

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

Par : Madame Stéphanie LEGRAND

Titre du mémoire : Modélisation du coefficient de mortalité en assurance collective

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

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

Membres présents du jury de la Signature : Entreprise : Groupama Gan Vie
filiale :

Nom : Vincent LAUDOU

Signature :



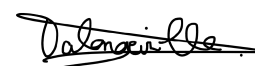
Directeur de mémoire en
entreprise

Membres présents du jury de
l'Institut des Actuares :

Signature :

Nom : Marie DALONGEVILLE

Signature :



Invité :

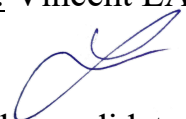
Nom :

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 : Vincent LAUDOU



Signature du candidat :



Table des matières

Résumé	1
Abstract	2
Note de synthèse	3
Executive summary	7
Introduction	11
Remerciements	13
1 Prévoyance collective	14
1.1 Généralités	14
1.2 Les garanties décès	15
1.2.1 Les prestations en capital	15
1.2.2 Les prestations en rente	16
1.2.3 Les allocations obsèques	16
1.3 Cadre juridique	16
1.3.1 Loi du 31 décembre 1989	17
1.3.2 Loi du 8 août 1994	17
1.3.3 Loi du 17 juillet 2001	17
1.3.4 Arrêté du 20 décembre 2005	18
1.3.5 Accord National Interprofessionnel du 11 janvier 2008	18
2 Modèle économique du décès et tarification	19
2.1 Modèle économique	19
2.1.1 Complexité du décès suite à arrêt de travail	19
2.1.2 Cartographie des types de décès	21
2.2 Modèle tarifaire	22
2.2.1 Contenu de la prime	22
2.2.2 Calcul de la prime	23
2.2.2.1 Taux technique	23
2.2.2.2 Tables de mortalité	23
2.2.2.3 Prime individuelle	23
2.2.2.4 Prime collective	24
2.2.3 Audit de l’outil actuel	26
3 Données et périmètre	27
3.1 Présentation des données	27
3.1.1 Données Insee	27

3.1.2	Données DSN	28
3.1.3	Données internes	29
3.2	Utilisation de ces données	29
3.3	Périmètre	29
3.4	Traitements	30
3.4.1	Contrats de travail	30
3.4.2	Arrêts de travail	31
3.4.3	Suspensions pour invalidité	31
3.4.4	Regroupement des tables	31
3.5	Comparaison avec les données Insee	32
3.5.1	Rapprochement des décès DSN avec les données Insee	32
3.5.2	Rapprochement des individus non décédés en DSN avec les données Insee	32
3.6	Derniers traitements	33
4	Statistiques descriptives et analyses	36
4.1	Statistiques descriptives	36
4.1.1	Données internes	36
4.1.2	Données DSN	37
4.2	Analyse de la répartition des décès	44
4.3	Analyse de l'évolution des décès sur un modèle simple	44
5	Modélisation	48
5.1	Présentation des modèles	48
5.1.1	Fonctions liées à la mortalité	48
5.1.1.1	Fonction de survie	48
5.1.1.2	Force de mortalité	48
5.1.1.3	Lien avec la probabilité de décès	48
5.1.2	Censure et Troncature	49
5.1.2.1	Censure	49
5.1.2.2	Troncature	49
5.1.3	Taux brut de mortalité	51
5.1.4	Kaplan-Meier	52
5.1.5	GLM	53
5.1.5.1	Modèle de Gompertz	54
5.1.5.2	GLM Poisson	54
5.1.5.3	GLM Poisson avec référence	55
5.1.5.4	Sélection des variables et regroupements des modalités	56
5.1.5.5	Sélection et validation du modèle	56
5.1.6	SMR par critère	57
5.1.6.1	Définition du SMR	58
5.1.6.2	Méthode itérative	58
5.1.6.3	Sélection et validation du modèle	60
5.1.7	Lissage P-splines	61
5.1.7.1	Définition des B-splines	61
5.1.7.2	Base de B-splines	61
5.1.7.3	Vraisemblance pénalisée	62
5.1.7.4	Implémentation	62
5.2	Résultats	62
5.2.1	Taux brut de mortalité	62

5.2.2	Kaplan-Meier	65
5.2.3	GLM Poisson	66
5.2.4	GLM Poisson avec référence	72
5.2.5	SMR par critère	78
5.2.6	Comparaison des résultats obtenus avec les différentes méthodes	83
6	Tarification	86
6.1	Outil actuel	86
6.2	Comparaison des lois de mortalité	86
	Conclusion	89
	Annexes	92
1	Analyse de la répartition des décès	92
2	Analyse du décès sur un exemple	94

Résumé

Ce mémoire a pour sujet l'étude des décès sans arrêt de travail en prévoyance collective dans un but de tarification. Le décès est un risque très volatile et difficile à tarifer. Des tables réglementaires sont en vigueur mais ne sont pas adaptées au portefeuille de chaque assureur. De plus, la distinction, entre les décès directs et les décès faisant suite à un arrêt de travail, n'est pas présente et peut être intéressante à évaluer. Ces deux risques sont également différents d'un point de vue comptable car la provision des décès suite à arrêt de travail doit être établie dès la connaissance de l'arrêt de travail par l'assureur. Un enjeu de cette étude est de mieux tarifer les décès directs dans un premier temps ainsi que de comprendre les avantages à rattacher les décès suite à arrêt de travail à la date de l'arrêt, dans les comptes afin de mobiliser les bonnes provisions.

Dans un premier temps, un travail important de préparation des données a été effectué. En assurance collective, peu de données sont généralement connus de l'assureur. Seules des informations agrégées sont connues. Cependant, avec la DSN les assureurs sont capables de récupérer les informations de chaque affilié sous leur responsabilité. Ainsi, ce détail d'information permet de réaliser des études détaillées sur le portefeuille assuré. Cette source d'information étant tout de même assez récente, un rapprochement avec les données de l'Insee a été effectué afin de vérifier la qualité des déclarations.

Une fois tous ces traitements réalisés, la modélisation a pu être effectuée. Afin de vérifier l'allure des données, des taux bruts de mortalité ont été calculés, puis le modèle de Kaplan-Meier a été implémenté. Cependant, l'objectif de ce mémoire n'est pas de construire une loi mais plutôt de calculer des coefficients d'abattement afin de se positionner par rapport aux lois de référence. Ainsi, des GLM Poisson ont été étudiés afin de connaître les critères significatifs. Puis cette méthode a de nouveau été implémentée mais en intégrant directement les tables de référence en offset afin d'obtenir des coefficients de passage entre les lois de mortalité. Ce qui ressort de ces analyses est le fait que le sexe et l'âge sont bien paramétrés dans les lois de référence mais que la mortalité du portefeuille est très inférieure à celle de la population générale sur laquelle ces lois ont été construites. De plus, la CSP s'est révélée être un critère très discriminant. La région est une variable significative mais pas dans tous les modèles et peut détériorer l'adéquation du modèle aux données. Afin, de confirmer ces observations une dernière méthode basée sur le SMR a été implémentée. Cette méthode permet d'étudier l'impact de chaque critère un par un en choisissant l'ordre de ces critères. Cette méthodologie a permis de confirmer les observations effectuées précédemment. Une comparaison des résultats des deux meilleurs modèles a révélé que peu d'écarts sont répertoriés. Ainsi, cela renforce la confiance que l'on a des estimations.

Pour finir, une comparaison des résultats avec les lois utilisées actuellement en tarification a été effectuée. Elle a révélé que des écarts assez importants sont présents sur certaines CSP comme les ouvriers. Ainsi, l'outil de tarification actuel sous-estime le risque sur cette catégorie. Cependant pour les cadres les écarts sont très faibles.

Mots clés : Prévoyance collective, Décès direct, Table de mortalité, Taux bruts, Kaplan-Meier, GLM Poisson, SMR, Positionnement, Tarification.

Abstract

The subject of this project is the study of deaths without medical leave in group insurance and the aim is to perform a pricing of this risk. Death is a volatile risk which is hard to price. Regulatory tables are in effect but are not adequate to the portfolio of every insurer. Moreover, distinction between direct deaths and deaths that are consequent to a medical leave, is not performed but could be interesting to evaluate. Those two risks are also different from an accounting stand point because the provision for deaths that are consequent to a medical leave has to be put in place as soon as the insurer is informed of the medical leave. One of the stakes of this project, is to improve the pricing of direct deaths, but also to understand the benefits of linking the deaths that are consequent to a medical leave with the initial leave's date, in the accounts, in order to utilize the right provisions.

First, a substantial work in order to prepare the data has been performed. In group insurance, often the insurer does not have a lot of data. Only aggregated information are known. However, with the "DSN", insurers are now able to retrieve new data on every affiliated person in the portfolio. Therefore, with this detailed data, studies can be performed on the insurer's portfolio. This data source being still pretty new, a check is performed using the Insee's data to make sure of the quality of the statements.

Once those are completed, models can be implemented. In order to check the structure of the data, raw mortality rates are calculated, then a Kaplan-Meier model is implemented. Nonetheless, the goal of this project is not to built a mortality law but rather to come up with abatement coefficients in order to position ourselves in relation to the reference tables. Therefore, GLM have been implemented in order to establish the significant criteria. Then, this method was used again but the reference tables have been directly used in the offset term in order to calculate abatement coefficients. What emerges from these analysis is that the gender and age are well taken into account in the reference tables, but the mortality of the portfolio is way lower than the mortality of the french population, which was used to determine the reference tables. Furthermore, socio-professional categories turned out to be very significant. Region is also a significant variable but not in every model and can even lower the adequacy of the model to the data. In order to confirm those conclusions, a last method based on the SMR ratio has been implemented. This method enables you to study the criteria one by one and in the desired order. This methodology confirmed the previous observations. A comparison, of the results obtained with the two best methods, has revealed that the results are very similar. Therefore, this back ups the trust we have in the predictions.

To conclude, a comparison of the results with the tables that are currently used in pricing has been performed. This analysis revealed that significant differences exist on certain socio-professional categories like the working class. Thus, the current pricing software underestimates the risk of this class. However, concerning the professional and managerial staff the differences were very small.

Key words : Group insurance, Direct death, mortality tables , Raw mortality rates , Kaplan-Meier, Poisson GLM, SMR, Positioning, Pricing.

Note de synthèse

Le risque décès est difficile à évaluer et cela est accentué en assurance collective. En effet, peu d'informations sur la population assurée sont connues de l'assureur. Aucune information individuelle n'est connue tant qu'un sinistre n'a pas eu lieu. Mais avec la DSN (Déclaration Sociale Nominative), les assureurs peuvent récupérer des informations sur les individus qu'ils assurent. Cela peut donc permettre aux assureurs de groupe de mieux tarifier leurs contrats et de mieux piloter les risques qu'ils assurent. Le risque décès est très volatile et dépend de nombreux facteurs. Par ailleurs, deux types de décès peuvent être répertoriés, les décès directs et les décès faisant suite à un arrêt de travail. Ces deux types de décès sont différents car les provisions associées ne sont pas les mêmes.

Ainsi, ce mémoire a pour but de tarifier le décès direct via les données de la DSN et de comparer ces résultats aux lois de référence.

Dans un premier temps, les réglementations en vigueur concernant le risque décès sont étudiées. Différentes prestations peuvent être servies en cas de décès que ce soit sous forme de capital ou de rentes. D'autre part, la loi Evin de juillet 2001 impose le maintien des garanties décès aux personnes en arrêt de travail et oblige à ce que cet engagement soit couvert par une provision spécifique à chaque instant. Cette provision se nomme la "provision de maintien de la garantie décès" ou PMGD.

Le décès suite à arrêt de travail est donc particulier. Une provision est mise en place dès la connaissance de l'arrêt afin d'anticiper l'éventuel décès de l'affilié. Mais dans les comptes de résultats deux façons de procéder sont possibles pour prendre en compte les décès suite à arrêt de travail. En effet, ces derniers peuvent être rattachés à la date du décès et on ne tient alors pas compte de l'arrêt de travail, ou bien ils peuvent être rattachés à la date de l'arrêt de travail qui est l'évènement générateur du décès. Cette dernière solution est la meilleure et elle permet de rattacher les bonnes provisions au bon risque. En effet, en cas de mouvement de portefeuille, si la mauvaise PMGD est utilisée alors des boni ou des mali apparaîtront.

Le principe de la tarification est d'égaliser les engagements de l'assureur avec ceux de l'assuré. Mais en assurance collective cela est difficile à évaluer car très peu d'informations sont connues sur la population assurée. Par ailleurs, en décès, les garanties peuvent être sous formes de rentes et la durée de paiement de ces dernières dépend alors des bénéficiaires de l'assuré. Or à la souscription aucune information n'est connue sur les bénéficiaires de l'affilié. Ainsi, des hypothèses sont prises mais peuvent s'avérer fausses. Enfin les lois réglementaires actuelles, à savoir les TH 00-02 et TF 00-02, ont été calculées sur la population française générale de 2000 à 2002. Cependant, en 20 ans la mortalité a évolué grâce aux améliorations des conditions de vie, des modes de vie et de la médecine. En outre, la population française comprend les inactifs non retraités. Or, selon plusieurs études cette catégorie de la population a une mortalité bien supérieure au reste. Ainsi, ces éléments font que ces lois ne sont pas nécessairement adaptées au portefeuille d'un assureur de prévoyance collective.

Par ailleurs, l'outil de tarification actuel date de plusieurs années et n'est donc plus nécessairement adapté malgré le fait qu'il ait été revu il y a quelques années. Il est basé sur les lois de mortalité de l'Insee de 96-98, ce

qui n'est pas adapté au risque actuel. L'outil prend en compte le sexe, l'âge, la CSP et la classe tarifaire qui est basée sur le secteur d'activité de l'entreprise.

En ce qui concerne les données, différentes sources sont utilisées afin de fiabiliser les résultats. Les données Insee sur les décès sont utilisées. Elles répertorient tous les décès des français ainsi que ceux des étrangers sur le territoire français. Elles proviennent des déclarations effectuées en mairie. L'avantage de ce fichier est la rapidité de déclaration des décès, les informations y étant présentes seulement quelques mois après le décès de l'individu. Les données DSN sont la source principale de l'étude. Les informations sont transmises via la fiche de paie des salariés aux services publiques et chaque assureur peut récupérer les informations des assurés de son portefeuille. Elles répertorient tous les événements survenus pendant le mois pour l'assuré. Ce niveau de détail d'information est très intéressant et novateur en assurance collective. Enfin, les données internes des dossiers décès sont utilisées afin de produire des statistiques sur les décès réellement traités en gestion.

De nombreux traitements ont été effectués sur les données DSN pour calculer les périodes d'arrêt de travail afin de recalculer les périodes d'observation des individus. Ainsi, des hypothèses ont dû être prises car les informations sur le contrat de chaque individu ne sont pas connues. Par exemple, la franchise n'est pas connue. Après analyse, une franchise de 90 jours a été prise en compte.

Une fois les données retraitées, une étude a été menée afin de vérifier la qualité des déclarations décès dans la DSN. Ainsi, les données DSN et Insee ont été mises en relation afin de vérifier si tous les décès avaient été déclarés et à l'inverse si ceux qui l'étaient étaient corrects. Cette analyse a révélé qu'il manquait 420 décès qui ont été rajoutés à l'étude. Mais que dans la globalité les déclarations DSN étaient fiables. Ce rapprochement avait cependant des limites, en effet, certains individus n'ont pas pu être rapprochés à cause des noms de jeune fille par exemple. De plus, dans les données de l'Insee les frontaliers, étrangers travaillant en France, ne sont pas présents alors qu'ils le sont dans les données DSN.

Afin de mieux appréhender les données, des statistiques descriptives ont été réalisées. Ces dernières ont été faites en utilisant les données internes puis les données retraitées de la DSN. Ces statistiques ont permis d'évaluer la proportion de décès faisant suite à arrêt de travail dans les sinistres décès. Cette proportion est de 43% et sera utilisée par la suite afin de permettre la comparaison avec les lois de mortalité réglementaires qui comprennent tous les types de décès. Il a aussi été observé que l'âge moyen des individus décédés suite à un arrêt de travail est supérieur à celui des individus décédés sans arrêt de travail. La composition du portefeuille a également pu être évaluée. Ainsi, le portefeuille comprend légèrement plus d'hommes que de femmes et une grande majorité de non-cadres est présente. En évaluant les décès, la surmortalité des hommes est frappante ainsi que celle des non-cadres.

D'autre part, un modèle artificiel et simplifié a été mis en place afin d'observer les changements d'états dans une population définie comme se comportant comme la moyenne. Les différents états et les coefficients de passage entre ces derniers ont été définis intelligemment selon d'autres études. Cela a permis de voir qu'après 10 ans de scénario les proportions entre les deux types de décès n'étaient pas encore les mêmes que celles observées actuellement dans le portefeuille mais qu'elles tendent vers ce résultat.

