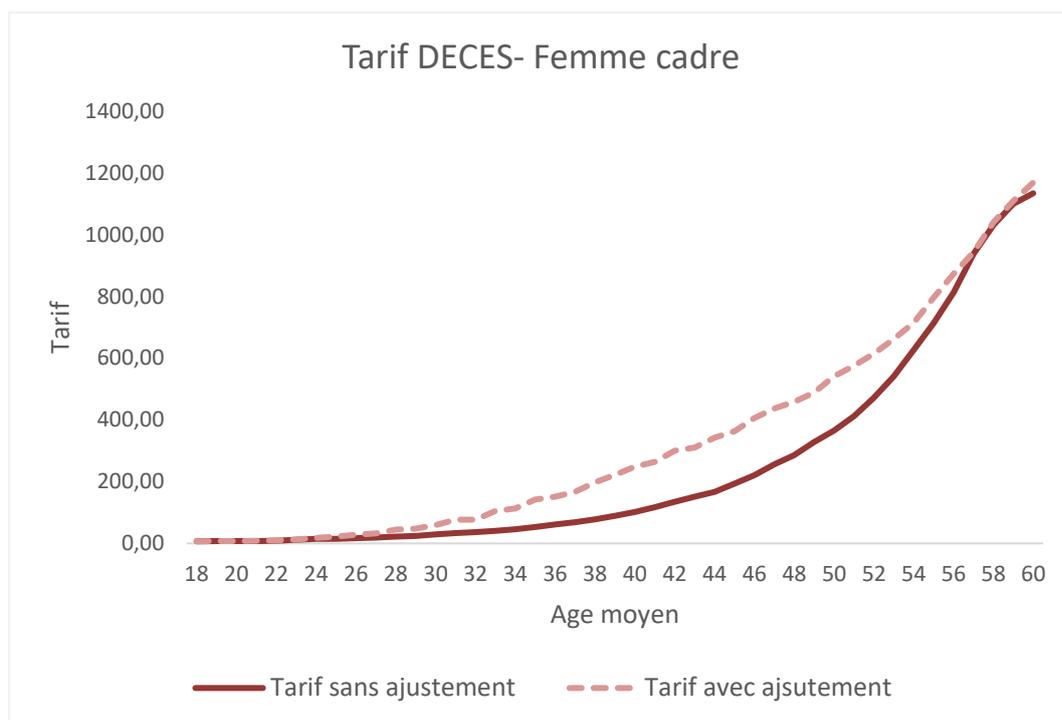


Figure 40: Tarif du décès avec et sans ajustement « âge actuariel » pour les femmes cadres

Enfin, l'âge actuariel est obtenu en recherchant l'âge pour lequel la prime est la plus proche de la prime moyenne groupe calculée.

Autrement dit, on recherche l'âge y pour lequel :

$$q_y * Capital\ garanti_y \approx \text{moyenne pondérée des coûts individuels.}$$

A l'aide de fonctions de recherche sur Excel, nous avons les correctifs « âge actuariel » moyens suivants, pour les cadres et les non-cadres :

Tableau 16: Correctif « âge actuariel » par sexe et CSP

Age	Correctif âge actuariel moyen non-cadre	
	H	F
23 ans et moins	0	0
24 à 27 ans	3	2
28 à 32 ans	5	5
33 à 40 ans	6	5
41 à 43 ans	6	4
44 à 46 ans	5	3
47 à 50 ans	4	3
51 à 54 ans	2	1
55 ans et plus	0	0

Age	Correctif âge actuariel moyen cadre	
	H	F
23 ans et moins	0	0
24 à 27 ans	3	2
28 à 32 ans	6	5
33 à 40 ans	7	6
41 à 43 ans	6	6
44 à 46 ans	5	5
47 à 50 ans	4	3
51 à 54 ans	2	1
55 ans et plus	0	0

En cas de nouvelle souscription, une entreprise d'âge moyen x aura pour q_x celui de l'âge moyen $x + \text{correctif âge actuariel}$ (décalage d'âge).

Exemple : une entreprise composée à 100% d'hommes non-cadres d'âge moyen 40 ans se fera appliquer le q_x de 46 ans pour la tarification.

L'application de ces correctifs « âge actuariel » permet d'avoir une vision « Best estimate » du tarif pour le groupe d'assuré. Cependant, leur détermination est très dépendante de la structure du portefeuille. Il serait donc utile de les actualiser régulièrement.

3.2 Impact dans le calibrage du choc catastrophe en vie

3.2.1 Rappel de la directive Solvabilité 2

Solvabilité 2 est une réforme réglementaire mise en place en 2016 par la commission européenne, qui vise à réguler le secteur de l'assurance avec une harmonisation du cadre de surveillance prudentielle au sein de la communauté européenne pour une protection du système financier et la prévention du risque systémique.

Elle repose sur trois piliers fondamentaux qui définissent des règles d'exigences quantitatives, qualitatives (management des risques et contrôle interne) et d'information vis-à-vis des autorités de contrôle mais aussi des assurés tel qu'illustré dans le schéma ci-après :



Figure 41: Les piliers de solvabilité 2

Pour la suite, nous allons nous intéresser uniquement au pilier 1 de solvabilité 2 sur le calcul des exigences quantitatives et ne donnerons pas plus de détails sur le fonctionnement des piliers 2 et 3.

3.2.2 Le pilier 1 de Solvabilité 2

Il définit les exigences quantitatives de la réforme. Les sociétés d'assurance ou de réassurance doivent disposer de suffisamment de capitaux pour honorer leurs engagements avec une probabilité de 99,5% à horizon 1 an : elles doivent être en mesure d'absorber les pertes résultant d'un choc bicentenaire, c'est-à-dire un risque se produisant statistiquement au moins tous les 200 ans, à horizon 1 an.

La quantification des capitaux propres est faite à partir du calcul des composantes du bilan prudentiel, qui inclut des ajustements par rapport au bilan comptable tels que la valorisation des actifs en valeur de marché, le calcul des provisions techniques en vision « *Best Estimate* » (*autrement appelées Best Estimate Liabilities (BEL)*), la marge de risque (ou coût du capital) et l'intégration des impôts différés.

La directive impose deux montants minimums de fonds propres :

- **Le solvency Capital requirement (SCR)** correspondant au niveau de capital nécessaire pour absorber une perte bicentenaire à un horizon d'un an, c'est à dire pour que la probabilité de ruine soit inférieure à 0,5%. Le STEC est l'équivalent du SCR dans le modèle interne d'AXA France.
- **Le Minimum Capital requirement (MCR)** correspondant au niveau minimum de fonds propres, pour pouvoir continuer à exercer l'activité d'assurance. En dessous de ce niveau de capital, l'Autorité de Contrôle Prudentiel et de Résolution (ACPR) retire l'agrément administratif autorisant l'exercice des opérations d'assurance. C'est l'exigence de capital cible correspondant à une probabilité de faillite de 15%.