Différents modèles ont été utilisés car différents avantages et inconvénients sont présents dans chaque méthode. Cette multiplication des approches permet aussi de fiabiliser les résultats et de les confronter en cas d'incohérence.

Les taux bruts de mortalité ont dans un premier temps été calculés afin d'obtenir une première approche de l'allure des données et de voir si la loi obtenue est bien en cohérence avec les autres lois de mortalité publiées. Les résultats ont été lissés en utilisant un lissage P-splines afin d'obtenir une courbe exploitable.

Comme attendu, aux âges extrêmes, les fluctuations sont importantes car l'exposition est plus faible et donc

l'incertitude plus élevée. La courbe lissée obtenue est bien croissante avec l'âge mais les coefficients semblent très inférieurs aux tables réglementaires.

De plus, en segmentant la population par CSP et sexe, il semble y avoir de très grandes disparités entre les différents groupes. La surmortalité des hommes est très visible ainsi que celle des ouvriers. Un résultat particulièrement intéressant est le fait que la mortalité des hommes cadres est légèrement supérieure au global à celle des femmes ouvrières alors que ces dernières ont la mortalité la plus élevée parmi les femmes tandis que celle des hommes cadres est la moins élevée parmi les hommes. Cela illustre le contraste important qui existe entre les hommes et les femmes mais également entre les différentes CSP.

Le modèle de Kaplan-Meier a ensuite été réalisé. Ce dernier est intéressant car il est non-paramétrique et permet donc de ne pas faire d'hypothèse sur les données. De plus, il prend en compte la censure en accordant plus de poids aux observations complètes.

Les courbes ont ici été séparés directement par sexe aux vues des observations précédentes. Les observations précédentes quant à l'allure des courbes sont bien confirmées ici également. Par exemple, il y a 6 ans d'écart entre le moment où la courbe masculine a un coefficient de mortalité de 0,5% et le moment où la courbe féminine atteint ce même niveau. Mais, cette méthode ne semble pas assez robuste pour supporter une segmentation plus fine, or d'autres critères semblent intéressants à étudier.

Ensuite un GLM Poisson a été implémenté. Ce modèle à l'avantage de permettre d'identifier rapidement les critères significatifs de la mortalité. La loi de Poisson a été utilisée car il s'agit de la loi employée habituellement dans la modélisation de la mortalité.

Dans un premier temps, toutes les variables explicatives ont été ajoutées au modèle. Les modèles ont été séparés en deux catégories avec ceux utilisant la variable âge et donc un coefficient de passage unique entre chaque âge, et ceux utilisant les tranches d'âge et donc des coefficients spécifiques pour chaque tranche d'âge. Une validation des modèles et de l'adéquation des modèles aux données est réalisée à l'aide de l'étude des résidus notamment.

Au final, le modèle par tranche d'âge retenant le sexe, les tranches d'âge, la CSP et la région avec des regroupement de modalités est celui qui est le mieux adapté aux données par rapport à la validation des données. Cependant, cela manque de précision car l'âge est un facteur très significatif et une loi par âge est souhaitable. De plus, un positionnement par rapport aux lois de référence est recherché. Ainsi, des coefficients de passages sont recherchés.

C'est pourquoi un modèle de Poisson utilisant les lois de référence, à savoir les TH 00-02 et TF 00-02, en offset a été implémenté. Ce modèle permet d'obtenir des lois de mortalité basées sur celles de référence ainsi que des coefficients de passage.

Au final, les variables significatives sont la CSP, la région et le croisement entre l'âge et la CSP. Le sexe et l'âge ne sont pas significatifs car ils sont déjà pris en compte dans les lois de référence et l'évolution en fonction de ces critères est la même dans les données étudiées et les lois de référence. Les disparités entre sexe et CSP sont encore très visibles. De plus, les régions sont regroupées selon 3 groupes ayant des mortalités différentes. Les résidus ont permis de valider le modèle.

Une comparaison avec les lois de référence en rajoutant de manière arbitraire la proportion des décès suite à arrêt de travail est effectuée. Cela démontre que la mortalité des salariés du portefeuille, est tout de même inférieure à celle de la population française générale. Les ouvriers ont la mortalité la plus élevée. Celle des employés est similaire mais un peu moins importante. La mortalité des professions intermédiaires est très faible aux âges jeunes mais similaire à celle des ouvriers aux âges élevés. Les cadres ont quant à eux la mortalité la plus faible.

Cependant, ce type de modèle impose des hypothèses assez fortes comme le fait de supposer que les données suivent une loi de Poisson. De plus, toutes les variables sont traitées ensemble dans le GLM, il est ainsi assez difficile d'observer les effets des différentes variables.

Une dernière méthode a été utilisée. Il s'agit, ici encore, d'une méthode de positionnement. Cette méthode est basée sur le SMR. Il s'agit du ratio entre le nombre de décès observés et le nombre de décès attendus si la population se comportait comme les lois de référence. Cette méthode possède plusieurs avantages. Les critères peuvent être étudiés un par un, indépendamment les uns des autres et dans l'ordre souhaité. L'effet de chaque critère est neutralisé à chaque étape. Cette méthode est donc itérative. Les résidus seront calculés, entre autres, afin d'étudier la qualité de l'ajustement final. Les critères étudiés sont ceux qui ont été retenus par les modèles précédents. L'ordre de ces critères est basé lui sur la significativité observée précédemment.

Ainsi, le premier critère étudié est le sexe car comme vu tout au long de cette étude la surmortalité masculine est très manifeste. Les deux coefficients obtenus sont très faibles mais très similaires. Cela indique que la mortalité du portefeuille est très inférieure à celle de la population française de 2000 à 2002 et que les lois de référence sont très bien paramétrées en fonction du sexe.

Ensuite, les tranches d'âge ont été étudiées. Ici encore, les variations entre générations ne sont pas très importantes ce qui permet de conclure que l'âge est bien pris en compte dans les lois de référence. De plus, ces observations sur les deux premiers critères étudiés sont cohérentes avec le fait que dans le GLM Poisson par référence, le sexe et l'âge n'étaient pas significatifs.

Pour finir, la CSP s'est révélée être un critère significatif dans tous les modèles c'est donc pour cette raison que cette variable est étudiée. Ici, l'effet est très visible et les disparités entre CSP très profondes. La mortalité croît de plus en plus dans l'ordre suivant : cadres, professions intermédiaires, employés et enfin ouvriers. Ces derniers ont la mortalité la plus élevée. La région a été étudiée mais n'a pas été retenue suite à l'étude des résidus.

Afin de fiabiliser les résultats, ceux des 2 meilleures méthodes ont été comparés. Les écarts observés sont très faibles, avec moins de 16% d'écart en valeur absolue. Ainsi, les résultats s'avèrent fiables et corrects. La méthode retenue est la dernière, celle utilisant le ratio SMR, car les coefficients sont facilement utilisables en tarification.

Pour finir, une comparaison des lois de mortalité obtenues avec celles utilisées dans l'outil de tarification actuel a été effectuée afin de voir les différences. Les décès suite à arrêt de travail ont été rajoutés de manière arbitraire afin de pouvoir comparer les lois. Cependant, cela fait partie d'une des limites de cette étude car cette approximation semble trop imprécise. A noter que l'outil de tarification actuel est basé sur les lois de 96-98 de l'Insee et non les lois TH 00-02 et TF 00-02 comme c'est le cas dans cette étude.

Les écarts les plus importants se situent sur les âges de plus de 60 ans et cela pour chaque sexe et CSP. Pour les cadres, les écarts sont très faibles. Les écarts sont plus importants pour les ouvriers avec au maximum environ 30% d'écart. Pour les employés et les professions intermédiaires l'analyse est plus délicate car l'outil de tarification actuel regroupe ces deux CSP. Ainsi, la mortalité des employés semble très sous-estimée actuellement alors que celle des professions intermédiaires est en adéquation avec la tarification actuelle.

A noter que l'outil comporte des classes de tarification qui dépendent du secteur d'activité de l'entreprise mais ce critère n'a pas été retenu dans l'étude car il n'était pas significatif. En effet, ce critère s'est avéré redondant par rapport à la catégorie socio-professionnelle qui capture déjà ces effets de classes.

Pour conclure, la mortalité du portefeuille de salariés étudié est très inférieure à celle de la population générale car les inactifs ont une surmortalité importante. Ainsi, les lois de référence ne sont pas adaptées à ce portefeuille et surestiment la mortalité. De plus, certains critères très discriminants comme la CSP ne sont pas pris en compte dans les lois de référence. Ainsi, cette étude a permis de mettre en évidence les critères impactant la mortalité des salariés et ceux qui ne le sont pas, et d'obtenir une loi ainsi que des coefficients de positionnement adaptés à la mortalité réelle du portefeuille. Cependant, seuls les décès directs ont été étudiés ici, ainsi une étude complémentaire sur les décès suite à arrêt de travail permettra d'obtenir une vision globale de la mortalité et de finaliser la tarification de ce risque.

Executive summary

The mortality risk is hard to evaluate and even more so in group insurance. Indeed, little information on the covered population is known by the insurer. No personal information is known until the risk has not occurred. But with the NSD (Nominative Social Declaration), insurers can collect information on the individuals that are covered by their policies. Therefore, this can enable group insurers to better evaluate the pricing of their policies and to better manage the covered risks. The mortality risk is very volatile and depends on many things. Moreover, two types of deaths can be defined, those without a medical leave and those with one. These two types of deaths are different because the associated provisions are not the same.

Therefore, the objective of this paper is to perform a pricing of deaths without a medical leave, by using the NSD's data and to compare these results with the regulatory tables .

First, the current laws regarding the mortality risk are studied. Different types of claims can be handed out in case of death, the claim can be an amount of money given once or an annuity. On the other hand, the Evin law of July 2001 enforces the maintenance of insurance cover regarding death, to the individuals in medical leave and enforces for this commitment to be covered by a specific provision at all times. This provision is called the "provision for the maintenance of the death guarantee".

Therefore, deaths that are consequent to a medical leave are particular. A provision is made as soon as a medical leave is brought to the attention of the insurer, in order to anticipate the potential death of the individual. However, in the profits and losses accounts, two ways to proceed are possible to take into account deaths that are consequent to a medical leave. Indeed, they can be linked to the date of death but then the medical leave is not taken into account, or they can be linked to the date of the medical leave which is the event at the root of the following death. This last solution is the best and it enables to link the right provisions to the matching risk. Indeed, in case of expansion or shrinkage of the portfolio, if the wrong provision is used then profits or losses will appear.

The goal of pricing is to even out the engagements of the insurer with those of the insured. But in group insurance that is difficult to evaluate because little information is known on the portfolio. Moreover, the claims can be annuities, and in that case the duration of the payments depends on the beneficiary and not the policyholder. However, when a company takes out a policy for its employees, no information is known on the beneficiaries. Therefore, hypothesis have to be taken but they can turn out to be false. Finally, the regulatory tables, which are the TH 00-02 and TF 00-02, are based on the French population from 2000 to 2002. But, in 20 years the mortality has evolved thanks to the improvement of life conditions, lifestyles and medicine. Moreover, the French population includes people that are unemployed. According to several studies this group has a higher mortality rate. Therefore, these factors explain why those tables are not necessarily adequate to the portfolio of a group insurer.

Moreover, the current pricing software has been used for several years now and is no longer well suited to the risk despite having been reviewed a few years ago. It is based on mortality tables from 96-98 published by the "Insee", which are not adapted to the current risk. The software takes into account the gender, the age, the

socio-professional category and the pricing category which is based on the line of business of the company.

In regards to the data, several sources have been used in order to reinforce the reliability of the results. Insee's death data has been used. Every death on the french soil is present as well as the ones of french people who died in a foreign country. They come from the statement made at the city hall by the relatives. The advantage of this file is the fact that the information is available only a few months after the death of the individual. NSD's data is the main data source of this project. The information is being transmitted through the employee's payslip to the public services and then each insurer can have access to the information of the people in its portfolio. Each employee statement lists every event that occurred during the month. This level of detail is very interesting and new in group insurance. Finally, internal data from death claims are used in order to produce statistics on deaths that are actually being handled by the insurer.

Lots of data processing was performed on the NSD's data to calculate the medical leave periods in order to deduce from that information the actual observation periods for each individual. Therefore, hypothesis had to be made because some information on the actual policy of each individual was unknown. For instance, the qualifying delay is unknown. After analysis, a qualifying delay of 90 days has been chosen.

Once the data was processed, a study was performed in order to check the quality of the statements regarding death. Therefore, data from the NSD and the Insee have been merged in order to check if all the deaths had been declared and on the contrary if those that had been declared were correct. This analysis showed that 420 deaths were missing and they were added to the study. But, globally the statements were reliable. This merger had limitations because some of the individuals were not merged successfully due to maiden names for example. Moreover, in the data from the Insee, border workers are not present even though there are in the NSD.

In order to better apprehend the data, descriptive statistics were produced. They were made using the internal data and then the processed NSD data. Those statistics allowed to evaluate the ratio of deaths that are consequent to a medical leave in all the death claims. This ratio is of 43% and will be used afterwards, in order to enable the comparison with the regulatory tables which are calculated on all types of deaths. Another observation was that the mean of the age of people who died consequently to a medical leave was higher than the one of people that died without a medical leave. The composition of the portfolio was also studied. Thus, the portfolio includes slightly more men than women and a high majority of non-executives is represented. When evaluating deaths, the high mortality of men is striking as well as the one of non-executives.

On the other hand, an artificial and simplified model was made in order to observe the transfers between the different states in a population defined as behaving like the average. The different states and factors to switch states have been defined carefully according to other studies. This allowed to see that after 10 years of running this scenario, the ratios between the two types of deaths were not those that are observed today in the real portfolio but that they tend to those.

Different models were used because they each have different qualities and flaws. This multiplication of approaches also allowed to obtain reliable results and to confront them in case of divergence.

Raw mortality rates were calculated, as a first approach, to evaluate the shape of the data and to see if the calculated mortality rates are coherent with the other published tables. The results were smoothed using P-splines in order to obtain a usable curve.

As expected, for extreme ages, the fluctuations are important because the exposure is lower and therefore the uncertainty is higher. The smoothed mortality rates obtained are increasing with age, but are quit lower than the regulatory ones.

Moreover, by segmenting the population by socio-professional category and gender, important differences appear between the different groups. The higher mortality rate of men is very visible as well as the one of wor-

kers. A particularly interesting result is the fact that the mortality rate of executive men is slightly higher than the one of female workers even though these workers have the highest mortality rate amongst the women and executive men have the lowest mortality rate amongst men. This shows the huge contrast that exists between men and women but also between the different socio-professional categories.

Then the Kaplan-Meier model was implemented. It is interesting because it is a non-parametric model and therefore allows to not make any hypothesis on the data. Moreover, it takes into account the censorship by giving more weight to whole observations.

The results were segmented right away by gender because of the previous observations. And those previous observations are confirmed here as well. For instance, there is a 6 year gap between the moment men have a mortality rate of 0,5% and the moment the women do. But, this method does not seem robust enough in order to allow a more precise segmentation, moreover other variables seem interesting to study.

Then a Poisson GLM was implemented. The advantage of this model is that it allows to quickly identify all the significant criteria for mortality. The Poisson's law was used because it is the usual law that is used for mortality.

First, every explanatory variable was added to the model. The models were divided into two categories, on one side those using the age variable and therefore a unique transition factor between each age, and on the other side those using age groups and therefore a specific factor for each group. A validation of the model and its adequacy to the data were performed using in particular residuals analysis.

Finally, the model with age groups in which the gender, age group, socio-professional category and the region with some aggregations of modalities is the best one according to the validation process. However, this model is lacking precision because the age is a very significant factor, moreover results by age are desirable. Furthermore, a positioning in regards to the regulatory tables is looked for. Therefore, transition factors are looked for.

This is why a Poisson model using regulatory tables, which are the TH 00-02 and TF 00-02 tables, in the offset was implemented. This model provides mortality tables based on the regulatory ones as well as transition factors.

In the end, the significant variables are the socio-professional category, the region and the crossing between the age and the socio-professional category. The gender and the age are not significant because they are already taken into account in the regulatory tables and the evolution according to these variables is the same in the studied data and in the regulatory tables. The disparities between gender and socio-professional category are still very visible. Moreover, the regions were grouped in 3 different groups with different levels of mortality. The residuals allowed to validate the model.

A comparison with the regulatory tables by adding arbitrarily the ratio of deaths that are consequent to a medical leave was performed. This showed that the portfolio's mortality is lower than the one of the french population. The workers have the highest mortality rate. The employees' mortality is similar but slightly lower. The intermediate professions' mortality is very low at young ages but similar to the workers' one at high ages. The executives have the lowest mortality rate.

Nevertheless, this type of model requires strong assumptions like assuming that the data follows a Poisson law. Moreover, every variable is processed with the other ones in the GLM, it is therefore quite hard to observe the effects of the different variables.

A last method was used. Here again, it is a positioning method. This method is based on the SMR ratio. It is the ratio between the number of observed deaths and the number of expected deaths if the population had the same behavior as in the regulatory tables. There are several benefits to this method. The variables can be studied one by one, independently from one and another and in the desired order. The effect of each criteria is neutralised at each step. This method is therefore iterative. The residuals will be calculated, among other indicators, in order to study the quality of the final adjustment. The studied variables are those that were

kept in the previous models. The order of the variables is based on the previously observed significance of the variables.

Therefore, the first studied criteria is the gender because, as seen throughout this study, the higher rate of men's mortality is very prominent. The two obtained coefficients are very low but similar. This indicates that the portfolio's mortality is quite lower than the one of the french population from 2000 to 2002 and that the regulatory tables are well configured in terms of the gender.

Then, the age groups were studied. Here again, the variations between generations are not very important which allows to deduce that the age is well taken into account in the regulatory tables. Moreover, these observations on the first two studied variables are consistent with the ones from the previous model.

Finally, the socio-professional category turned out to be a significant criteria in every model, this is why this is the next studied variable. The effect is very visible and the disparities between the different categories are very deep. The mortality rate increases in the following order : executives, intermediate professions, employees and finally workers. Workers have the highest mortality rate. The region was also studied but was not kept after the analysis of the residuals.

In order to make sure that the results are reliable, those from the two best methods are confronted. The observed differences are very low, with less than 16% of difference in absolute value. Therefore, the results are reliable and right. The final method, using the SMR ratio, is the one that is kept, because the coefficients are easily usable in a pricing software.

Finally, a comparison of the obtained mortality tables with those used in the current pricing software was performed in order to study the differences. The proportion of deaths that are consequent to a medical leave was added arbitrarily in order to be able to compare the different tables. However, this is a drawback of this study because this approximation seems too inaccurate. It is worth noticing that the current pricing software is based on tables published by the Insee from 96-98 and not the TH 00-02 and TF 00-02 like it is the case in this project.

The most important differences are located on the over 60 years old age group and that is the case for each gender and socio-professional category. For executives, the disparities are very small. The differences are bigger for workers, with approximately 30% of difference at the highest. For employees and intermediate professions the analysis is more tricky because the current pricing software gathers these two categories as one. Therefore, the mortality rate of employees seems very underestimated currently even though the one of intermediate professions is in adequacy to the current pricing results.

Furthermore, the current pricing software also has pricing groups which are based upon the line of business of the company but this criteria was not significant in this study so it was not kept. Indeed, this criteria was redundant to the socio-professional category which already captures these class effects.

To conclude, the portfolio's mortality rate is much lower to the one of the general population because the unemployed have the highest mortality rate. Therefore, the regulatory tables are not adequate to this portfolio and overestimate the mortality. Furthermore, some variables that are very discriminant like the socio-professional category are not taken into account in the regulatory tables. Therefore, this study allowed to put light upon the criteria that have a significant impact on the mortality of employed people and those which are not, and to built a table as well as position factors which are adequate to the real mortality of the portfolio. However, only deaths without a medical leave were studied here, therefore this study is complementary to the one that will be performed later on deaths that are consequent to a medical leave which will provide a global vision of the portfolio's mortality and will allow the completion of this risk's pricing.

Introduction

En prévoyance collective, la garantie décès est généralement tarifiée à l'aide des tables réglementaires recommandées par le code des assurances, à savoir les tables TH 00-02 et TF 00-02. Ces tables ont été construites par l'Insee en utilisant des données concernant la population française globale, observée entre les années 2000 et 2002. Aucune nouvelle table officielle n'a été publiée depuis l'arrêté du 29 décembre 2005¹. Or, avec les avancées technologiques et médicales, l'espérance de vie tend à s'accroître en France, en outre, l'écart existant entre les hommes et les femmes se réduit. En effet, entre 2000 et 2018, l'espérance de vie des femmes a augmenté de 2,6 années et celle des hommes de 4,3 années. Cependant, depuis 2014 l'espérance de vie semble stagner². De surcroît, en prévoyance collective la population assurée n'a pas les mêmes caractéristiques que la population française globale. En effet, l'âge, des personnes assurées, est compris principalement dans un intervalle dépendant de l'âge légal de travail et de l'âge de départ à la retraite. Ainsi, la connaissance de la mortalité en dehors de cet intervalle est d'une importance moindre. D'autre part, la population assurée est alors salariée, par conséquent, les individus reçoivent un revenu régulier. Or, dans la population française générale, des études démontrent que la population avec la plus forte mortalité est celle des inactifs. Par conséquent, il est rationnel de penser que les tables réglementaires ne sont pas adaptées à la tarification du risque décès d'une population salariée, car elles surestiment ce risque. Le risque décès étant volatile et onéreux pour l'assureur, il est donc important de bien le tarifer.

Par ailleurs, depuis la loi Evin de 2001³, le maintien des garanties décès aux personnes en arrêt de travail est imposé. Et cela, malgré le fait que ces individus ne paient plus de cotisations, car ils ne font plus partie de la population active étant donné qu'ils perçoivent des prestations arrêt de travail. Cette loi impose aux assureurs la création d'une provision spécifique, appelée provision pour maintien de la garantie décès, pour couvrir ce risque à tout moment.

De ce fait, cela complexifie le risque décès. Ce dernier se décompose alors en deux parties : les décès sans arrêt de travail, appelés décès directs, et les décès faisant suite à un arrêt de travail. Dans ces deux cas de figures les causes de décès ne semblent pas les mêmes. Ainsi, il semble raisonnable de distinguer ces deux types de décès dans la tarification du risque. Cela est néanmoins difficile à réaliser car peu de données sont disponibles. En effet, le décès est un événement rare. De plus, en prévoyance collective, peu de données sur la population assurée sont connues de l'assureur. Cependant, la mise en place de la Déclaration Sociale Nominative (DSN) depuis 2017 a permis aux assureurs de récolter des données sur la situation de chaque individu présent dans la population assurée de manière simplifiée et fiable. Le risque peut donc être mieux évalué et maîtrisé afin d'adapter la tarification au mieux.

L'objectif de ce mémoire est de comparer le comportement du portefeuille de Groupama Gan Vie par rapport aux lois de référence mais également de déterminer les critères explicatifs des décès qui ne font pas suite à un arrêt de travail. Pour cela, plusieurs approches vont être utilisées. L'utilité finale étant la tarification, l'approche par positionnement par rapport aux lois de référence va être privilégiée, plutôt que la construction d'une loi de

1. <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000000636581>

2. <https://www.spac-actuaire.fr/lexique/esperance-de-vie/>

3. https://www.legifrance.gouv.fr/loda/article_lc/LEGIARTI000006756630/2001-07-18/#LEGIARTI000006756630

mortalité.

Dans un premier temps, une présentation du risque décès et du cadre juridique est réalisée. Les différentes garanties y sont présentées ainsi que les principales lois ayant fait évoluer la réglementation de ce risque.

Dans un second temps, le modèle économique et la tarification seront exposés. Des précisions sur la complexité de la prise en compte du décès suite à arrêt de travail dans les comptes des assureurs seront développées. Les différentes trajectoires d'un individu pouvant amener un assureur à verser des prestations pour décès seront détaillées. Le tarif, son calcul et son contenu seront analysés. L'outil de tarification actuel et son fonctionnement seront également présentés.

Puis, les données disponibles pour réalisées l'étude seront décrites avec les hypothèses prises et les traitements réalisés sur les données. Une étude permettant de vérifier la qualité des données est également réalisée.

Ensuite, des statistiques sur les données sont exposées afin de mieux cerner la population étudiée. Un modèle simple permettant de voir l'évolution d'une population dans les différents états possibles est exposé.

Finalement, une partie sur la modélisation, avec la présentation des différents modèles et des résultats obtenus, sera développée. Les résultats obtenus seront également comparés afin de sélectionner la meilleure méthode.

Pour finir, ces résultats seront également comparés avec les lois utilisées par l'outil de tarification actuel afin de mesurer les impacts.

Remerciements

Tout d'abord, je tiens à remercier particulièrement ma tutrice Marie DALONGEVILLE pour son soutien, ses conseils et le temps qu'elle m'a dédié. Je remercie également Mathilde LOMBARD, ma responsable, ainsi que Vincent LAUDOU, responsable de la Direction Technique Prévoyance et Santé Collectives, pour m'avoir apporté un suivi régulier, des conseils précieux et m'avoir aidé à appréhender le sujet de ce mémoire.

Je souhaite également remercier toute l'équipe de la Direction Technique pour leurs conseils et leur accueil chaleureux, et plus particulièrement Adel KOBTAN.

Enfin, je souhaite remercier mon tuteur pédagogique, Guillaume BIESSY, pour ses recommandations et son suivi. J'adresse également mes remerciements à l'ensemble du corps professoral de l'ISUP.

Chapitre 1

Prévoyance collective

1.1 Généralités

L'article L-140-1¹ du code des assurances définit le contrat d'assurance de groupe comme étant « un contrat souscrit par une personne morale ou un chef d'entreprise, en vue de l'adhésion d'un ensemble de personnes » qui doivent avoir « un lien de même nature avec le souscripteur ». Les produits de prévoyance sont destinés à couvrir les salariés, anciens salariés et leurs ayants droit contre les aléas de la vie portant atteinte à l'intégrité physique de la personne (maladie, accident), la maternité, l'incapacité de travail temporaire, l'invalidité et le décès. Ces garanties permettent de compléter les prestations versées par la Sécurité Sociale. Plusieurs types d'entreprises peuvent délivrer un contrat de prévoyance collective :

- Une société d'assurance régie par le Code des Assurances
- Un institut de prévoyance régi par le Code de la Sécurité Sociale
- Une mutuelle régie par le Code des mutuelles

Ces contrats sont donc tripartites :

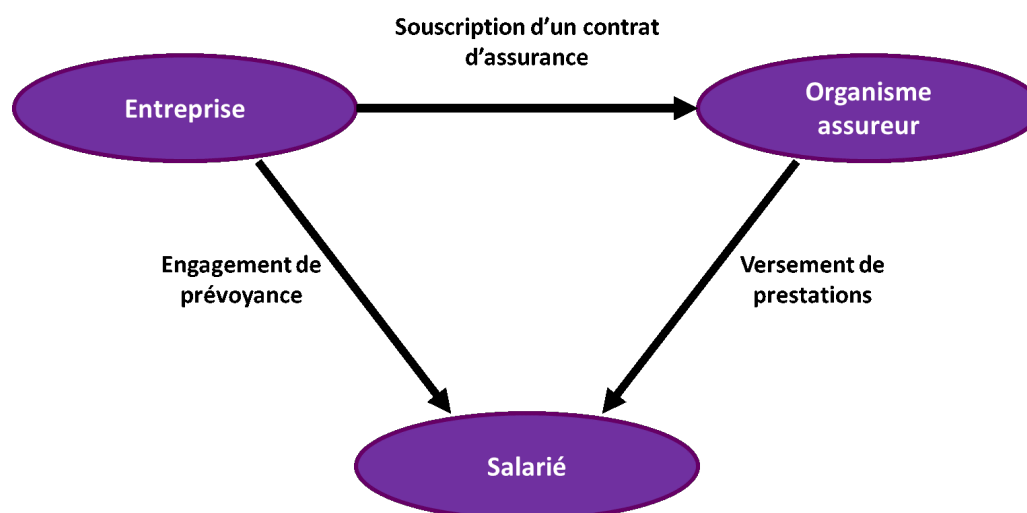


FIGURE 1.1 – Schéma de la gestion d'un contrat d'assurance de groupe

Le caractère collectif du contrat impose que l'ensemble des salariés ou une partie des salariés appartenant à une même catégorie objective, doit être couvert par les mêmes garanties. Il existe 5 catégories objectives qui

1. https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000006793485

sont les suivantes :

- Statut cadre ou non-cadre
- Catégories par tranches de rémunération AGIRC-ARRCO
- Catégories et classification CCN et accords professionnels ou interprofessionnels
- Catégories définies par rapport au niveau de responsabilité, type de fonction ou autonomie au regard des CCN ou accords professionnels ou interprofessionnels
- Usages en vigueur dans la profession.

La plupart des régimes d'entreprise sont des régimes à adhésion obligatoire toutefois des garanties optionnelles peuvent être proposées afin d'améliorer la couverture du salarié si celui-ci le désire.

La mise en place d'un régime de prévoyance collective peut se faire soit par un accord collectif, soit par référendum, soit par décision unilatérale du chef d'entreprise. A noter qu'en cas de décision unilatérale, cela ne permet pas d'imposer les cotisations aux salariés.

1.2 Les garanties décès

Les différentes garanties décès ont pour but de palier la perte de revenu du ménage due au décès de l'affilié. On distingue trois types de garanties pour le risque décès :

- Les capitaux
- Les rentes (conjoint, éducation)
- Les allocations obsèques

1.2.1 Les prestations en capital

Le montant versé par la Sécurité Sociale est forfaitaire et fixé par décret, il est revalorisé chaque année depuis la loi de financement de 2015. Ce montant était de 3 539 € en avril 2022. Ce capital est cependant soumis à des conditions d'éligibilité. En effet, l'assuré devait soit exercer une activité salariée moins de 3 mois avant son décès, soit être titulaire d'une pension d'invalidité ou d'une rente allouée en vertu de la législation sur les accidents du travail et maladies professionnelles, soit bénéficier du maintien de ses droits à l'assurance décès au moment de son décès.

Ce montant du régime de base étant très faible, il est bénéfique d'avoir recours à une assurance complémentaire afin d'obtenir un capital plus conséquent. Ce « capital décès de base » est délivré en cas de décès de l'affilié. Il est calculé en fonction du salaire de l'affilié, de sa situation familiale ... Ce montant est versé au(x) ayant(s) droit sous la forme d'un versement unique.

Des garanties annexes peuvent être souscrites :

- Décès accidentel : majoration du capital de base en cas de décès dans les 12 mois suivant l'accident (en général doublement du capital)
- Capital double-effet : en cas de décès de l'affilié et du décès simultané ou postérieur du conjoint lorsque celui-ci a encore des enfants à charge et qu'il n'a pas atteint l'âge de 65 ans

- En cas d'IAD : le capital de base peut être délivré de manière anticipée en cas d'IAD (Invalidité Absolue et Définitive) de l'affilié à la demande de celui-ci, la garantie décès prend alors fin.
- Décès du conjoint avant l'assuré (« Pré-décès du conjoint ») : versement d'un capital à l'assuré

1.2.2 Les prestations en rente

Il s'agit de paiements récurrents (mensuels, trimestriels, semestriels ou annuels) pouvant être temporaires ou viagers. Ces versements peuvent remplacer ou venir en complément des versements en capitaux énoncés ci-dessus. On distingue deux types de rentes en fonction du bénéficiaire :

- La rente conjoint : elle peut être temporaire (en attente du versement de la pension de réversion des régimes de retraite obligatoires) ou viagère (versée jusqu'au décès du conjoint, afin de compenser le fait que le salarié décédé n'a pas cotisé jusqu'à sa retraite et la pension de réversion sera donc amoindrie). Ces deux types de rentes peuvent être cumulées. Elles sont calculées sur le dernier salaire de l'assuré décédé.
- La rente éducation : une rente est versée à chaque enfant à charge de l'assuré au moment de son décès. Elle est temporaire, avec, dans la plupart des contrats, un arrêt des versements à 18 ans. Cependant, elle peut être prolongée jusqu'à 26 ans maximum, en général, en cas de poursuite des études. Cette rente peut être constante ou évoluer en fonction de l'âge de l'enfant (on dit alors qu'elle est par palier). Cette rente est également calculée en fonction du dernier salaire de l'assuré décédé. Si l'enfant est handicapé alors cette rente devient viagère.

1.2.3 Les allocations obsèques

Il s'agit d'un montant forfaitaire destiné à couvrir les frais liés aux obsèques. Ce versement est effectué le plus souvent rapidement après la déclaration du décès car son but est de couvrir le règlement des funérailles. Ce capital est versé à la personne ayant réglée les frais d'obsèques.

Synthèse des prestations garanties pour le risque décès	
Capitaux	Capital décès
	Capital accidentel (en cas de décès suite à un accident)
	Capital double effet (si décès simultané ou postérieur du conjoint)
Rentés	Rente conjoint
	Rente éducation (si l'assuré a au moins un enfant à charge, une rente par enfant)
Allocations	Allocations obsèques
Autres	Frais d'honoraires
	Capital décès en cas de décès du conjoint avant l'affilié (option)
	Capital de revalorisation (si le décès fait suite à un arrêt de travail)
	Capital anticipé en cas d'IAD (met fin à la garantie)

FIGURE 1.2 – Tableau récapitulatif des prestations décès

1.3 Cadre juridique

Tous les engagements des assureurs doivent être en accord avec les différents textes juridiques. Ces derniers visent à protéger les affiliés en améliorant leurs droits et en renforçant la capacité des assureurs à respecter leurs

engagements. Plusieurs lois ont été publiées au fil des années faisant évoluer les garanties décès avec notamment l'extension de la garantie aux personnes en arrêt de travail.