L'ACPR utilise le SCR comme principal indicateur de solvabilité des compagnies d'assurance.

Pour le calcul des exigences quantitatives, le choix peut être fait entre la formule standard basée sur le profil de risque d'un assureur européen moyen, ou un modèle interne adapté aux spécificités propres de la compagnie d'assurance. La mise en place d'un modèle interne nécessite des développements spécifiques permettant d'adapter les calculs de besoin en capital pour une meilleure optimisation. Il doit être validé par l'autorité de contrôle avant d'être adopté.

3.2.2.1 La Formule standard

La Formule Standard prend en considération les risques majeurs auxquels, en général, les compagnies d'assurance sont exposées comme le risque de marché, de souscription vie et non-vie, de santé et de crédit. Le calcul du SCR selon la formule standard est basé sur une approche modulaire et repose sur une méthode d'agrégation des risques par matrice de corrélation dont les coefficients sont définis par l'autorité de contrôle. Cela permet de tenir en compte l'effet de diversification entre les risques.

En effet, l'hypothèse sous-jacente de cette diversification matérialise le fait que tous les risques ne peuvent se produire en même temps. Le graphe suivant détaille la structure du SCR dans la formule standard :

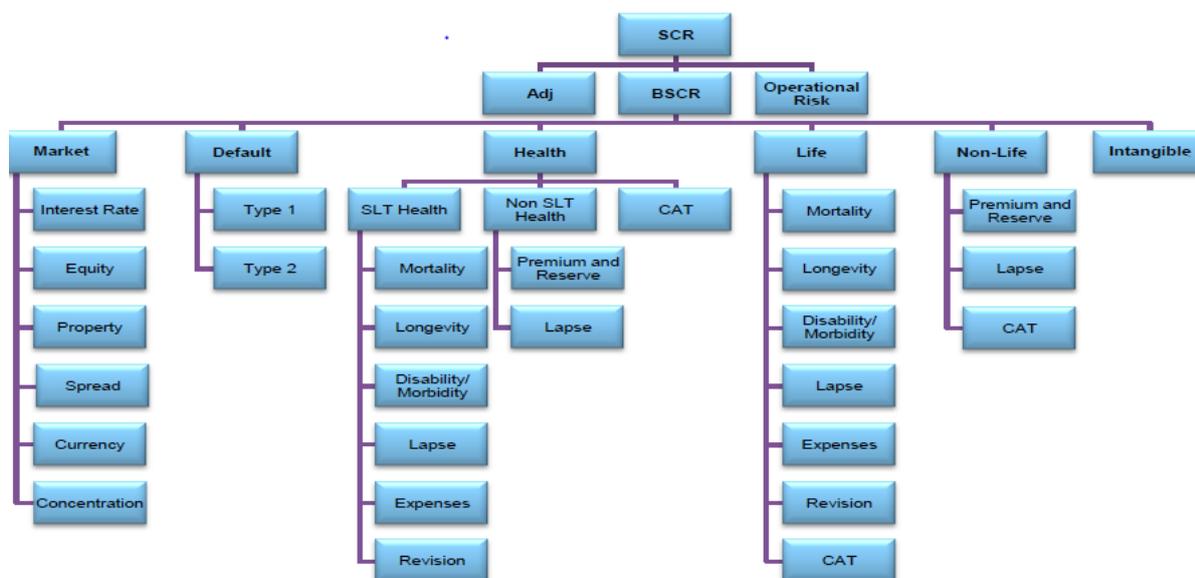


Figure 42: Structure du SCR par la formule standard

Ainsi, le montant du SCR total est déterminé par la somme du SCR de base, du SCR opérationnel, et des ajustements pour impôts différés et capacité d'absorption des pertes par les provisions techniques :

$$SCR = BSCR + SCR_{\text{opérationnel}} - \text{Ajustement} .$$

Où le SCR opérationnel correspond au capital à immobiliser pour couvrir les risques opérationnels. Le risque opérationnel est défini comme le risque de perte qui résulte d'une inadéquation ou d'une défaillance d'un processus, d'une personne ou d'un système, ou d'un événement externe.

3.2.2.2 Le modèle interne d'AXA France

Compte tenu de la diversité de son activité (vie, non-vie, gestion d'actifs...), de sa taille et de sa présence géographique à l'international, le Groupe AXA dispose d'un modèle interne adapté à ses spécificités et validé par l'Autorité de contrôle prudentiel et de résolution (ACPR).

En particulier, nous allons nous intéresser au modèle interne de l'entité AXA France. En effet, dans le cadre de Solvabilité 2, le groupe d'assurance doit être en mesure de prouver la solvabilité de ses entités prises individuellement mais aussi collectivement.

Le SCR sera remplacée par le terme STEC (*Short Term economic capital*) dans le modèle interne d'AXA France. Celui-ci repose sur une structure de cinq modules du risque de base (BSCR) et des briques de sous-risques adaptés à ses activités : par exemple pour le risque de Marché, nous notons la prise en compte des risques relatifs aux *hedge funds*, au *private equity*, à la volatilité et à l'inflation en lieu et place du risque de concentration.

Le schéma suivant illustre la structure du STEC dans le modèle interne d'AXA France :