1.3.1 Loi du 31 décembre 1989

L'article 1 de cette loi indique que l'auto-assurance est interdite en prévoyance, ainsi le risque est transmis à un organisme assureur afin d'assurer le versement des prestations même en cas de défaillance de l'entreprise contractante.

L'article 2, quant à lui, précise le caractère collectif du contrat. Ainsi, l'assureur a la possibilité de renoncer à assurer le groupe mais s'il s'y engage alors il doit couvrir tous les salariés sans exclusions individuelles possibles même si certains ayants droit ont des facteurs de risque plus important que d'autres.

L'article 7, quant à lui, impose qu'en cas de résiliation, le paiement des prestations en cours (notamment les rentes) doit être maintenu au moins au niveau atteint lors de la résiliation. Cet engagement doit être couvert à tout moment par des provisions mathématiques.

1.3.2 Loi du 8 août 1994

Tout d'abord, cette loi complète l'article 7 précédemment cité en imposant la poursuite de la revalorisation des prestations en cours de service en cas de résiliation du contrat.

Ensuite, elle indique qu'en cas de changement d'assureur le maintien de la garantie décès pour les personnes en incapacité ou en invalidité doit être prévu, il est donc soit pris en charge par l'ancien assureur, soit il est transféré vers le nouvel assureur.

1.3.3 Loi du 17 juillet 2001

Cette loi vient éclaircir le dernier point cité de la loi précédente, en imposant le maintien des garanties décès aux personnes en arrêt de travail et cela même en cas de résiliation ou de non-renouvellement du contrat. Cet engagement doit être couvert à tout moment dans les comptes de l'assureur par une provision spécifique appelée « provision de maintien de la garantie décès » ou PMGD.

Ainsi, en cas d'arrêt de travail l'assuré est exonéré du paiement de ses cotisations mais sa garantie décès est maintenue. En revanche, lors de son arrêt de travail l'assuré ne voit pas son salaire évoluer alors que le coût de la vie évolue, c'est pourquoi il est normal de refléter cela en revalorisant le capital décès lorsque l'assuré vient à décéder. Cette revalorisation du salaire de base est calculée par rapport à la valeur de l'unité de référence en vigueur à la date du décès et celle en vigueur à la date de l'arrêt de travail. Ce type de garantie est donc à traiter de manière différente que le décès sans arrêt de travail car les risques ne sont pas les mêmes. C'est ce que l'on nomme le décès suite à arrêt de travail ou suite à exonération.

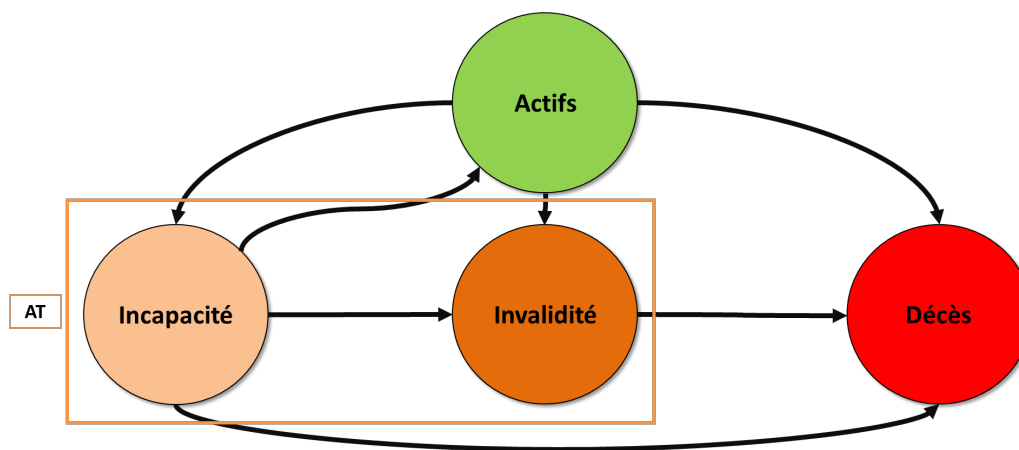


FIGURE 1.3 – Représentation des différents états et transitions possibles

1.3.4 Arrêté du 20 décembre 2005

L'article 5 de cet arrêté impose l'utilisation des tables de mortalité homologuées TH 00-02 et TF 00-02 pour les contrats autres que les rentes viagères. On les appelle les tables réglementaires. Ces dernières ont été construites à l'aide des données de l'Insee issues d'observations réalisées entre 2000 et 2002 sur la population française dans sa globalité. La table TH 00-02 est basée sur la population masculine alors que la table TF 00-02 est basée sur la population féminine. Cette distinction est nécessaire dû à l'écart important entre la mortalité des femmes et des hommes.

Ainsi, pour la construction de son tarif l'assureur a le choix entre les tables réglementaires et une table de mortalité homologuée par un actuaire certifié, appelée table d'expérience (article A132-18 du Code des assurances²). Cependant, en assurance collective dans le cas de contrats décès résiliables annuellement cette homologation n'est pas nécessaire. L'avantage des tables d'expérience, est qu'elles ne sont pas basées sur la population nationale mais sur celle du portefeuille de l'assureur et reflètent par conséquent, mieux le risque réellement porté par l'assureur.

1.3.5 Accord National Interprofessionnel du 11 janvier 2008

L'article 14 de cet accord prévoit la portabilité des garanties prévoyance et frais de santé en cas de rupture du contrat de travail sous réserve que certaines conditions soient réunies. La loi de sécurisation de l'emploi du 14 juin 2013 a modifié ces conditions. Ainsi, l'ensemble des entreprises est concerné par ce droit de conservation des bénéficiaires des garanties prévues au contrat collectif en vigueur dans l'ancienne entreprise. Ce maintien est effectif pour une période égale à la durée des derniers contrats de travail s'ils sont consécutifs chez le même employeur et ne peut tout de même excéder 12 mois. Les conditions requises sont le non-licenciement pour faute lourde, avoir le droit à l'assurance chômage et avoir ouvert des droits à couverture complémentaire chez le dernier employeur. Ce droit est accordé dès le 1^{er} jour de travail. L'ancien salarié a la possibilité de renoncer à ce maintien des garanties sinon le maintien des garanties s'achève lors de la signature d'un nouveau contrat de travail ou au maximum à l'issue de la période de maintien des droits.

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

Chapitre 2

Modèle économique du décès et tarification

2.1 Modèle économique

Dans cette partie, la complexité des comptes de résultats décès, due au décès faisant suite à arrêt de travail et au lien existant entre l'arrêt et le décès, est analysée. Puis, tous les parcours pouvant mener l'assureur à verser des prestations décès sont répertoriés.

2.1.1 Complexité du décès suite à arrêt de travail

Dans cette partie, les comptes de résultats concernant le décès suite à exonération sont expliqués.

Tout d'abord, les comptes de résultats des assureurs peuvent être effectués par exercice comptable ou par exercice de survenance. Dans le premier cas, tous les sinistres déclarés l'année N sont comptabilisés dans les comptes de résultats de l'année N , cela ne tient donc pas compte de l'année de survenance. Dans le second cas, tous les sinistres survenus l'année N sont comptabilisés dans les comptes de résultats de l'année N , même si l'année de déclaration est ultérieure.

Le cas du décès suite à arrêt de travail représente un cas particulier. En effet, deux sinistres ont lieu, l'arrêt de travail et le décès. De ce fait, deux points de vue existent, les assureurs peuvent déclarer le décès en considérant l'arrêt de travail comme le fait générateur du sinistre, ou alors en considérant le décès comme le fait générateur du sinistre.

Par conséquent, dans ce dernier cas, le décès suite à exonération n'est pas traité différemment dans les comptes de résultats par rapport au décès sans arrêt de travail. Le décès n'est pas rattaché à l'évènement déclencheur qui est, ici, l'arrêt de travail.

Or cela n'est pas souhaitable car la loi oblige les assureurs à constituer la provision pour maintien de la garantie décès (PMGD) en cas d'arrêt de travail. Cette provision est donc mise en place lors de la survenance de l'arrêt de travail, comme l'indique la loi. Ainsi au moment de l'arrêt de travail une provision mathématique, pour les prestations incapacité/invalidé (PM AT), est créée ainsi que la provision pour maintien de la garantie décès. Cette dernière étant généralement calculée comme un pourcentage de la provision mathématique pour arrêt de travail.

Notons N l'année de cet arrêt de travail. Si le décès n'est pas survenu l'année N et que l'assuré se maintient en arrêt, alors ces provisions sont maintenues (avec évolution de leurs montants). Si le décès survient l'année $N + k$ ($k > 0$, entier) et que l'assuré était encore en arrêt au moment du décès, alors le décès est rattaché à l'année $N + k$ (année de survenance du décès) et c'est alors la PMGD des assurés tombés en arrêt de travail

l'année $N + k$ qui sert à payer les prestations décès de cet assuré. Ainsi, il y a une mutualisation de la PMGD. Le sinistre n'est donc pas rattaché à la survenance de la PMGD émise pour couvrir ce même sinistre.

De cette manière les décès déclarés comme survenus l'année $N + k$ se décomposent de la façon suivante :

$$\begin{aligned} \text{Décès}_{N+k} &= \text{Décès directs}_{N+k} + \text{Décès suite à AT}_{N+k} + \text{Décès des personnes en AT au 31/12/N+k-1} \\ &= \text{Décès directs}_{N+k} + \text{Décès suite à AT}_{N+k} + \text{Décès suite à AT}_{N+k-1} + \text{Décès suite à AT}_{N+k-2} + \dots \end{aligned}$$

Où l'année en indice représente l'année de survenance du fait générateur.

Or pour que le ratio S/P ne soit pas impacté par cette vision, cela nécessite que le portefeuille de l'assureur reste stable dans le temps.

Le taux d'incidence en arrêt de travail, le taux de maintien en arrêt de travail et le taux de décès suite à arrêt de travail sont supposés constants par année de survenance et vision comptable.

Alors, en cas de croissance du portefeuille entre les années N et $N + k$, il y a augmentation dans le temps du nombre d'arrêts de travail survenus. Il y a donc également une augmentation du nombre d'individus en maintien. Or, la PMGD, dans ce cas, a été calculée l'année $N + k$ par rapport au portefeuille $N + k$. Ainsi, les décès suite à exonération, dont l'arrêt de travail est survenu avant l'année $N + k$, sont sous représentés par rapport à la PMGD calculée pour couvrir les décès suite à arrêt de travail en $N + k$. Le S/P sera donc déformé, et un boni de charge apparaîtra sur l'année de survenance de l'arrêt de travail car la PMGD calculée apparaît comme trop conséquente. La charge du décès est alors portée par l'année $N + k$.

Bien évidemment, en cas de décroissance du portefeuille le phénomène inverse se produit. Un mali de charge apparaîtra sur l'année de survenance de l'arrêt de travail.

Par ailleurs, dans l'autre vision, le décès est rattaché à l'année de survenance de l'arrêt de travail. Cette vision représente le cadre idéal car le décès suite à exonération est bien rattaché au fait générateur qui est ici l'arrêt de travail. Dès lors, le décès est bien rattaché à la PMGD émise pour ce risque. Ainsi, le ratio S/P n'est pas déformé par une potentielle croissance ou décroissance du portefeuille.

Prenons l'exemple d'un salarié qui tombe en arrêt de travail l'année N et qui se maintient en arrêt de travail jusqu'à l'année $N + 4$, année où il décède.

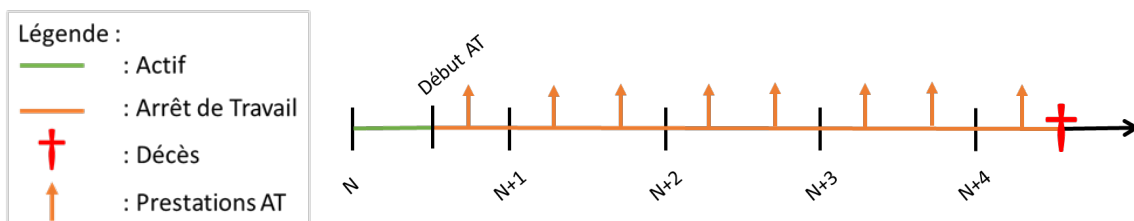


FIGURE 2.1 – Exemple d'un décès suite à arrêt de travail

Dans le cas où le décès suite à exonération est traité de la même manière que le décès direct, représenté par la figure de gauche ci-dessous, le décès est comptabilisé dans les comptes de l'année $N + 4$ avec pour année de survenance $N + 4$. Le décès suite à exonération est donc dissocié de l'arrêt de travail.

Dans l'autre cas, le décès est rattaché à la survenance de l'arrêt de travail. Ainsi, le décès est comptabilisé l'année $N + 4$ pour l'année de survenance N .

Décès	Année de survénance				
	N	N+1	N+2	N+3	N+4
Année comptable N					Décès + PMGD
Année comptable N+1		Prestations + PM			
Année comptable N+2					
Année comptable N+3					
Année comptable N+4					

AT	Année de survénance				
	N	N+1	N+2	N+3	N+4
Année comptable N	AT				
Année comptable N+1	AT	Prestations + PM			
Année comptable N+2	AT				
Année comptable N+3	AT				
Année comptable N+4	Décès + PMGD				

FIGURE 2.2 – Représentation des 2 différentes visions, dans le cadre d'un exemple, sur les triangles de charges

Cela permet donc de voir que dans le cas de droite, le triangle de charges comprend la vision complète du déroulement de l'arrêt de travail, avec la survénance, le maintien et le décès. En revanche, la vision de gauche ne permet pas de connaître l'historique du sinistre.

2.1.2 Cartographie des types de décès

Dans cette partie, les différents cas de figures concernant un sinistre décès, où l'assureur est amené à verser une prestation, vont être répertoriés. A noter que les cas présentés peuvent être précédés d'autres arrêts de travail avec rétablissements.

La légende utilisée est la même que celle de la partie précédente (fig. 2.1).

Les différents parcours menant au versement de prestations décès sont donc :

- Décès seul : salarié actif qui décède, ce cas est pris en compte dans cette étude car il s'agit bien d'un cas de décès direct.



FIGURE 2.3 – Schéma décès direct seul

- Décès après arrêt de travail :
 - Décès avant franchise : salarié actif qui tombe en arrêt de travail et qui décède pendant la période de franchise. L'assuré continue de cotiser ainsi son décès est comptabilisé comme un décès direct.

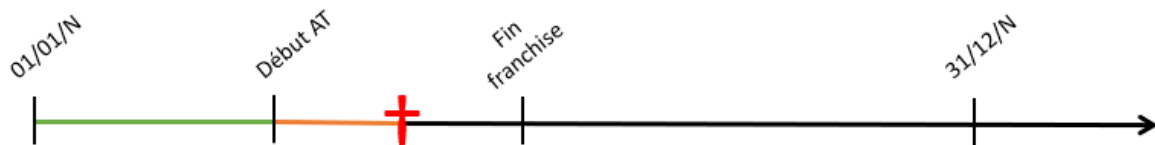


FIGURE 2.4 – Schéma décès direct pendant la franchise

- Décès après franchise et contrat en vigueur (individu dans les effectifs) : salarié actif qui tombe en arrêt de travail et qui décède après la période de franchise (la personne reçoit donc des prestations arrêt de travail de la part de l'assureur)

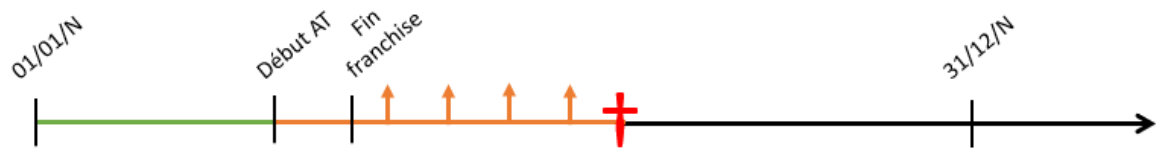


FIGURE 2.5 – Schéma décès suite à exonération

- Décès après franchise et contrat résilié ou individu sorti des effectifs au moment du décès : personne active qui tombe en arrêt de travail (la personne reçoit donc des prestations arrêt de travail de la part de l'assureur) et qui décède postérieurement à sa sortie des effectifs ou à la résiliation du contrat

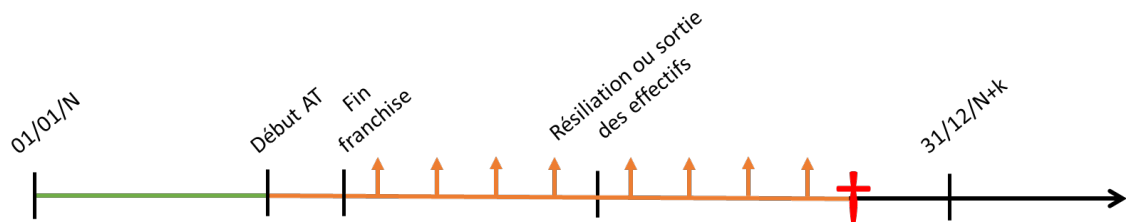


FIGURE 2.6 – Schéma décès suite à exonération après la sortie des effectifs

Les deux derniers cas correspondent donc à des décès suite à arrêt de travail et ne seront dès lors pas pris en compte dans cette étude.