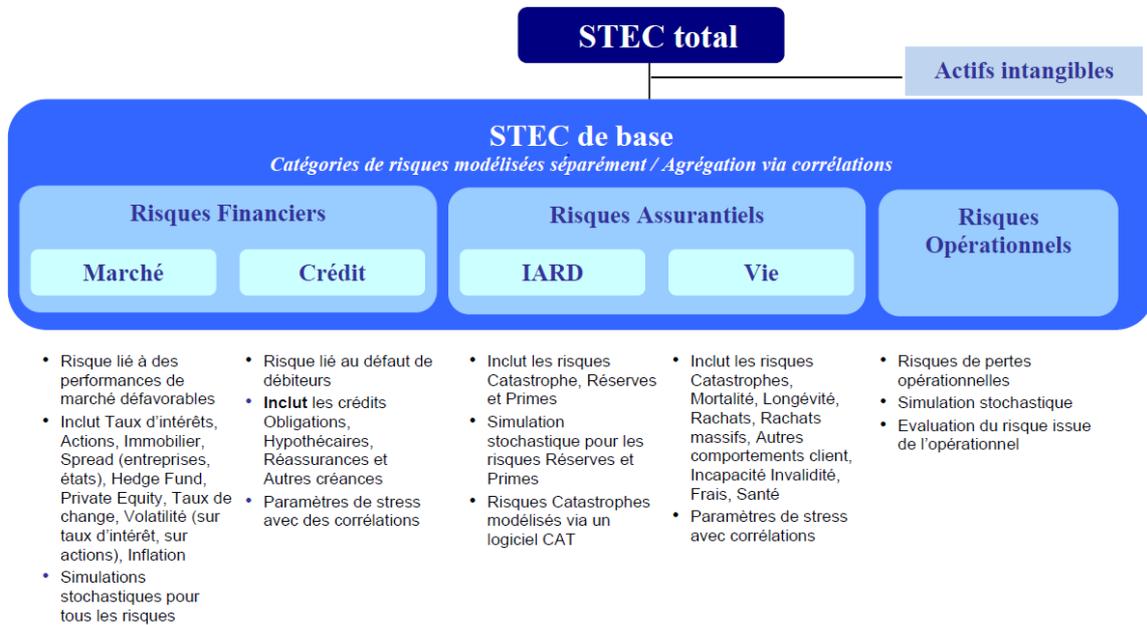


Figure 43: Structure du STEC dans le modèle interne d'AXA France

Tous les types de risques individuels sont agrégés via une matrice de corrélation pour tenir compte de l'effet de diversification selon le principe de base que tous les événements ne se produisent pas en même temps : par exemple, la probabilité qu'un événement catastrophique bicentenaire se produise en même temps qu'un événement financier bicentenaire est beaucoup plus faible que 1/200.

Pour chaque type de risque, les chocs bicentennaires appliqués pour le calcul du STEC sont définis à partir des modèles statistiques, des données de marché, d'événements extrêmes observés et de jugement d'expert.

Le STEC est la variation de l'EOF (*Eligible Own Funds*) à la suite de l'application du choc (stress) correspondant. Le schéma suivant illustre son calcul :

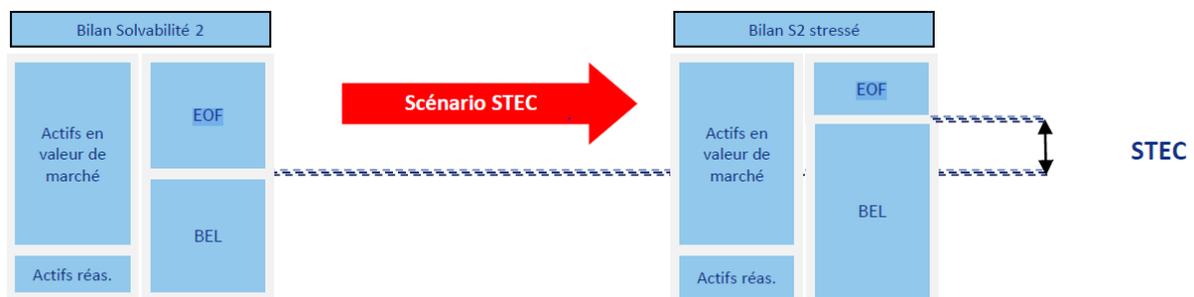


Figure 44: Illustration du calcul du STEC

EOF (*Eligible Own Funds*) représente le montant du capital immédiatement disponible pour honorer les engagements envers les assurés dans un contexte défavorable. Il correspond à la différence entre l'actif et le passif en valeur de marché.

BEL (*Best estimate liabilities*) correspond au montant des provisions en vision économique.

Dans la suite, nous allons nous intéresser au calibrage du choc et au calcul du STEC pour le module vie, en particulier le risque Catastrophe en vie sur le périmètre de la Prévoyance collective.

3.2.3 Le risque Catastrophe en vie (CAT)

Le risque catastrophe en vie correspond à un événement extrême qui fait augmenter significativement la mortalité, les frais médicaux et le risque d'invalidité. Dans le modèle interne d'AXA France, ce risque est basé sur 3 scénarios : la pandémie, le terrorisme et les tremblements de terre. Le risque pandémie étant lui-même segmenté en deux composantes : le risque « *CAT mortality* » et le risque « *CAT disability et medical expenses* ».

3.2.3.1 Calibrage du choc CAT Mortality

Le choc appliqué au risque « *CAT mortality* » correspond à une hausse additionnelle de la mortalité sur la première année de projection appliquée sur les contrats vie (santé, prévoyance, épargne). Une étude interne indépendante de ce mémoire a permis d'estimer ce taux à 0,08% (hausse de q_x).

Cependant, sur le périmètre de la prévoyance collective, le modèle de projection fonctionne sous forme de *S/C* (ou *Loss ratio*) et la hausse additionnelle de 0,08% de mortalité doit être converti en impact *S/C*.

Ainsi, la formule retenue pour traduire ce choc est :

$$c\% = \left(\frac{q_x + \text{choc}}{q_x} \right) - 1,$$

q_x est le taux de mortalité moyen best estimate calibré sur le portefeuille.

Une hausse additive de 0,08% des taux de mortalité correspond à une augmentation de $c\%$ du loss ratio. Le *S/C* choqué correspond au $S/C_{de\ base} \times (1 + c\%)$.

Exemple : pour $q_x = 0,20\%$, nous avons $c\% = \left(\frac{0,20\% + 0,08\%}{0,20\%} \right) - 1 = 40\%$.

Le choc appliqué au *S/C* de base sera donc de 40%.

Aujourd'hui, le calcul du q_x moyen sur le portefeuille est basé sur des tables de mortalité prenant en compte uniquement l'effet âge et sexe des assurés et est estimé à 0,15%.

$$q_x \text{ moyen BE} = \sum_{x=age_{min}}^{x=age_{max}} \frac{CSR_x^H \times q_x^H}{CSR_{Totaux}^H + CSR_{Totaux}^F} + \frac{CSR_x^F \times q_x^F}{CSR_{Totaux}^H + CSR_{Totaux}^F},$$

avec $CSR = \text{capitaux garantis sous risque}$.

Le choix du $q_x \text{ moyen BE}$ se justifie par la relation entre celui-ci et la charge totale de sinistres S représentant le numérateur dans le calcul du S/C .

Lorsque les paramètres sont bien calibrés, $S = q_x \text{ moyen BE} \times CSR_{\text{Totaux}}$.

Toute variation du $q_x \text{ moyen BE}$ devrait impacter le S/C de base. Cependant, on peut avoir certains écarts dus au calcul des S/C pour chaque model point⁶ dans le modèle de projection et non au global. De plus, les effets de la non-linéarité des comptes de participation bénéficiaire ne sont pas captés dans le calcul du S/C .

3.2.3.2 Impact de la segmentation du q_x par CSP dans le calibrage du choc

Pour affiner le calcul du choc « CAT mortality », nous allons recalculer le $q_x \text{ moyen BE}$ en prenant en compte la segmentation fine des q_x par CSP obtenus dans notre étude (cf paragraphe 2.3.2.8.). La formule de calcul du $q_x \text{ moyen BE}$ est réécrite sous la forme suivante :

$$q_x \text{ moyen BE} = \frac{1}{CSR_{\text{Totaux}}} \sum_{x=age_{\min}}^{x=age_{\max}} CSR_x^{HC} \times q_x^{HC} + CSR_x^{FC} \times q_x^{FC} + CSR_x^{HNC} \times q_x^{HNC} + CSR_x^{FNC} \times q_x^{FNC},$$

$$\text{Avec } CSR_{\text{Totaux}} = CSR_{\text{Totaux}}^{HC} + CSR_{\text{Totaux}}^{FC} + CSR_{\text{Totaux}}^{HNC} + CSR_{\text{Totaux}}^{FNC} .$$

La prise en compte de l'effet catégorie socio-professionnelle conduit ainsi à un $q_x \text{ moyen BE}$ plus faible estimé à 0,13%. Ce qui entraîne une hausse mécanique du choc traduit en S/C de 16%.

A iso S/C , nous obtenons les montants de STEC estimés suivants :

Tableau 17: Estimation des montants de STEC pour le risque « CAT mortality »

	Tables actuelles	Nouvelles tables d'expérience avec prise en compte de la CSP
Choc additif	0,08%	0,08%
Qx moyen BE	0,15%	0,13%
Choc traduit en impact S/C	54%	62%
Variation		16%
Montant estimé de STEC (en M€)	238	276
Variation (en M€)		38

⁶ Model point = Regroupement de contrats ayant les mêmes caractéristiques pour la modélisation du passif du portefeuille

Nous remarquons que la variation du STEC est sensible à la variation du q_x moyen BE. Une baisse de 0,02 point de ce paramètre fait augmenter le choc de 16% et le montant du STEC du même ordre de grandeur à iso S/C.

Cependant, compte tenu du poids très réduit du STEC « CAT mortality » sur le STEC total vie et des effets de diversification entre risques (matrice de corrélation), l'impact sur les métriques S2 du bilan global de l'entreprise ne sera pas significatif.

Conclusion

Le contexte de notre étude s'inscrit dans l'optique d'une meilleure maîtrise des risques d'un portefeuille de prévoyance collective, à la fois pour une tarification optimale, mais également dans le cadre de solvabilité 2, pour ajuster les calculs de capitaux économiques réglementaires.

La mise en place des DSN depuis 2017 a permis de collecter progressivement les données des assurés, dans le respect de la réglementation RGPD et de mener des études qui pallient le manque de données détaillées au moment de la souscription des affaires nouvelles, sur le périmètre de la prévoyance collective.

Nous nous sommes intéressés à l'étude de la garantie décès et à la construction de tables de mortalité d'expérience adaptées au profil de risque des assurés.

En effet, l'outil actuel de tarification en décès se base sur des tables réglementaires ajustées avec des correctifs définis à la suite d'études portefeuille (correctif par sexe et par catégorie socio-professionnelle).

La construction de tables d'expérience par sexe et par catégorie socio-professionnelle est proposée pour affiner la tarification et supprimer ces nombreux correctifs. Ces tables sont construites en utilisant les méthodes de modélisation paramétriques et non paramétriques usuelles dans l'étude des phénomènes de survie et un historique de données DSN sur la période 2017-2019.

La mortalité constatée chez les hommes non-cadres a conduit à la considérer comme celle de référence et à étudier l'influence du sexe et de la catégorie socio-professionnelle sur la mortalité du portefeuille avec le modèle de Cox.

Après validation statistique du modèle retenu pour définir la mortalité des hommes non-cadres, celles des autres groupes (femmes non-cadres, hommes cadres et femmes cadres) en sont déduites avec les coefficients du modèle relationnel de Cox.

Pour les femmes cadres, les faibles expositions et sinistralité génèrent une volatilité importante et conduisent à sous-estimer la mortalité sur ce segment avec un SMR supérieur à 1. Cependant, la marge obtenue sur les autres segments devrait permettre d'absorber cette potentielle sous-estimation.

La comparaison des tables d'expérience et des tables de tarification actuelles montre une surestimation de la mortalité par ces dernières. Même si opérationnellement, la tarification de la garantie décès est ajustée par des correctifs commerciaux, pour tenir compte de la réalité du risque, de l'environnement concurrentiel, ainsi que du comportement du risque arrêt de travail avec lequel la garantie décès est en général combinée, ce mémoire a permis d'objectiver le choix des tables pour une tarification adéquate et plus concurrentielle.

Afin d'optimiser la tarification du décès, des correctifs « âge actuariel » ont été introduits car les données de tarification sont agrégées à la maille groupe : l'entreprise fournit des paramètres de tarification pour la tête moyenne du groupe (âge moyen arithmétique, salaire moyen, nombre moyen d'enfants...). Ces correctifs ont été calibrés sur l'expérience du portefeuille pour prendre en compte la répartition des âges et des capitaux garantis pour autour de leurs moyennes.

Cependant, la construction de ces correctifs est très sensible à la structure du portefeuille, il serait donc utile d'actualiser régulièrement les conclusions de cette étude dont la robustesse sera amenée à

progresser au fur et à mesure de l'extension de l'utilisation des DSN et de la profondeur de l'historique des données disponibles.

Enfin, la prise en compte de l'effet catégorie socio-professionnelle dans le calcul du q_x moyen BE du portefeuille a permis d'affiner le calibrage du choc du risque « CAT mortality ». Ainsi, son montant de STEC estimé est ajusté à la hausse, mais l'impact sur les métriques S2 du bilan global de l'entreprise n'est pas matériel avec les effets de diversification entre risques.

Néanmoins, afin d'optimiser le calcul du STEC sur ce risque, il serait techniquement plus judicieux de proposer un q_x moyen BE calibré pour chaque model point et donc neutraliser l'effet variation du STEC induit par la variation des chocs traduits en S/C . Toutefois, la quantité limitée de données disponibles à ce niveau de granularité rend cet exercice difficile.