2.2 Modèle tarifaire

Dans cette section, la tarification de la garantie décès est décrite. Le contenu des primes décès et arrêt de travail est analysé puis, le calcul explicite de la prime décès est illustré. Enfin, l'outil de tarification actuellement en place est décrit.

2.2.1 Contenu de la prime

Ici, le contenu de la prime d'une année comptable est explicité suivant le risque qu'elle couvre. L'arrêt de travail est étudié ici pour faire le lien avec le décès suite à exonération. On rappelle également, qu'une prime est déterminée de manière à équilibrer les engagements de l'assureur (prestations, frais) avec ceux de l'assuré (primes). La loi française interdit de proposer un tarif différent pour une femme et un homme toutes choses égales par ailleurs.

La prime de l'année N couvrant l'arrêt de travail doit permettre de couvrir les prestations des arrêts de travail survenus dans l'année N .

La prime de l'année N couvrant le risque décès se décompose, elle, en 2 parties. Elle doit, de la même manière que la prime couvrant l'arrêt de travail, couvrir les prestations des décès directs survenus dans l'année N . Mais elle doit également couvrir les décès suite à exonération dont l'année de survenance de l'arrêt de travail est l'année N .

Cela démontre donc que la tarification du risque décès est complexe, car elle doit prendre en compte des décès qui auront lieu potentiellement des années plus tard.

2.2.2 Calcul de la prime

Dans cette section, le calcul de la prime d'un contrat individuel puis d'un contrat collectif, couvrant le risque décès est détaillé. Le calcul du tarif d'un contrat collectif peut être segmenté par rapport aux sous-populations présentes dans l'entreprise à assurer mais le tarif final devra lui être unique pour tous les salariés de cette entreprise.

2.2.2.1 Taux technique

Le taux technique utilisé pour la tarification est règlementé par le code des assurances, avec l'article A132-1¹. Ce dernier prévoit une distinction en cas de contrat vie ou non-vie. Le cas considéré dans ce mémoire est le cas d'un contrat vie.

Le taux technique est calculé par rapport au TME (Taux Moyen des Emprunts de l'Etat français) qui est publié chaque mois.

Par conséquent, pour le risque décès, la loi oblige que le taux technique retenu n'excède pas le plus bas des deux taux suivants : 3,5% ou 60% de TME.

2.2.2.2 Tables de mortalité

Ici encore, le code des assurances régit l'utilisation des tables de mortalité, avec l'article A132-18². Il existe deux types de tables :

- Tables homologuées par arrêté du ministère de l'économie et des finances
- Tables d'expériences établies par l'assureur et certifiées par un actuaire indépendant agréé

De plus, deux catégories de tables existent :

- Tables de mortalités instantanées : elles prennent en compte uniquement l'âge de l'individu
- Tables par génération ou prospectives : elles sont bidimensionnelles, non seulement l'âge est pris en compte mais également l'année de naissance de l'individu. Cela permet de prendre en compte l'évolution de la mortalité suivant la génération de l'individu.

La loi précise également qu'en cas de tarification des rentes viagères avec des tables d'expériences, le tarif ne peut être inférieur à celui calculé en utilisant les tables réglementaires adaptés.

Les tables réglementaires actuelles, pour la tarification du risque décès, sont les TH 00-02 et TF 00-02 pour les capitaux et les TGH05 et TGF05 pour les rentes. Ces dernières sont des tables prospectives, car les rentes représentent un engagement à long terme.

2.2.2.3 Prime individuelle

La prime individuelle est calculée en fonction des caractéristiques spécifiques de l'individu à assurer.

1. https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000035514601/
2. https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000035514715/

La prime pure individuelle Π , est égale à :

$$\Pi = \text{Capital sous risque} \times \text{Probabilité de décès} \times \text{Correctif}$$

En effet, cette prime a pour but de couvrir le risque de décès au cours de l'année pour un individu spécifique.

Le coefficient de correction peut, par exemple, prendre en compte le sexe, l'âge, la catégorie socio-professionnelle

...

Le capital sous risque correspond à un capital dans le cas où la garantie prévoit le versement d'un capital ou à des capitaux constitutifs de rente dans le cas où la garantie prévoit le versement d'une ou plusieurs rente(s) (conjoint, éducation).

Pour tarifier les capitaux associés aux garanties annexes, tel que le décès accidentel ou le double effet, le plus souvent une probabilité d'occurrence du risque est utilisée par garantie à la place de la probabilité de décès. Celle-ci correspond à la probabilité que cet événement se réalise et ne prend pas en compte d'autres variables tel que l'âge par exemple. Une correction peut tout de même être mise en place. Par exemple, dans le cas du décès accidentel, le secteur d'activité influe sur la probabilité d'occurrence.

En ce qui concerne les rentes, les capitaux constitutifs doivent être calculés avec un taux d'actualisation i , qui est défini selon les explications vu précédemment (section 2.2.2.1) .

La formule pour un bénéficiaire d'âge x , dans le cas viager, pour 1€ de rente versée annuellement à terme échu est :

$$CC_{RV} = \sum_{k=1}^{\infty} {}_k p_x \times \frac{1}{(1+i)^k} = a_x$$

La somme possède en théorie un nombre infini de termes, mais en réalité cette somme s'arrête pour $k = w - x$ où w correspond à l'âge limite des tables de mortalité.

Et pour le cas d'une rente temporaire versée annuellement pendant n années à terme échu :

$$CC_{RT} = \sum_{k=1}^n {}_k p_x \times \frac{1}{(1+i)^k} = {}_n a_x$$

Le capital sous risque en cas de rente correspond donc au capital constitutif adéquat multiplié par le montant de la rente.

Les rentes peuvent être versées plusieurs fois par an. Un terme de fractionnement est alors à ajouter aux formules des capitaux constitutifs précédentes. Il est égal à $\frac{m-1}{2m}$, où m correspond au nombre de paiements dans l'année.

2.2.2.4 Prime collective

Dans le cas d'un contrat collectif, les données individuelles sont généralement inconnues au moment de la tarification, seules des informations synthétiques tel que l'âge moyen actuariel, la proportion d'hommes ou encore le secteur d'activité sont connus.

L'âge moyen actuariel est utilisé, et non l'âge moyen, car le risque décès n'est pas linéaire mais exponentiel. Ainsi, le quotient de mortalité du groupe est différent de celui du salarié d'âge moyen.

Les primes sont généralement exprimées en pourcentage du salaire de l'assuré dans le cadre d'un contrat collectif.

Pour le capital décès toutes causes, la prime pure en pourcentage du salaire est donc égale à :

$$\Pi = \frac{\text{Capital moyen garanti}}{\text{Salaire moyen}} \times \text{Probabilité de décès} \times \text{Correctif}$$

Ici, un correctif peut être appliqué en fonction de certaines variables comme la catégorie socio-professionnelle (CSP) par exemple. La probabilité de décès est celle correspondant à l'âge moyen actuariel du groupe.

De la même manière, pour les garanties annexes qui sont versées sous forme de capital, la même formule s'applique en remplaçant la probabilité de décès par une valeur fixe ou une probabilité ne dépendant pas de l'âge mais plutôt de la CSP ou du secteur d'activité par exemple.

Le cas des rentes est plus compliqué à tarifier. En effet, les rentes conjoint et éducation seront versées si le salarié vient à décéder, ce qui est calculé à l'aide d'âges moyens, et donc avec peu de précision. Outre cela, la durée de versement de la rente dépend elle de l'ayant-droit, et notamment, de l'âge de ce dernier. Or, ce type d'information est inconnu de l'employeur en général et ne peut donc pas être transmis à l'assureur, même sous forme agrégée tel qu'un âge moyen des conjoints, ou un nombre moyen d'enfants à charge par salarié par exemple. Ces informations ne sont donc pas connues lors de la tarification. Elles seront seulement connues en cas de réalisation du risque pour un individu, afin de permettre le règlement de la prestation.

C'est pourquoi, des hypothèses supplémentaires doivent être prises.

Rappelons également l'expression de la valeur actuelle probable d'1€ de rente viagère à terme échu, avec m paiements dans l'année, $a_x^{(m)}$, qui est définie par :

$$a_x^{(m)} = \sum_{k=1}^{\infty} {}_kE_x + \frac{m-1}{2m} = \sum_{k=1}^{\infty} \frac{l_{x+k}}{l_x} \frac{1}{(1+i)^k} + \frac{m-1}{2m}$$

Où, l_x correspond au nombre de survivants dans une cohorte à l'âge x .

De même, l'expression de la valeur actuelle probable d'1€ de rente temporaire à verser pendant n années, à terme échu, avec m paiements dans l'année, ${}_na_x^{(m)}$ est définie par :

$${}_na_x^{(m)} = \sum_{k=1}^n {}_kE_x + \frac{m-1}{2m} = \sum_{k=1}^n \frac{l_{x+k}}{l_x} \frac{1}{(1+i)^k} + \frac{m-1}{2m}$$

Tout d'abord, le cas de la rente conjoint est explicité.

Si aucune information n'est connue sur les conjoints, leurs sexes et leurs âges sont estimés par rapport aux sexes et aux âges des salariés.

Pour le cas de la rente viagère, la prime pure exprimée en pourcentage du salaire est donc :

$$\Pi_{RCV} = \frac{\text{Rente moyenne}}{\text{Salaire annuel moyen}} \times \left(\%Hommes \times q_x^H \times a_y^{(m)} + \%Femmes \times q_x^F \times a_z^{(m)} \right) \times \%Conjoints$$

où y et z correspondent respectivement à l'âge pris en hypothèse pour les conjoints des hommes et pour les conjoints des femmes. Ils sont définis par rapport à x , l'âge moyen actuariel de la population. Dans la pratique, une différence de 3 ans est généralement retenue en considérant l'homme comme étant le plus âgé des deux.

De même pour le cas de la rente temporaire, la prime pure exprimée en pourcentage du salaire est :

$$\Pi_{RCT} = \frac{\text{Rente moyenne}}{\text{Salaire annuel moyen}} \times \left(\%Hommes \times q_x^H \times {}_{r-y}a_y^{(m)} + \%Femmes \times q_x^F \times {}_{r-z}a_z^{(m)} \right) \times \%Conjoints$$

où r correspond à l'âge légal de départ à la retraite.

A présent, intéressons-nous à la rente éducation. A noter qu'ici, des paliers sont souvent mis en place, ce qui

complexifie encore les calculs. Des hypothèses concernant le nombre d'enfants à charge ainsi que leurs âges sont nécessaires si elles ne sont pas transmises à l'assureur. Dans la pratique, les âges des enfants sont approximés en considérant x , mais peuvent également dépendre de la CSP par exemple du parent. Sous ces hypothèses, la prime pure vaut :

$$\Pi_{RE} = \frac{\text{Rente moyenne}}{\text{Salaire annuel moyen}} \times (\%Hommes \times q_x^H + \%Femmes \times q_x^F) \times_n a_z^{(m)} \times \text{nombre moyen d'enfants à charge}$$

où z correspond à la moyenne d'âge des enfants et n correspond au nombre d'années restant jusqu'à la fin du paiement de la rente ou du palier.

A noter qu'en cas de tarification avec plusieurs paliers, la valeur actuelle probable sur les autres paliers est alors différée.

2.2.3 Audit de l'outil actuel

Dans cette partie, l'outil de tarification actuellement en place est présenté. Cet outil est en place depuis plusieurs années et n'est donc très certainement plus adapté à la mortalité actuelle. Le cas d'une prestation versée sous forme de capital est utilisé afin d'illustrer le fonctionnement de l'outil.

Premièrement, la prime est calculée en fonction de la situation familiale en utilisant l'approche "fréquence x coût". Les informations nécessaires à l'établissement d'un tarif décès pour un capital en cas de décès toutes causes par exemple, sont assez classiques. En effet, afin d'établir la prime par tranche de salaire, la situation familiale, la prestation exprimée en pourcentage du salaire par tranche et par situation de famille, la masse salariale de la situation de famille ainsi que la probabilité de décès sont nécessaires. Cette dernière dépend de plusieurs critères comme le sexe, l'âge actuariel qui correspond ici à l'âge moyen décalé en fonction de la CSP et du sexe, la CSP ainsi que la classe tarifaire.

L'objet de l'étude n'étant pas le calcul du tarif, ce dernier ne sera pas plus détaillé. En revanche, les probabilités de décès sont analysées plus en détail.

Les tables de référence utilisées sont celles publiées par l'Insee et calculées sur la population française de 1996 à 1998. Ainsi, ces tables sont encore plus anciennes que les tables réglementaires en vigueur, à savoir les TH 00-02 et TF 00-02 calculées sur la population française de 2000 à 2002. Ces tables sont abattues, dans un premier temps, en fonction du sexe et de la CSP. Ces coefficients ayant été définis à partir de différentes études de l'Insee. Les catégories socio-professionnelles définies sont les cadres, les ETAM (employés, techniciens et agents de maîtrise), les ouvriers et les non-cadres. A noter que les âges vont de 20 à 70 ans.

A cela s'ajoute un abattement général de 10% qui a été mis en place afin d'ajuster les tarifs. En effet, une étude a été réalisée afin de mettre à jour les tables de mortalité abattues qui dataient de plusieurs années et n'étaient donc plus adaptées au risque actuel.

Enfin, des ajustements par classe de risque sont effectués. Il existe 3 classes de risque qui sont déterminées en fonction du code NAF de l'entreprise, c'est-à-dire en fonction du secteur d'activité. La classe 1 étant la moins risqué, un coefficient de 1 y est appliqué. Ce dernier évolue pour les autres classes tarifaires.

Par conséquent, cet outil ne distingue pas les types de décès. De plus, les lois utilisées bien qu'abattues datent de plusieurs années et ne sont plus forcément adaptées au risque actuel. C'est pourquoi, cette étude va permettre de mettre à jour ces normes tout en ajoutant la distinction entre les différents types de décès.

Des études sur le portefeuille étaient difficilement réalisables auparavant car les données individuelles ne pouvaient être connues. Mais, cela est maintenant possible.

Chapitre 3

Données et périmètre

3.1 Présentation des données

Cette section vise à présenter les différentes sources utilisées ainsi que les informations présentes dans celles-ci. La multiplication des sources de données permet d'enrichir et de fiabiliser les informations récoltées afin de pouvoir tirer des conclusions crédibles. Les données sont traitées à l'aide du logiciel SAS.

En effet, le risque décès est complexe. Il est difficile à modéliser car c'est un évènement rare mais coûteux. Ainsi, pour réaliser une étude assez robuste il est nécessaire de travailler sur un portefeuille assez large afin d'avoir assez de sinistres décès. Des contextes, tel que la pandémie de Covid-19, peuvent venir aggraver les risques de décès.

D'autre part, la réglementation distingue les décès directs des décès suite à exonération. De plus, les causes de décès direct et suite à exonération ne sont sûrement pas les mêmes. Ainsi, il y a là encore une disparité des cas.

3.1.1 Données Insee

Ces données contiennent tous les décès qui ont eu lieu sur le territoire français ainsi que ceux des français décédés à l'étranger. Elles ont été récupérées sur le site [data.gouv](https://www.data.gouv.fr)¹. Il manque donc les frontaliers c'est-à-dire les personnes travaillant en France mais n'habitant pas en France et qui sont étrangères. Elles ont été collectées par l'Insee qui reçoit les décès enregistrés dans les communes. L'Insee publie des fichiers mensuels, trimestriels et annuels. Ce sont ces derniers qui ont été récupérés pour les années 2018 à 2021. Le fichier du premier trimestre de 2022 a également été récupéré car il peut contenir des décès survenus en fin d'année 2021 et qui n'avaient pas encore été déclarés au moment de la publication du fichier annuel de 2021. Ces fichiers contiennent tous les décès qui ont été portés à la connaissance de l'Insee sur ladite période ainsi certains décès peuvent être antérieurs à cette date s'ils ont été reçus tardivement. L'Insee étant soumise à une obligation légale de diffusion, les lois concernant le caractère personnel des données au bénéfice des ayants droit ou héritiers ne s'appliquent pas ici.