Par ailleurs, il serait également intéressant de raffiner l'étude de la mortalité sur ce périmètre en intégrant de nouvelles variables discriminantes telles que le secteur d'activité, la zone géographique d'implantation des entreprises et de résidence des salariés, avec l'augmentation progressive de la profondeur et du volume des données DSN.

Références

Boice, & Rothman. (1979), Epidemiologic Analysis with a Programmable Calculator, Washington DC: US Government Printing.

Breslow, & Day. (1987), Statistical methods in cancer research (vol 2), Lyon: IARC Scientific Publ n°82.

Decamps R. (2015), Création et étude d'une table d'incidence en invalidité, ENSAE.

F D K Liddel (1984), Simple exact analysis of the standardised mortality ratio, Journal of Epidemiology and Community Health.

Goblet M. (2018), Evaluation de la mortalité en assurance collective, ISUP.

Gorrand R. (2019), Solvabilité 2 hors BEL, Centre d'études actuarielles.

Gorrand R. (2019), Bilan Solvabilité 2, Centre d'études actuarielles.

Khalidi N. (2017), Construction de tables de mortalité fonctionnaire et impacts tarifaires sur le produit Obsèques, Centre d'études actuarielles.

Lopez O. (2019), Survival analysis and other topics, Sorbonne Université, Laboratoire de Probabilités, Statistique et Modélisation.

Ndiaye El H. (2016), Construction d'une table de mortalité et lissage par positionnement, ISFA.

Petauton P. (1996), Théorie et pratique de l'assurance vie, Dunod.

Planchet F., Thérond P.E. (2011), Modélisation statistique des phénomènes de durée – applications actuarielles, Paris : Economica (première édition en 2006).

Saint-pierre P. (2015), Introduction à l'analyse des durées de survie, Université Pierre et Marie Curie.

Sauveplane P. (2019), Assurances de personnes, Centre d'études actuarielles.

Staad F. (2018), Les engagements d'assurance vie, Centre d'études actuarielles.

<https://www.verkinder-courtage.com/sant%C3%A9-pr%C3%A9voyance-placements/en-savoir-prev>

<https://www.franceassureurs.fr/>

<https://droit-finances.commentcamarche.com/faq/24576-accord-de-branche-definition>

<https://www.ameli.fr/>

<http://www.dsn-info.fr/pourquoi.htm>

https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000035514715/

<http://www.ressources-actuarielles.net/>

Annexes

Annexe 1 : Critère de validation de Cochran

Pour construire nos tables d'expérience, il est important de disposer de données suffisantes pour obtenir des résultats robustes. Le critère de validation de Cochran nous permet de valider ce point et repose sur la théorie suivante :

$$\begin{cases} N_x \times \hat{q}_x \geq 5, \\ N_x \times (1 - \hat{q}_x) \geq 5. \end{cases}$$

- Avec \hat{q}_x le taux de mortalité à l'âge x ,
- N_x l'exposition à l'âge x .

Autrement dit, pour chaque âge, le nombre de décès et le nombre de survivants doit être supérieur ou égal à 5. Ce critère de Cochran est respecté pour les hommes non-cadres qui représente notre groupe de référence pour les âges allant de 19 à 70 ans.

Age	$N_x \times \hat{q}_x$	$N_x \times (1 - \hat{q}_x)$
19	5	11 677
20	5	17 622
21	16	23 398
22	6	28 063
23	12	30 956
24	7	32 800
25	13	34 212
26	8	35 522
27	9	36 774
28	18	37 688
29	20	38 567
30	19	39 439
31	15	40 082
32	13	40 638
33	21	40 445
34	16	40 917
35	25	42 334
36	27	44 103
37	18	45 121
38	35	44 598
39	39	43 677
40	37	42 949
41	38	42 799
42	36	43 680
43	45	45 143
44	51	47 453

Age	$N_x \times \hat{q}_x$	$N_x \times (1 - \hat{q}_x)$
45	71	48 943
46	73	49 276
47	96	48 575
48	96	47 425
49	98	46 573
50	98	46 130
51	111	45 890
52	113	45 662
53	133	45 198
54	139	44 588
55	163	43 489
56	171	41 563
57	168	39 217
58	230	36 370
59	243	31 963
60	131	16 263
61	97	11 028
62	43	6 314
63	28	4 344
64	17	3 254
65	20	2 417
66	9	1 618
67	13	1 230
68	6	1 006
69	4	850
70	8	651

Annexe 2 : Les modèles à hasards proportionnels (cf. référence : SAINT-PIERRE Philippe)

a- Modèle général

Ces modèles s'écrivent sous la forme d'une structure multiplicative exprimant un effet multiplicatif de différents covariables sur la fonction de hasard. Ils sont caractérisés par la relation suivante pour tout $t > 0$:

$$\mu(t/Z) = \mu_0(t) \times h(\theta^T \times Z).$$

Avec $\mu_0(t)$ une fonction de hasard de base commune à tous les individus, Z un vecteur de covariables, θ le paramètre d'intérêt et h une fonction positive. La fonction de hasard μ s'exprime ainsi sous la forme d'un produit d'une fonction dépendant uniquement du temps et d'une autre qui ne dépend pas de celui-ci. Ils sont dits à hasard proportionnels car pour tous individus i et j ayant pour covariables Z_i et Z_j , le rapport des fonctions de hasard ne varie pas au cours du temps, ce qui implique que les fonctions de hasard sont proportionnelles. Nous avons la relation suivante :

$$\frac{\mu(t/Z_i)}{\mu(t/Z_j)} = \frac{h(\theta^T \times Z_i)}{h(\theta^T \times Z_j)}.$$

Il s'agit d'une hypothèse importante qu'il faut vérifier pour valider le modèle.

b- Cas particulier du modèle de Cox :

Dans le cas particulier du modèle de Cox, la fonction h positive, présentée précédemment qui a été choisie, est celle exponentielle. Le modèle s'écrit alors comme suit :

$$\mu(t/Z) = \mu_0(t) \times \exp(\theta^T \times Z).$$

Où Z est un vecteur de covariables de dimension $p \times 1$ et θ un vecteur de coefficient de régression de dimension $p \times 1$.

c- Vraisemblance partielle de Cox :

L'idée générale est d'estimer uniquement le coefficient de régression θ en considérant la fonction μ_0 comme un paramètre de nuisance qu'on ne cherche pas à estimer. Cox considère qu'aucune information ne peut être donnée sur θ par les intervalles sur lesquels aucun événement n'a lieu, car μ_0 peut être considérée comme nulle dans ceux-ci. Nous travaillons donc conditionnellement à l'ensemble des instants où un décès a lieu.

soit :

- D le nombre de décès observés parmi les n individus observés,
- $T_1 < T_2 < \dots < T_D$, les temps de décès distincts,
- $(1), (2), \dots, (D)$, les indices des individus décédés respectivement en T_1, T_2, \dots, T_D ,
- Z_i la valeur des covariables de l'individu i ,
- $R(T_i)$, l'ensemble des individus encore à risque à T_i^- (juste avant T_i).