De plus, l'Insee ne garantit ni l'exhaustivité ni l'exactitude des données ce qui peut se révéler problématique pour cette étude. En effet, la qualité des données est un enjeu majeur en assurance. Les données ayant pour but d'être rapprochées avec celles du portefeuille, la qualité de celles-ci est donc importante afin de bien réconcilier les 2 sources d'information.

Les données récupérées sont donc les décès survenus entre 2018 et 2021 avec les noms et prénoms, sexe, date de naissance, code, localité et pays de naissance (si le pays n'est pas la France), date de décès et code du lieu de décès. Ces données collectées sont en accord avec le RGPD (Règlement Général sur la Protection des Données)

1. <https://www.data.gouv.fr>

car les informations concernant des personnes décédées ne sont pas des données à caractère personnel et ne relèvent pas du secret de la vie privée.

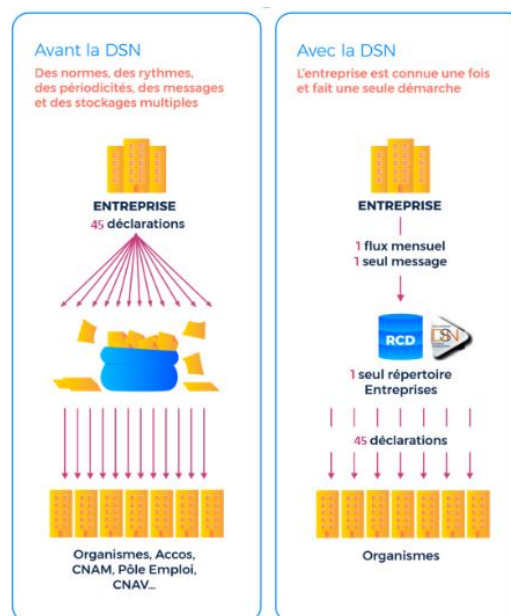
3.1.2 Données DSN

Suite à la loi de simplification du 22 mars 2012, la Déclaration Sociale Nominative (DSN) a été mise en place et remplace la plupart des déclarations sociales à la charge des employeurs. Cette déclaration est obligatoire pour toutes les entreprises du secteur privé depuis 2017 et depuis le 1er janvier 2022 pour le secteur public. L'employeur déclare ainsi de manière unique, mensuelle et dématérialisée les données reflétant la situation de ses salariés au moment de la paie. Ainsi, tous les événements survenus dans le mois sont déclarés (arrêt de travail, décès, fin de contrat de travail ...). Il s'agit d'une norme nationale mais chaque assureur reçoit uniquement les données des entreprises qu'il assure. Ainsi, dans le cadre de cette étude seules les données concernant le portefeuille de GGVIIE ont été utilisées.

Tous les salariés des entreprises assurées sont donc présents dans ces données ce qui permet d'étudier la population assurée avec précision. Ces éléments permettent ainsi d'obtenir l'exposition de l'assureur, donnée qui était inconnue auparavant en assurance collective. Cette nouvelle information permet ainsi d'utiliser les mêmes méthodes qu'en assurance individuelle. Cela représente donc une réelle innovation dans le domaine de l'assurance collective.

Les informations disponibles sont divisées en plusieurs catégories :

- Données générales concernant l'individu
 - Numéro de sécurité sociale
 - Nom et prénom
 - Sexe
 - Date de naissance
 - Lieu et département de naissance
 - Département de résidence actuelle
- Données concernant l'entreprise
 - SIRET
 - Code APE
 - Département
- Données concernant le contrat de travail de l'individu
 - Statut (CSP)
 - Date de début du contrat
 - Motif de fin de contrat
 - Date de fin de contrat
 - Nature du contrat
 - Salaire
 - etc.
- Données concernant les arrêts de travail de l'individu
 - Motif



- Dernier jour travaillé
- Date de fin prévisionnelle de l'arrêt
- Date de reprise

Le motif de fin de contrat de travail permet de reconnaître les personnes décédées. Les informations transmises via la DSN sont considérées fiables car elles reflètent ce qui est inscrit sur les bulletins de salaire des employés.

3.1.3 Données internes

Une base de données répertoriant tous les sinistres décès, à partir des informations de gestion interne, a été construite sous SAS. Elle répertorie par sinistre les différentes informations du dossier comme l'année de naissance de l'assuré, l'année de survenance du décès, des informations sur le contrat, un indicateur en cas de décès suite à arrêt de travail, un indicateur pour les dossiers comportant seulement une allocation d'obèques et les différentes prestations en distinguant les prestations échues et payées. En effet, le terme « échu » désigne le fait que lors de l'arrivée d'un dossier les montants sont estimés et provisionnés, puis ces montants peuvent être révisés. Les prestations ont été segmentées par garantie afin d'avoir le détail des prestations versées.

3.2 Utilisation de ces données

L'objectif est d'utiliser les données DSN afin de construire une loi de décès et de la comparer avec celles de l'Insee (TH 00-02 et TF 00-02). L'utilisation des données internes n'est pas possible car l'exposition n'est pas présente dans ces données. De plus, les données sont peu fiables car elles sont saisies manuellement et sont incomplètes (seule l'année de survenance du décès est connue et non la date exacte).

C'est pourquoi ce sont les données DSN qui sont utilisées pour réaliser l'étude. En effet, une des problématiques en assurance collective est le fait d'avoir peu d'informations disponibles, dû au fait du caractère collectif du contrat, sur les individus présents dans la population assurée. En outre, la gestion de la majorité des contrats de GGVIÉ étant déléguée, peu de données sont disponibles sur les populations assurées. Lors de la souscription du contrat avec l'entreprise, des informations générales sur la population salariale sont récoltées, afin de tarifier le contrat, mais ces dernières ne permettent pas une discrétisation précise de la population. Cependant, les données DSN permettent de récupérer des informations sur chaque salarié, comme sa date de naissance, son statut etc. et de constituer un historique de ses sinistres. Cela permet donc également d'avoir connaissance de l'exposition de la population assurée ce qui était auparavant, en assurance collective, inconnu de l'assureur.

Les données internes vont être utilisées afin de décrire le risque décès et la répartition de ces décès au sein du portefeuille de GGVIÉ.

Les données Insee vont être utilisées afin de fiabiliser les données DSN. Ainsi une fusion des deux bases est effectuée afin de vérifier la fiabilité de l'indicateur décès en DSN (tous les individus reportés comme décédés en DSN sont présents dans le fichier Insee) et également vérifier si certains décès n'ont pas été déclarés en DSN (individus présents dans le fichier Insee et en DSN mais non déclarés comme décédé en DSN).

3.3 Périmètre

L'étude concerne la période allant du 1^{er} janvier 2018 au 31 décembre 2021. Cette période a été choisie car les données DSN dans les bases de GGVIÉ ne sont fiables qu'à partir de 2018. De plus, afin d'éviter des problèmes de saisonnalité on étudie des années entières. Une période de 4 ans permet également de lisser des effets particuliers, tel que la hausse des décès due à la pandémie de Covid-19 à partir de 2020.

Le but de ce mémoire est de construire une loi de mortalité pour les salariés en ne prenant en compte que les décès directs (ne faisant pas suite à un arrêt de travail). Ainsi, quelques hypothèses, comme la durée de la franchise pour les arrêts de travail, sont fixées. En effet, la franchise prévue au contrat de chaque individu n'étant pas connue, une franchise de 90 jours a donc été retenue pour l'ensemble des arrêts de travail. Cette durée est la plus courante dans les conditions générales et est assez longue afin de ne pas écarter de décès considérés comme « directs ». De surcroît, les arrêts de travail pour cause de maternité, paternité ou adoption ont été écartés car il ne s'agit pas d'arrêts indemnisés par l'assureur, ce sont des arrêts prévus dans le code du travail. Ainsi, seul la Sécurité Sociale indemnise ces arrêts. Par ailleurs, en cas d'invalidité l'hypothèse d'impossibilité de retour à l'état d'actif est prise.

Par conséquent, l'exposition est retraitée en recalculant les périodes d'observation de chaque individu par rapport aux arrêts de travail de plus de 90 jours et aux invalidités observées. L'exposition est calculée par contrat. Ainsi, un individu sera présent plusieurs fois dans la base s'il possède plusieurs contrats différents durant la période d'étude. Enfin, les contrats d'intérim sont éliminés car ce type de contrat représente un volume important mais les durées de ces contrats sont très courtes. Ce sont donc des contrats à traiter à part car ils ne possèdent pas les mêmes risques que les autres.

3.4 Traitements

Dans cette partie, les différents traitements effectués sur les données afin de les fiabiliser et de les formater vont être détaillés. Les traitements ont été réalisés en plusieurs étapes. Dans un premier temps, la table concernant les contrats de travail est retraitée. Puis, les arrêts de travail sont retraités afin de ne garder que ceux ayant dépassé la franchise. Ensuite, les suspensions pour invalidité sont retraitées afin de conserver la première date de suspension. Finalement, la table contenant les contrats de travail et donc les périodes d'exposition est révisée en fonction des dates d'arrêts de travail et de suspensions.

3.4.1 Contrats de travail

Au départ, chaque contrat actif possède une ligne par mois, c'est-à-dire correspondant à chaque déclaration DSN mensuelle. Différents traitements sont réalisés afin d'obtenir la période d'observation de chaque contrat et de ne retenir qu'une seule ligne par contrat.

Des variables sont créées tel qu'une indicatrice décès valant 1 en cas de décès renseigné et 0 sinon, les variables *min_observation* et *max_observation* renseignent les bornes de la période d'observation de l'individu, elles ont été choisies de manière prudente afin de ne pas surestimer la période de visualisation. L'âge au début de l'observation a également été calculé.

Enfin, les contrats avec moins de 3 jours d'écart et concernant le même individu, Siret, statut, nature de contrat et statut conventionnel ont été regroupés. En effet, cela crée un volume important avec de petites périodes d'observation et ainsi gonfle artificiellement le nombre de contrats observé. Cette durée de 3 jours permet de prendre en compte les week-ends par exemple, afin de regrouper des contrats qui se terminent en fin de semaine et qui reprennent en début de semaine.

Par la suite, les contrats dont la durée est de moins de 15 jours ont été écartés car il s'agit là de contrats qui peuvent être assimilés à de l'intérim. Ces derniers ayant été écartés auparavant de l'étude, il est donc cohérent d'isoler ces contrats également.

Par ailleurs, les contrats dont l'âge au début de l'observation est inférieur à 16 ans et ceux dont l'âge au début de l'observation est supérieur à 79 ans ont également été écartés, car dans le premier cas 16 ans est

considéré comme l'âge légal pour travailler, et dans le deuxième cas cela permet de prendre en compte le cas des cumuls emploi/retraite. Néanmoins, une limite a dû être fixée afin de conserver une exposition assez conséquente.

La table d'exposition obtenue comporte 4 310 768 lignes dont 5 921 décès. A noter que pour l'instant les décès suite à arrêt de travail sont présents. Il faut maintenant retraiter cette exposition afin de recalculer les périodes d'observation en cas d'arrêt de travail de plus de 90 jours.

3.4.2 Arrêts de travail

De même que pour les contrats de travail, au départ, chaque déclaration DSN sur laquelle l'arrêt de travail est présent est incluse dans la table *arrêt de travail*. Ainsi, des traitements sont effectués afin d'obtenir une seule ligne par arrêt, cette dernière comprend donc la période totale de l'arrêt. Pour rappel, les arrêts ayants pour motif un congé maternité, paternité ou adoption ont été préalablement écartés.

Les arrêts de travail qui ont moins de 7 jours d'écart sont regroupés. En effet cela est considéré comme une rechute et donc comme un seul et même arrêt. De plus, cela permet également de regrouper les arrêts pour cause de mi-temps thérapeutique, pour lesquels chaque jour non travaillé est parfois déclaré comme un arrêt.

Par ailleurs, l'hypothèse est prise qu'en cas de contrats de travail simultanés si un arrêt de travail survient alors il survient sur tous les contrats en vigueur à la date de cet arrêt.

A la suite de ces regroupements, seuls les arrêts de plus de 90 jours et dont la période d'arrêt coïncide au moins en partie avec la période d'observation sont conservés, la franchise retenue étant de 90 jours. La base obtenue contient 206 581 arrêts de travail.

3.4.3 Suspensions pour invalidité

Enfin, pour les suspensions pour cause d'invalidité, la première date de suspension est retenue par contrat. Ainsi, l'individu ne sera plus observé à partir de cette date. L'hypothèse de non-retour à l'état d'actif suite à un passage en invalidité ayant été prise. De plus, en cas de plusieurs suspensions par individu, la date de suspension la plus ancienne sera retenue afin d'être prudent. La base obtenue pour l'invalidité contient 2 500 lignes.

3.4.4 Regroupement des tables

Suivant les dates d'arrêt de travail et d'invalidité, les périodes d'observation de chaque individu pour chaque contrat sont recalculées. Cela est effectué par étape en considérant d'abord le cas des arrêts de travail puis le cas des invalidités. Ainsi, en cas d'arrêts de travail multiples au cours d'un même contrat, les périodes d'observation sont alors multiples et elles sont entrecoupées de périodes de non observation qui correspondent aux arrêts de travail. Ainsi, un traitement est effectué afin d'obtenir une ligne par période d'observation continue. C'est pourquoi, le nombre de lignes de la base peut augmenter au final par rapport à la table des contrats initiale.

A noter, que l'indicatrice indiquant les décès est également recalculée car le décès peut alors avoir lieu en dehors de la période retenue (cas des décès suite à arrêt de travail). Cette table contient donc seulement les décès ne faisant pas suite à un arrêt de travail. Pour finir, les périodes d'observation de moins de 15 jours sont supprimées. Et cela, afin d'être cohérent avec les retraitements effectués précédemment.

Ainsi, la table obtenue comprend 4 241 décès directs pour 4 401 999 périodes d'observation continues. En effet, comme expliqué auparavant, un même contrat peut maintenant être découpé en plusieurs lignes si un ou plusieurs arrêts de travail ont eu lieu. Ainsi, la maille de la base n'est plus seulement l'individu et le contrat de travail comme dans la table des contrats initiale. Cette nouvelle table est donc à la maille individu, contrat de travail et période d'observation.

3.5 Comparaison avec les données Insee

Les données Insee sont utilisées ici afin de vérifier la qualité des déclarations DSN reçues. Cette étude est réalisée en deux parties. Dans la première partie les décès directs déclarés en DSN sont rapprochés des décès Insee. Dans la deuxième partie, tous les individus non considérés comme décédés en DSN sont rapprochés des données Insee, ainsi les individus dont le décès fait suite à un arrêt de travail sont présents dans cette liste. Ce deuxième cas permet de vérifier si des décès n'ont pas été déclarés en DSN.

3.5.1 Rapprochement des décès DSN avec les données Insee

Cette première étude a pour but de vérifier la qualité des données DSN, tel que le délai de déclaration et la date de fin de contrat qui en cas de décès doit correspondre à la date de décès.

Un travail de mise en forme des données a été effectué afin notamment d'enlever les caractères spéciaux et les accents, ces traitements ont été effectués de la même manière sur les données DSN et Insee afin d'harmoniser les variables tel que le nom et prénom. A noter que pour les données Insee, tous les prénoms de l'individu sont saisis dans une même variable ainsi cela ne permet pas dans certains cas de reconnaître les prénoms composés. De plus, le nom d'usage n'est pas toujours renseigné dans la DSN et les données Insee ne font pas la distinction entre le nom de naissance et le nom marital. Ainsi, ces différences compliquent le rapprochement des données.

Comme les décès DSN sont peu nombreux, peu d'homonymes sont présents ici. De plus, si les déclarations ne sont pas frauduleuses et ne concernent pas des travailleurs frontaliers, elles doivent être présentes dans les données Insee. Ainsi, 8 niveaux de clés ont été utilisés afin de rapprocher les 2 bases. La clé la plus précise nécessite l'égalité du nom de famille, prénom, date de naissance, département de naissance et sexe. Les autres clés sont basées sur cette première clé de rapprochement avec des différences comme par exemple le nom d'usage à la place du nom de famille ou encore la suppression de l'égalité nécessaire entre les départements de naissance. Des vérifications sur les rapprochements avec les clés moins précises ont été effectuées, ainsi que des vérifications entre les dates de décès déclarées.

Finalement, plus de 96% des décès déclarés en DSN ont été retrouvés dans les données Insee. Cela permet de confirmer la qualité des déclarations DSN. Il y a 85% des dates de décès qui sont exactement les mêmes et 97% qui ont moins de 7 jours d'écart.

De plus, pour les 4% non rapprochés plusieurs explications sont possibles. Une première raison est qu'il y a trop d'erreurs de saisies dans les données et le rapprochement n'a donc pas pu avoir lieu. Une autre explication est que les données Insee pour les personnes étrangères ne contiennent que les décès ayant lieu sur le sol français. Ainsi les travailleurs frontaliers décédés (vivant à l'étranger et travaillant en France) seront déclarés en DSN mais ne seront pas présents dans la base de l'Insee. Enfin, en cas de nom de naissance et de nom marital, les deux ne sont pas forcément connus de l'employeur et un seul est communiqué dans les données de l'Insee. Ainsi, certains individus peuvent ne pas avoir été rapprochés pour cette raison étant donné que le nom constitue une information capitale pour le rapprochement.