On va supposer dans un premier temps qu'il n'y a qu'un seul décès à chaque temps d'événement. La probabilité qu'il y ait un décès en T_i dans l'intervalle $[T_i, T_i + \Delta t]$ s'écrit :

$$\sum_{j \in R(T_i)} \mu_0(T_i) \times \exp(\theta^T \times Z_j),$$

La probabilité que l'individu i décède en T_i sachant qu'un décès a eu lieu en T_i vaut :

$$\frac{\mu_0(T_i) \times \exp(\theta^T \times Z_{(i)})}{\sum_{j \in R(T_i)} \mu_0(T_i) \times \exp(\theta^T \times Z_j)} = \frac{\exp(\theta^T \times Z_{(i)})}{\sum_{j \in R(T_i)} \exp(\theta^T \times Z_j)}.$$

En simplifiant par $\mu_0(T_i)$, nous obtenons une expression de la probabilité qui dépend uniquement du paramètre θ . Etant donné qu'il y a des contributions à la vraisemblance qu'à chaque temps de décès, la vraisemblance partielle de Cox est définie comme le produit des probabilités sur les temps de décès et s'écrit sous la forme suivante :

$$L_{cox}(\theta) = \prod_{i=1}^D \frac{\exp(\theta^T \times Z_{(i)})}{\sum_{j \in R(T_i)} \exp(\theta^T \times Z_j)}.$$

Par conséquent, le paramètre θ peut être estimé par maximisation de la vraisemblance partielle de Cox sans connaître la fonction de hasard de base $\mu_0(t)$.

A noter que la vraisemblance partielle n'est pas une vraisemblance dans le sens statistique du terme cependant elle se comporte comme telle. On peut donc développer une théorie asymptotique similaire et l'utiliser pour estimer et tester les coefficients de régression θ .

d- Estimation des coefficients de régression θ

A partir de la vraisemblance partielle, il est possible d'obtenir une estimation du vecteur θ de dimension $p \times 1$. Notons :

$$l(\theta) = \log(L_{cox}(\theta)) = \sum_{i=1}^D \left[\theta^T \times Z_{(i)} - \log \left(\sum_{j \in R(T_i)} \exp(\theta^T \times Z_j) \right) \right].$$

La log vraisemblance partielle et $U(\theta)$ la fonction score vecteur des dérivées premières de $l(\theta)$ de dimension $p \times 1$

$$U(\theta) = \frac{\partial l(\theta)}{\partial \theta} = \left(\frac{\partial l(\theta)}{\partial \theta_1}, \dots, \frac{\partial l(\theta)}{\partial \theta_p} \right) = \sum_{i=1}^D \left[Z_{(i)} - \frac{\sum_{j \in R(T_i)} Z_j \times \exp(\theta^T \times Z_j)}{\sum_{j \in R(T_i)} \exp(\theta^T \times Z_j)} \right].$$

$$U(\theta) = \left(\sum_{i=1}^D \left[Z_{(i),1} - \frac{\sum_{j \in R(T_i)} Z_{j,1} \times \exp(\theta^T \times Z_j)}{\sum_{j \in R(T_i)} \exp(\theta^T \times Z_j)} \right], \dots, \sum_{i=1}^D \left[Z_{(i),p} - \frac{\sum_{j \in R(T_i)} Z_{j,p} \times \exp(\theta^T \times Z_j)}{\sum_{j \in R(T_i)} \exp(\theta^T \times Z_j)} \right] \right).$$

L'estimateur $\hat{\theta}$ du coefficient de régression est solution de l'équation $U(\theta) = 0$.

Il n'y a pas de solution exacte à ce problème, cependant des algorithmes implémentés dans certains logiciels tel que celui de Newton-Raphson sont utilisés pour obtenir une solution.

Par ailleurs, un estimateur consistant de la matrice de variance-covariance de θ peut être obtenu à partir de l'inverse de la matrice d'information de Fisher :

$$\widehat{Var}(\hat{\theta}) = \{I(\hat{\theta})\}^{-1} \text{ où le terme } (i, j) \text{ de la matrice } I(\theta) \text{ s'écrit } [I(\theta)]_{i,j} = -\frac{\partial^2 l(\theta)}{\partial \theta_i \partial \theta_j}.$$

a- Test de significativité des coefficients de régression θ

Pour valider certaines hypothèses des modèles de régression, des tests statistiques sont utilisés. Pour les coefficients de régression, les tests portent sur l'hypothèse suivante :

$$H_0: \theta = \theta_0$$

Les trois statistiques suivantes qui suivent chacune une loi du khi-deux à p degrés de liberté sous l'hypothèse H_0 sont utilisées :

- Statistique du rapport de vraisemblance qui mesure la distance entre $\log(L_{cox}(\hat{\theta}))$ et $\log(L_{cox}(\theta_0))$:

$$X_{LRT}^2 = 2 \times [\log(L_{cox}(\hat{\theta})) - \log(L_{cox}(\theta_0))] \xrightarrow{H_0} X^2(p),$$

- Statistique de Wald qui mesure l'écart entre $\hat{\theta}$ et θ_0 :

$$X_W^2 = (\hat{\theta} - \theta_0)^T \times I(\hat{\theta}) \times (\hat{\theta} - \theta_0) \xrightarrow{H_0} X^2(p),$$

- Statistique du score qui mesure la pente de la tangente en θ_0 :

$$X_S^2 = (U(\theta_0))^T \times (I(\theta_0))^{-1} \times (U(\theta_0)) \xrightarrow{H_0} X^2(p).$$

La statistique du rapport de vraisemblance ne nécessite pas de calculer les dérivées secondes de la log-vraisemblance et celle du score ne nécessite pas l'estimation de $\hat{\theta}$.

De ces statistiques, peuvent être déduits des tests partiels permettant de tester des hypothèses uniquement sur certaines coordonnées du vecteur θ . Supposons par exemple, vouloir tester si l'ajout d'une nouvelle variable Z_p dans un modèle avec $p - 1$ variables, apporte plus d'informations sur la distribution des durées de vie que le modèle sans cette variable.

Dans ce cas, le test d'hypothèse est : $H_0: \theta = \theta_0$ où $\theta = (\theta_1, \dots, \theta_p)$ et $\theta_0 = (\theta_1, \dots, \theta_{p-1}, 0)$.

Autrement dit, on teste $H_0: \theta_p = 0$.

Les statistiques de tests se réécrivent ainsi sous la forme suivante et suivent une loi du khi-2 à 1 degré de liberté :

- Statistique du rapport de vraisemblance

$$X_{LRT}^2 = 2 \times [\log(L_{cox}(\hat{\theta})) - \log(L_{cox}(\theta_0))] \xrightarrow{H_0} X^2(1),$$

- Statistique de Wald :

$$X_W^2 = (\hat{\theta} - \theta_0)^T \times I(\hat{\theta}) \times (\hat{\theta} - \theta_0) \xrightarrow{H_0} X^2(1),$$

- Statistique du score :

$$X_S^2 = (U(\theta_0))^T \times (I(\theta_0))^{-1} \times (U(\theta_0)) \xrightarrow{H_0} X^2(1),$$

Où $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_p)$ et $\hat{\theta}_0 = (\hat{\theta}_1, \dots, \hat{\theta}_{p-1}, 0)$.

Annexe 3 : Ajustement du modèle de Makeham par maximum de vraisemblance discret

a- Méthode générale :

On suppose que la vraie probabilité de décéder à l'âge x notée $q_x(\theta)$ dépend d'un paramètre vectoriel θ . La méthode suivante permet de déterminer ce vecteur de paramètres pour ajuster au mieux la mortalité réelle aux taux bruts \hat{q}_x estimés par la méthode de Kaplan Meier.

Soit \tilde{d}_x le nombre de décès estimé à l'âge x obtenu à partir de l'estimation des \hat{q}_x .

Nous avons : $\tilde{d}_x = N_x \times \hat{q}_x$ où N_x représente le nombre de personnes exposées au risque de décès sur l'intervalle $[x; x + 1]$.

Soit X une variable aléatoire représentant le décès de l'individu d'âge x , X vaut 1 si l'individu décède avec la probabilité $q_x(\theta)$ et 0 sinon. X suit une loi de Bernoulli de paramètre $q_x(\theta)$.

Par conséquent, en supposant l'indépendance des décès des N_x individus d'âge x exposés, nous pouvons approximer \tilde{d}_x par une loi binomiale $B(N_x, q_x(\theta))$.

Nous allons utiliser l'approximation gaussienne d'une loi binomiale pour une taille d'échantillon assez grande et ce grâce au critère de Cochran (cf. annexe 1).

Nous avons alors $\tilde{d}_x \approx N(N_x \times q_x(\theta), N_x \times q_x(\theta) \times (1 - q_x(\theta)))$. Pour rappel lorsqu'une variable aléatoire Y suit une loi binomiale $B(n, p)$, nous avons $E(Y) = n \times p$ et $V(Y) = n \times p \times (1 - p)$.

Nous pouvons alors en déduire la loi de \hat{q}_x puisque $\tilde{d}_x = N_x \times \hat{q}_x$. Nous avons alors :

$$\hat{q}_x \approx N\left(q_x(\theta), \sigma^2(\theta) = \frac{q_x(\theta) \times (1 - q_x(\theta))}{N_x}\right),$$

Pour l'ensemble des observations $(\hat{q}_n, \dots, \hat{q}_m)$, la fonction de vraisemblance dépendant du paramètre θ peut donc s'écrire sous la forme suivante :

$$L(\hat{q}_n, \dots, \hat{q}_m / \theta) = \prod_{x=n}^{x=m} \frac{1}{\sigma(\theta)\sqrt{2\pi}} \exp\left(-\frac{1}{2} \times \frac{(q_x(\theta) - \hat{q}_x)^2}{\sigma^2(\theta)}\right).$$

On en déduit la log -vraisemblance :

$$\ln(L(\hat{q}_n, \dots, \hat{q}_m / \theta)) = \sum_{x=n}^{x=m} \ln\left(\frac{1}{\sigma(\theta)\sqrt{2\pi}}\right) - \frac{1}{2} \times \sum_{x=n}^{x=m} \frac{(q_x(\theta) - \hat{q}_x)^2}{\sigma^2(\theta)}.$$

Ainsi, la maximisation de la vraisemblance en θ est équivalente à la minimisation de $C(\theta)$ avec

$$C(\theta) = \sum_{x=n}^{x=m} \frac{(q_x(\theta) - \hat{q}_x)^2}{\hat{\sigma}^2} = \sum_{x=n}^{x=m} \frac{N_x}{\hat{q}_x \times (1 - \hat{q}_x)} \times (q_x(\theta) - \hat{q}_x)^2.$$

Avec $\hat{\sigma}^2 = \frac{\hat{q}_x \times (1 - \hat{q}_x)}{N_x}$ estimation empirique de $\sigma^2(\theta)$ inconnue.

Le choix de des valeurs initiales est très important pour l'application et la convergence du modèle.

b- Application au modèle de Makeham :

Nous rappelons que pour le modèle de Makeham, nous avons :

$q_x(\theta) = 1 - \exp\left[-a - \frac{b}{\ln(c)} \times c^x \times (c - 1)\right]$ avec $\theta = (a, b, c)$, et $\theta_0 = (a_0, b_0, c_0)$ vecteur de paramètres initiaux du modèle.

$$\mu_x = a + b \times c^x, \quad p_x = 1 - q_x = \exp\left[-a - \frac{b}{\ln(c)} \times c^x \times (c - 1)\right].$$

Posons $s = \exp(-a)$ et $g = \exp\left(\frac{-b}{\ln(c)}\right)$, nous avons alors $\ln(p_x) = \ln(s) + c^x \times (c - 1) \times \ln(g)$.

Le choix des paramètres initiaux du modèle $\theta_0 = (a_0, b_0, c_0)$ se fait en utilisant l'ordonnée à l'origine et la pente de la droite de régression de $\ln|q_{x+1} - q_x|$ et obtenir ainsi g et c .

Nous pouvons réécrire s sous la forme suivante : $s = \frac{(1-q_x)}{\exp[(c-1) \times \log(g) \times c^x]} = \exp(-a)$.

Cette méthode permet d'obtenir des ordres de grandeur pour s , c et g et en déduire a , b et c pour initialiser le problème d'optimisation de l'ajustement par maximum de vraisemblance.