3.5.2 Rapprochement des individus non décédés en DSN avec les données Insee

Dans cette partie, les individus dont aucun décès direct n'a été observé en DSN sont rapprochés avec les données Insee. Ici, le problème des homonymes est présent car le volume de données dans les 2 bases est très important. Ainsi, les clés utilisées pour le rapprochement sont plus restrictives et seulement 3 clés différentes sont utilisées. Finalement, 6 208 individus sont rapprochés. Ces fusions se divisent en 3 catégories.

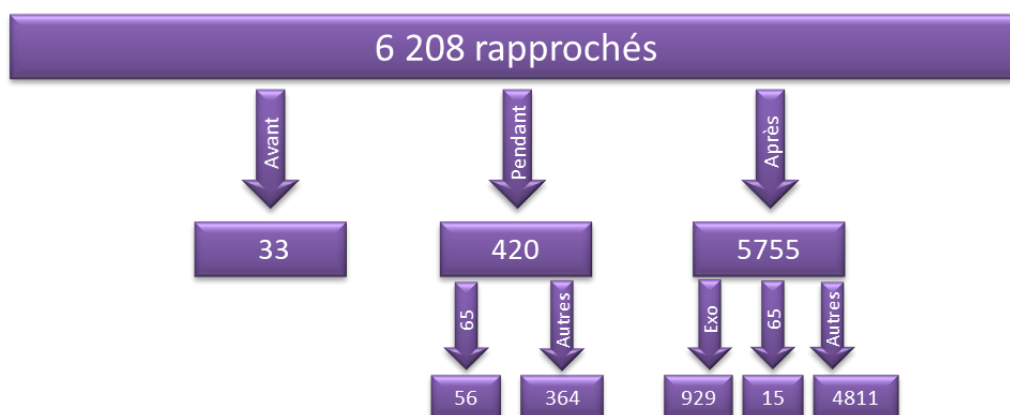


FIGURE 3.1 – Schéma des résultats du rapprochement des individus non décédés en DSN avec les données Insee

Premièrement, il y a 33 individus dont la date de décès Insee est antérieure au début d’observation en DSN. Il s’agit donc d’anomalies. Ces cas peuvent correspondre à de la fraude ou à une déclaration tardive en DSN. Ainsi, ces cas sont supprimés de notre étude et feront par ailleurs l’objet d’investigations.

Deuxièmement, il y a 420 individus dont la date de décès Insee se trouve pendant la période d’observation DSN. Sur ces cas, 56 ont un motif de fin de contrat correspondant au décès de l’employeur. Ainsi, il est raisonnable de penser qu’il s’agit de rapprochements corrects et que l’employeur et l’employé sont dans ces cas une seule et même personne. Ainsi, il s’agit d’une déclaration DSN incomplète, car le décès de l’employeur a été déclaré mais pas celui du salarié, or il s’agit dans ce cas de la même personne. En ce qui concerne les autres cas, aucune explication n’a été trouvée, car il est difficile d’investiguer des cas concrets car peu de données sont connues sur le contexte des décès. Pour la suite de cette étude, ces 420 décès sont rajoutés à l’étude.

Troisièmement, les 5 755 autres rapprochements ont une date de décès Insee qui est postérieure à la période d’observation en DSN. Cela semble donc cohérent avec les analyses effectuées. De plus, au moins 929 de ces décès sont en réalité des décès faisant suite à un arrêt de travail. Ces derniers avaient été écartés de l’étude et leur période d’observation avait été recalculée. Et 15 de ces décès sont comme vu précédemment des décès de l’employeur. Ici, les autres cas sont très certainement des décès suite à arrêt de travail qui n’avaient pas été identifié en DSN, mais également des décès directs ayant eu lieu en dehors des périodes d’observation en DSN.

Cette étude a permis de vérifier et de renforcer la qualité des données DSN. Ainsi, tous les décès observés en DSN sont conservés pour l’étude et la période d’observation est tronquée par la date de décès présente dans les données Insee. Cette date étant plus fiable car elle est déclarée à partir de l’acte de décès Cela permet finalement de supprimer 10 décès car la durée d’observation devient négative. Il s’agit en fait de décès faisant suite à un arrêt de travail mais déclarés en retard.

3.6 Derniers traitements

A la suite de ces changements dans les données, des traitements permettant d’actualiser les indicateurs précédemment construits sont réalisés. De plus, de nouveaux indicateurs sont créés afin d’enrichir les dimensions d’étude.

Premièrement, des informations comme le salaire moyen mensuel et le secteur d’activité sont rajoutées afin d’enrichir l’étude. Le salaire moyen mensuel n’est rajouté que si la période totale sur laquelle le salaire est

connue est assez conséquente. De plus, des valeurs aberrantes sont également écartées. Ces traitements sont effectués afin de fiabiliser cette variable.

Afin d'étudier l'effet du salaire sur la mortalité, des niveaux de salaire sont créés à partir des salaires moyens mensuels. Deux variables différentes sont calculées.

Une première évaluant le niveau de salaire par rapport à l'ensemble des individus. Une deuxième calculant le niveau de salaire par rapport aux individus ayant la même CSP. Les deux variables possèdent 5 niveaux différents. La deuxième variable permet de prendre en compte le salaire en corrigeant l'impact de la CSP. Les niveaux sont créés en fonction de l'exposition. Ainsi, chaque classe possède le même pourcentage d'exposition. La classe 1 représente donc les 20% les plus pauvres tandis que la classe 5 représente les 20% les plus riches.

A noter, que des temps partiels sont présents dans la base mais que le salaire de ces individus a été recalculé afin qu'il soit équivalent à un temps plein. Cependant, il a été difficile de recalculer un salaire équivalent à un temps plein car le temps travail n'est pas disponible dans les données. En effet, le champ correspondant à ce taux de travail est peu renseigné. Seule une variable précisant si le contrat est en temps plein ou non est disponible. Plus de 88% des cadres travaillent à temps plein alors que seulement 77,5% des non-cadres travaillent à temps plein. Pour ces derniers la disparité entre sexe est bien visible avec seulement 64% des femmes non-cadres qui travaillent à temps plein contre 89% des hommes non-cadres. Finalement, une hypothèse assez forte supposant que le taux moyen d'un temps partiel est de 66,66% a été prise. Ce taux se base sur un article de l'Insee². Au regard de cela, une prudence supplémentaire devra être accordée aux impacts possibles de cette hypothèse trop globale.

En ce qui concerne le secteur d'activité, des classes assez générales sont utilisées afin de voir la significativité de ce critère. Le secteur est obtenu à partir du code APE, également appelé NAF, en utilisant le regroupement par section de l'Insee³. Ainsi, 21 secteurs d'activités sont obtenus.

Par ailleurs, les données sont également séparées par année calendaire afin de permettre l'étude de l'impact de cette variable dans la partie modélisation. De plus, cela permet de pouvoir utiliser l'âge comme critère plus facilement. Par conséquent, les périodes d'observation, l'indicatrice de décès, ainsi que les âges au début de chaque nouvelle période d'observation sont recalculés. Enfin, un coefficient d'exposition est calculé. Il correspond à la durée d'observation en jours divisée par le nombre de jours dans l'année. C'est donc un coefficient se situant entre 0 et 1 qui indique le temps d'observation d'un individu sur une année.

La région de résidence de l'individu est également retenue comme variable explicative. Elle est calculée à partir des départements de résidence des individus. Les départements qui ne correspondent pas à des départements de la France métropolitaine hors Corse, sont regroupés dans une modalité 'Autres'. En effet, l'exposition n'est pas assez conséquente pour discrétiser la Corse, des départements outre-mer, des pays étrangers.

Enfin, afin d'étudier l'âge, des tranches d'âge ont été créées afin de pouvoir analyser cet effet tout en gardant une exposition assez conséquente. Cinq tranches sont créées :

- moins de 30 ans
- entre 30 et 40 ans
- entre 40 et 50 ans
- entre 50 et 60 ans
- plus de 60 ans

2. <https://www.insee.fr/fr/statistiques/fichier/2891765/ecrt17j-f4-temps-travail.pdf>

3. <https://www.juripresse.fr/liste-des-codes-naf-ape/>

Cette segmentation a été réalisée arbitrairement, tout en s'assurant que l'exposition serait assez conséquente dans chaque tranche d'âge et assez équilibrée. A noter, que celle des plus de 60 ans a tout de même une exposition beaucoup plus faible que les autres. Mais cela n'est pas gênant pour l'étude car les plus de 60 ans ne sont pas la préoccupation principale de cette étude.

Cependant, les tranches d'âge auraient pu être calculées afin d'optimiser les regroupements et d'homogénéiser au maximum les différentes tranches. En outre, le nombre de tranche d'âge pourrait lui aussi faire l'objet d'une étude afin d'être optimisé.

Finalement, quelques traitements sont effectués afin de conserver des expositions assez conséquentes et de supprimer les données aberrantes. Pour les non-cadres un filtre est appliqué sur les catégories socio-professionnelles pour retenir les 3 principales, les autres comportant moins de 0,4% de l'exposition. Les 3 catégories socio-professionnelles retenues sont :

- les ouvriers qualifiés et non qualifiés y compris les ouvriers agricoles
- les professions intermédiaires tels que les techniciens, contremaîtres, agents de maîtrise et clergés
- les employés administratifs d'entreprise, de commerce et les agents de service.

Enfin, si la variable salaire est inconnue ou aberrante, alors les individus sont écartés de l'étude. Cela résulte à une exposition totale de 5 303 209,81 pour 4 422 décès.

Cela se décompose de la façon suivante selon le statut. Pour les cadres il y a finalement 840 décès pour une exposition de 1 359 099,01. Pour les non-cadres, il y a 3 582 décès pour une exposition de 3 944 110,8.

A présent, les données ont été fiabilisées et formatées. L'analyse de ces données peut donc commencer. Tout d'abord des statistiques descriptives sont étudiées afin de mieux appréhender la population assurée.

Chapitre 4

Statistiques descriptives et analyses

4.1 Statistiques descriptives

Afin de mieux appréhender les données de cette étude des statistiques descriptives sont produites. Ces dernières permettent d'obtenir des premiers résultats sur la population étudiée et ainsi de mieux comprendre les différentes caractéristiques des individus composant ce portefeuille.

4.1.1 Données internes

Dans cette partie, les données internes des sinistres décès ont été utilisées. Cela permet de mettre en évidence les différences entre les décès directs et ceux faisant suite à un arrêt de travail. En effet, dans les données internes un indicateur permet de distinguer les 2 types de décès. Afin de rester sur des données comparables, seuls les dossiers avec une année de survenance entre 2018 et 2021 sont utilisés.

Le graphique suivant représente la proportion de dossiers décès faisant suite à un arrêt de travail par année de survenance par rapport au nombre de dossiers décès total¹. Ce pourcentage est assez constant et vaut en moyenne 42,6%. Ainsi, l'hypothèse que les décès suite à arrêt de travail représentent 43% des décès sera prise dans la suite de cette étude afin de faciliter la comparaison avec les autres lois de mortalité dans certains cas.

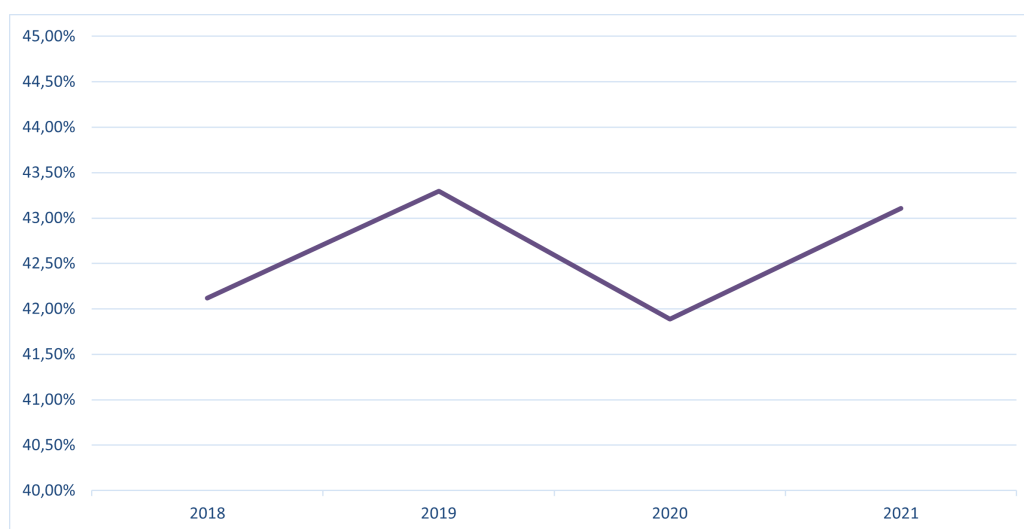


FIGURE 4.1 – Proportion de dossiers décès faisant suite à arrêt de travail par année de survenance

1. en ne prenant pas en compte les dossiers où seule une allocation obsèques est versée

Ensuite, l'âge entre les 2 groupes est étudié. L'âge moyen de chaque groupe est calculé dans un premier temps. L'âge moyen des décès directs est inférieur à celui des décès faisant suite à un arrêt de travail. Néanmoins, il y a moins de 2 ans d'écart ce qui est relativement peu.

	Age moyen
Décès directs	52,578
Décès suite à arrêt de travail	54,354

FIGURE 4.2 – Ages moyens suivant le type de décès

Ensuite, une répartition par tranches d'âge est effectuée afin d'examiner plus en détail la répartition des décès en fonction des âges et des types de décès. Cela semble en accord avec les observations précédentes car la proportion des décès suite à arrêt de travail est plus importante sur les tranches d'âge les plus élevées. Une autre remarque est que presque 85% des décès sont entre 40 et 70 ans. Cela s'explique par le fait qu'il s'agit des âges avec la plus grande exposition dans le portefeuille.

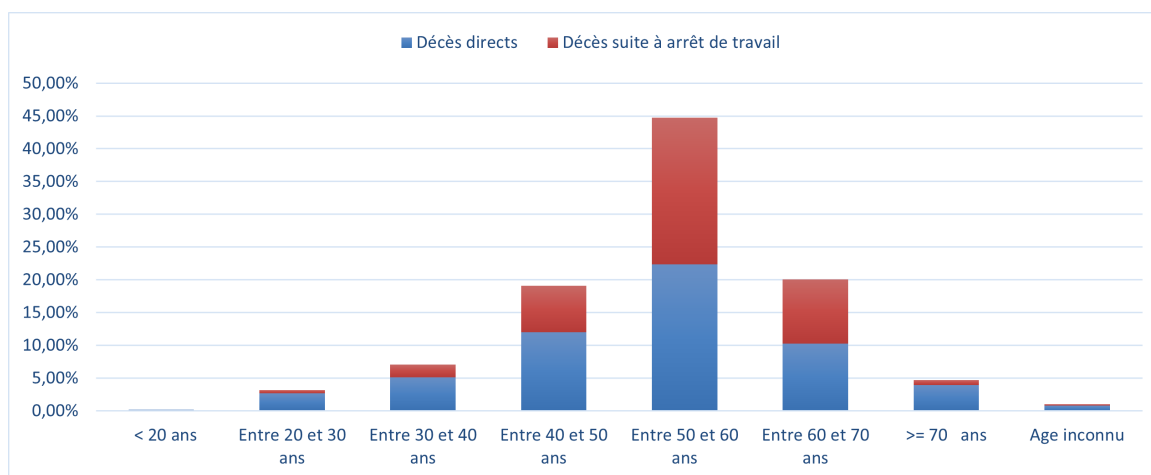


FIGURE 4.3 – Répartition des décès par type et tranches d'âge

4.1.2 Données DSN

Dans cette partie, les données DSN retraitées à l'aide des données Insee ont été utilisées, cela correspond donc aux données de la table d'exposition finale. Dans un premier temps, l'exposition est analysée. Puis, les décès seront analysés et enfin les différents critères d'étude seront analysés afin de voir la répartition de la population dans les différentes modalités.

L'exposition moyenne est intéressante à analyser. Cela correspond au temps moyen d'observation sans interruption d'un individu sur une année. Les interruptions étant dues aux arrêts de travail, démissions, passages à la retraite, changements d'entreprise ou de statut par exemple ou à des déclarations DSN manquantes. Les résultats sont segmentés par sexe et tranches d'âge afin de voir les différents impacts de ces critères sur le temps d'observation sans interruption.