Annexe 4 : intervalle de confiance des q_x - approche par approximation gaussienne (validation)

Makeham

Age	\hat{q}_x	N_x	$38\,416 \times \hat{q}_x \times (1 - \hat{q}_x)$
19	0,00018	11 682	7
20	0,00020	17 627	8
21	0,00021	23 414	8
22	0,00022	28 069	8
23	0,00023	30 968	9
24	0,00025	32 807	10
25	0,00027	34 225	10
26	0,00029	35 530	11
27	0,00031	36 783	12
28	0,00033	37 706	13
29	0,00036	38 587	14
30	0,00039	39 458	15
31	0,00042	40 097	16
32	0,00045	40 651	17
33	0,00049	40 466	19
34	0,00053	40 933	21
35	0,00058	42 359	22
36	0,00063	44 130	24
37	0,00069	45 139	27
38	0,00076	44 633	29
39	0,00083	43 716	32
40	0,00090	42 986	35
41	0,00099	42 837	38
42	0,00108	43 716	42
43	0,00119	45 188	46
44	0,00130	47 504	50
45	0,00143	49 014	55
46	0,00157	49 349	60
47	0,00173	48 671	66
48	0,00190	47 521	73
49	0,00209	46 671	80
50	0,00230	46 228	88
51	0,00253	46 001	97
52	0,00278	45 775	107
53	0,00306	45 331	117
54	0,00338	44 727	129
55	0,00372	43 652	142
56	0,00410	41 734	157
57	0,00452	39 385	173
58	0,00498	36 600	190
59	0,00549	32 206	210
60	0,00606	16 394	231
61	0,00668	11 125	255
62	0,00737	6 357	281
63	0,00813	4 372	310
64	0,00897	3 271	342
65	0,00990	2 437	376
66	0,01092	1 627	415
67	0,01205	1 243	457
68	0,01330	1 012	504
69	0,01468	854	556
70	0,01621	659	613

Whittaker-Henderson

Age	\hat{q}_x	N_x	$38\,416 \times \hat{q}_x \times (1 - \hat{q}_x)$
19	0,0004	11 682	15
20	0,0004	17 627	15
21	0,0004	23 414	14
22	0,0004	28 069	14
23	0,0003	30 968	13
24	0,0003	32 807	13
25	0,0003	34 225	13
26	0,0003	35 530	13
27	0,0004	36 783	13
28	0,0004	37 706	14
29	0,0004	38 587	15
30	0,0004	39 458	15
31	0,0004	40 097	16
32	0,0004	40 651	16
33	0,0005	40 466	17
34	0,0005	40 933	19
35	0,0005	42 359	20
36	0,0006	44 130	22
37	0,0006	45 139	24
38	0,0007	44 633	26
39	0,0007	43 716	28
40	0,0008	42 986	31
41	0,0009	42 837	34
42	0,0010	43 716	37
43	0,0011	45 188	41
44	0,0012	47 504	46
45	0,0013	49 014	52
46	0,0015	49 349	57
47	0,0017	48 671	64
48	0,0018	47 521	70
49	0,0020	46 671	77
50	0,0022	46 228	84
51	0,0024	46 001	93
52	0,0027	45 775	104
53	0,0031	45 331	118
54	0,0035	44 727	133
55	0,0040	43 652	151
56	0,0045	41 734	172
57	0,0051	39 385	193
58	0,0056	36 600	216
59	0,0062	32 206	237
60	0,0067	16 394	255
61	0,0071	11 125	271
62	0,0075	6 357	285
63	0,0078	4 372	297
64	0,0081	3 271	309
65	0,0084	2 437	319
66	0,0087	1 627	330
67	0,0089	1 243	340
68	0,0092	1 012	350
69	0,0095	854	360
70	0,0097	659	370

Annexe 5 : Taux de mortalité par sexe et par CSP

Age	HNC	HC	FNC	FC
19	0,04%	0,02%	0,02%	0,01%
20	0,04%	0,02%	0,02%	0,01%
21	0,04%	0,02%	0,02%	0,01%
22	0,04%	0,01%	0,02%	0,01%
23	0,03%	0,01%	0,02%	0,01%
24	0,03%	0,01%	0,02%	0,01%
25	0,03%	0,01%	0,02%	0,01%
26	0,03%	0,01%	0,02%	0,01%
27	0,04%	0,01%	0,02%	0,01%
28	0,04%	0,02%	0,02%	0,01%
29	0,04%	0,02%	0,02%	0,01%
30	0,04%	0,02%	0,02%	0,01%
31	0,04%	0,02%	0,02%	0,01%
32	0,04%	0,02%	0,03%	0,01%
33	0,05%	0,02%	0,03%	0,01%
34	0,05%	0,02%	0,03%	0,01%
35	0,05%	0,02%	0,03%	0,01%
36	0,06%	0,02%	0,03%	0,01%
37	0,06%	0,03%	0,04%	0,02%
38	0,07%	0,03%	0,04%	0,02%
39	0,07%	0,03%	0,04%	0,02%
40	0,08%	0,03%	0,05%	0,02%
41	0,09%	0,04%	0,05%	0,02%
42	0,10%	0,04%	0,06%	0,02%
43	0,11%	0,05%	0,06%	0,03%
44	0,12%	0,05%	0,07%	0,03%
45	0,13%	0,06%	0,08%	0,03%
46	0,15%	0,06%	0,09%	0,04%
47	0,17%	0,07%	0,10%	0,04%
48	0,18%	0,08%	0,11%	0,04%
49	0,20%	0,08%	0,12%	0,05%
50	0,22%	0,09%	0,13%	0,05%
51	0,24%	0,10%	0,14%	0,06%
52	0,27%	0,11%	0,16%	0,07%
53	0,31%	0,13%	0,18%	0,08%
54	0,35%	0,15%	0,20%	0,09%
55	0,40%	0,17%	0,23%	0,10%
56	0,45%	0,19%	0,26%	0,11%
57	0,51%	0,21%	0,30%	0,13%
58	0,56%	0,24%	0,33%	0,14%
59	0,62%	0,26%	0,36%	0,15%
60	0,67%	0,28%	0,39%	0,17%
61	0,71%	0,30%	0,42%	0,18%
62	0,75%	0,32%	0,44%	0,18%
63	0,78%	0,33%	0,46%	0,19%
64	0,81%	0,34%	0,48%	0,20%
65	0,84%	0,35%	0,49%	0,21%
66	0,87%	0,37%	0,51%	0,21%
67	0,89%	0,38%	0,52%	0,22%
68	0,92%	0,39%	0,54%	0,23%
69	0,95%	0,40%	0,56%	0,23%
70	0,97%	0,41%	0,57%	0,24%