Sexe	Tranche d'âge	W moyen	
Hommes	< 30 ans	0,573	0,719
	30-40 ans	0,731	
	40-50 ans	0,777	
	50-60 ans	0,806	
	>= 60 ans	0,713	
Femmes	< 30 ans	0,499	0,653
	30-40 ans	0,677	
	40-50 ans	0,716	
	50-60 ans	0,750	
	>= 60 ans	0,686	

FIGURE 4.4 – Exposition moyenne par sexe et tranche d'âge

Ainsi, l'exposition moyenne est assez élevée, mais celle des hommes est supérieure à celle des femmes. Cela peut être due en partie au congé maternité des femmes mais cela ne peut être la seule explication car l'exposition des femmes est inférieure à celle des hommes quelque soit la tranche d'âge.

Par ailleurs, la répartition par tranches d'âge semble cohérente avec ce qui peut être attendu. Les moins de 30 ans ayant une exposition moyenne faible car les jeunes travaillent pendant leurs vacances par exemple ou ont des contrats avec une date de fin déterminée.

Afin de voir si des valeurs extrêmes influencent les résultats, la médiane est également étudiée.

Sexe	Tranche d'âge	Exposition médiane	
Hommes	< 30 ans	0,542	1,000
	30-40 ans	1,000	
	40-50 ans	1,000	
	50-60 ans	1,000	
	>= 60 ans	1,000	
Femmes	< 30 ans	0,404	0,805
	30-40 ans	0,852	
	40-50 ans	1,000	
	50-60 ans	1,000	
	>= 60 ans	0,915	

CSP	Exposition médiane
Cadre	1
Profession intermédiaire	1
Employé	0,734246575
Ouvrier	0,871232877

FIGURE 4.5 – Exposition médiane par sexe et tranche d'âge puis par CSP

Des résultats intéressants sont obtenus. En effet, beaucoup de tranches d'âge ont une médiane égale à 1. Cela est attendu pour les âges correspondants à ceux de la vie active c'est-à-dire entre 30 et 60 ans ici. Les mêmes arguments que précédemment expliquent les médianes faibles observées pour les moins de 30 ans. En revanche, il est intéressant d'observer que la médiane des femmes entre 30 et 40 ans est inférieure à 1. Cela peut avoir un lien avec les congés maternités. Cela indiquerait que les femmes changent de poste ou prennent une pause dans leur carrière professionnelle pour s'occuper de leurs jeunes enfants. Cet argument étant cohérent avec le fait que les femmes plus âgées ont bien une médiane égale à 1. A noter, que les congés maternité n'ont pas été retraités et que donc les femmes sont bien considérées comme présentes dans l'entreprise durant cette période. Une autre possibilité est la mauvaise qualité des déclarations DSN durant cette période.

En ce qui concerne les catégories socio-professionnelles, les médianes des cadres et des professions intermédiaires sont égales à 1 alors que celles des employés et des ouvriers sont inférieures à 1. Cela suggère donc que les employés et ouvriers ont plus d'arrêts de travail ainsi qu'une plus grande mobilité sur leurs postes.

La répartition de l'exposition par sexe et catégorie socio-professionnelle est représentée sur le graphique suivant. Ainsi, le portefeuille est composé en majorité d'hommes et ce sont les non-cadres qui sont le plus représentés et plus spécifiquement les employés.



FIGURE 4.6 – Répartition de l'exposition par sexe et CSP

La pyramide des âges de l'exposition est tracée en fonction du sexe. L'exposition se trouve principalement entre 25 et 55 ans.

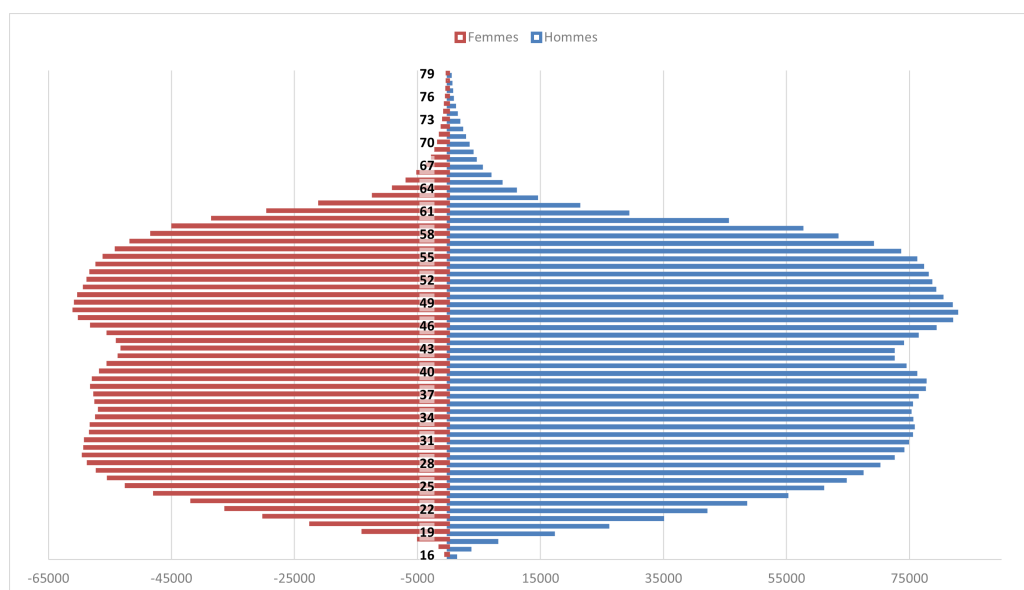


FIGURE 4.7 – Pyramide des âges de l'exposition

Les mêmes graphiques sont réalisés pour les décès directs.

Ici, 60% des décès sont des hommes non-cadres, et plus de la moitié de ces hommes non-cadres sont des ouvriers. Cela semble en accord avec les normes, même si l'écart de mortalité entre les hommes et les femmes semble très accentué ici. En effet, les lois de l'Insee démontrent une mortalité plus élevée des hommes.

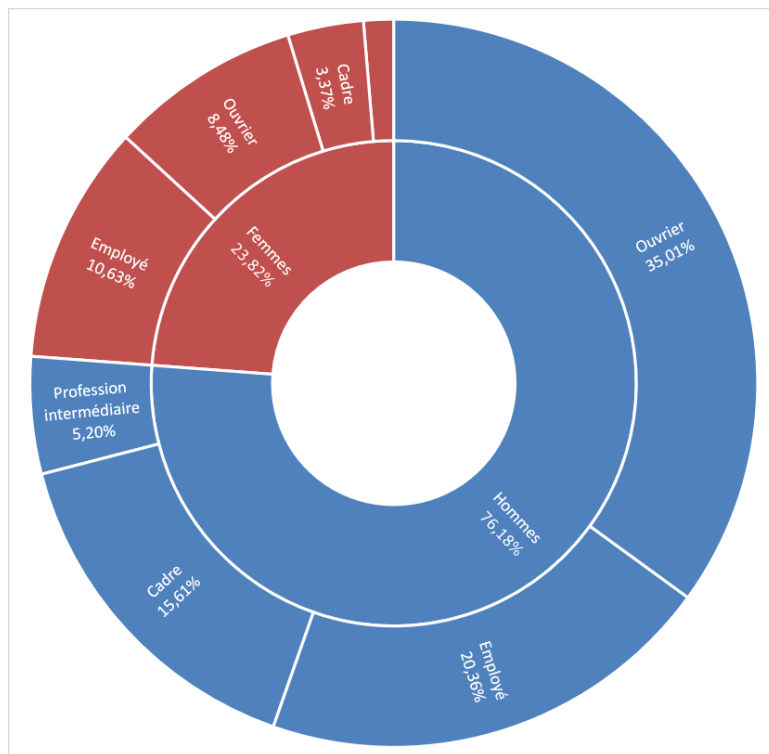


FIGURE 4.8 – Répartition des décès directs par sexe et statut

La pyramide des âges des décès sans arrêt de travail est tracée en fonction du sexe. Les décès se trouvent principalement entre 52 et 61 ans pour les hommes et entre 52 et 59 ans pour les femmes.

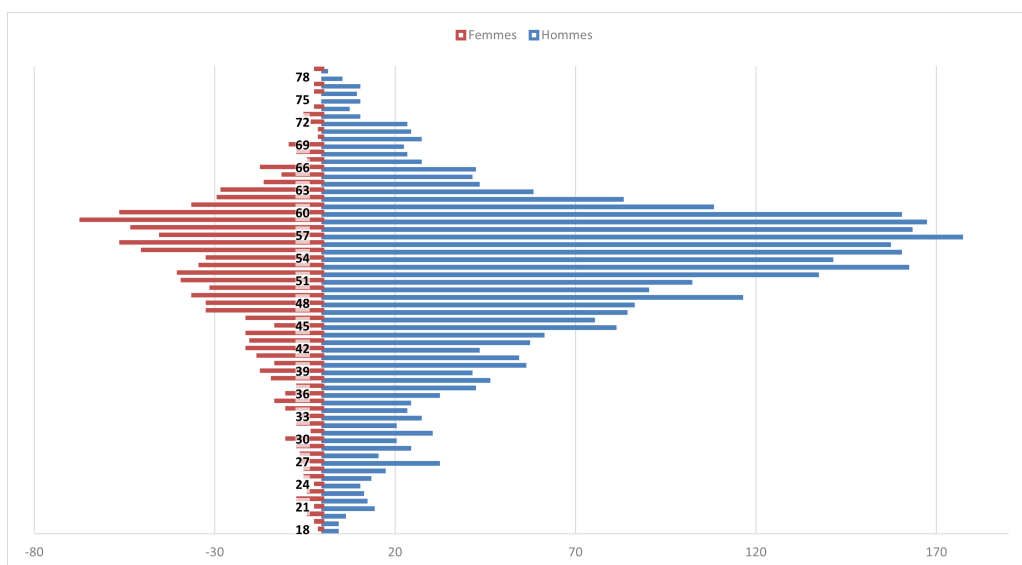


FIGURE 4.9 – Pyramide des âges des décès directs

Maintenant que la répartition de la population par sexe et âge est déterminée, les critères explicatifs sont considérés.

Afin de prendre en compte la variable salaire de la meilleure manière, plusieurs pistes ont été envisagées. Comme expliqué précédemment, finalement 2 variables sont retenues.

Premièrement, une variable représentant des niveaux de salaire est calculée sur tous les individus. Celle-ci comprend 5 niveaux différents et chaque niveau contient 20% de l'exposition. Ainsi, le niveau 1 représente les 20% les plus pauvres et le niveau 5 les 20% les plus riches. Cependant, cette variable vient capturer les effets des catégories socio-professionnelles. En effet, dans le cinquième niveau la grande majorité des individus sont des cadres. Ainsi, cette variable c'est avérée biaisée quand elle était utilisée dans les modèles.

Les salaires moyens de chaque niveau sont représentés ci-dessous. La classe 5 se distingue particulièrement des autres avec un salaire moyen correspondant à plus de deux fois celui de la classe 4. Cela confirme que de nombreuses valeurs erronées sont présentes.

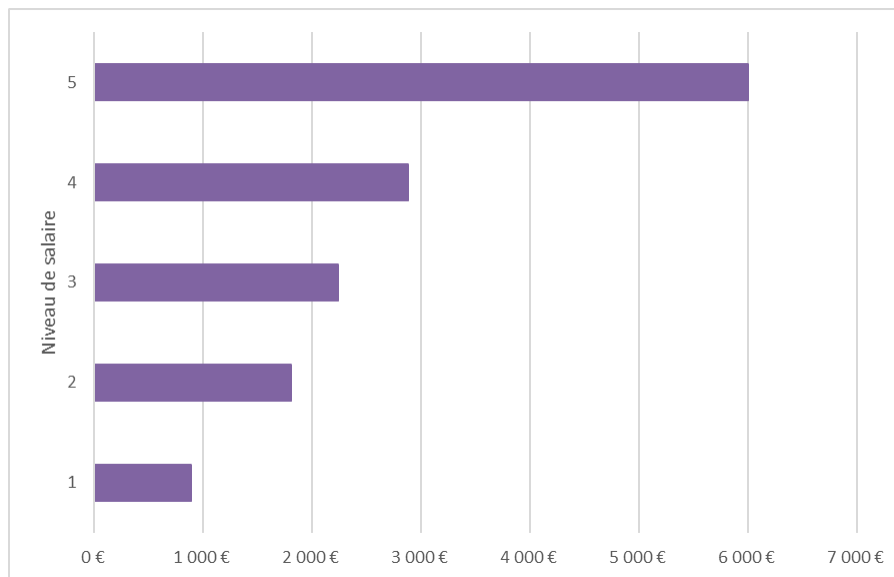


FIGURE 4.10 – Représentation du salaire moyen par niveau de salaire

Ainsi, une deuxième variable est calculée. Elle reprend le même principe que la première avec également 5 niveaux de salaire. Mais dans ce cas les niveaux sont calculés en fonction de la CSP. Ainsi, les niveaux sont propres à chaque CSP.

Les salaires moyens de chaque niveau par CSP sont représentés ci-dessous. Le salaire moyen de la première classe des ouvriers semble très faible, car il se situe bien en-dessous du SMIC. En opposition, celui de la dernière classe des cadres semble extrêmement élevé. Cela laisse envisagée que la qualité de la variable salaire est mauvaise, mais les classes peuvent permettre d'atténuer l'effet des valeurs extrêmes.

A noter, que comme expliqué précédemment une partie importante des individus ne sont pas à temps plein, et leur taux d'activité n'est pas connu. Une remise à niveau de manière globale a été effectuée sur les salaires des individus qui sont en temps partiels afin de ramener ces salaires à un équivalent temps plein. Ainsi, l'hypothèse que le taux moyen de travail en temps partiel est de 66,66% a été prise. Cela reste une hypothèse assez générale et peu précise. Par ailleurs, la variable salaire présente des erreurs de saisies observées, qui malgré certains traitements afin de ne pas prendre en compte les valeurs les plus absurdes et le calcul d'une moyenne pondérée peuvent subsister et altérer la qualité de la variable.

Cependant le regroupement par niveaux de salaire permet de gommer l'effet de ces valeurs absurdes.

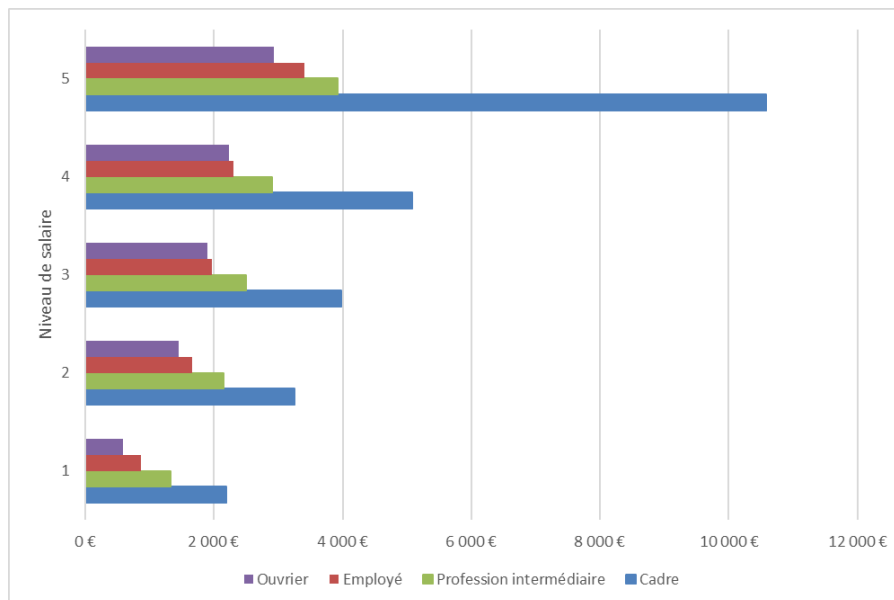


FIGURE 4.11 – Représentation du salaire moyen par CSP et par niveau de salaire

En outre, l'âge moyen par CSP et niveau de salaire par CSP est étudié. En effet, de manière générale le salaire augmente avec l'âge. Cela se confirme bien dans nos données comme on peut le voir dans le tableau suivant. Ainsi, les données semblent avoir la dynamique attendue dans l'ensemble.

CSP	Niveau Salaire	Age moyen
Cadre	1	40
	2	40
	3	42
	4	44
	5	48
Profession intermédiaire	1	36
	2	38
	3	40
	4	42
	5	44
Employé	1	34
	2	36
	3	37
	4	39
	5	41
Ouvrier	1	43
	2	41
	3	40
	4	42
	5	44

FIGURE 4.12 – Représentation de l'âge moyen par CSP et par niveau de salaire

Finalement, la région de résidence est également retenue comme critère explicatif. Cependant, une inquiétude face à ce critère est de ne pas avoir assez d'exposition dans les différentes régions de France. La catégorie "AUTRE" regroupe les valeurs non renseignées, les pays étrangers ainsi que les régions hors France métropolitaine ainsi que la Corse.

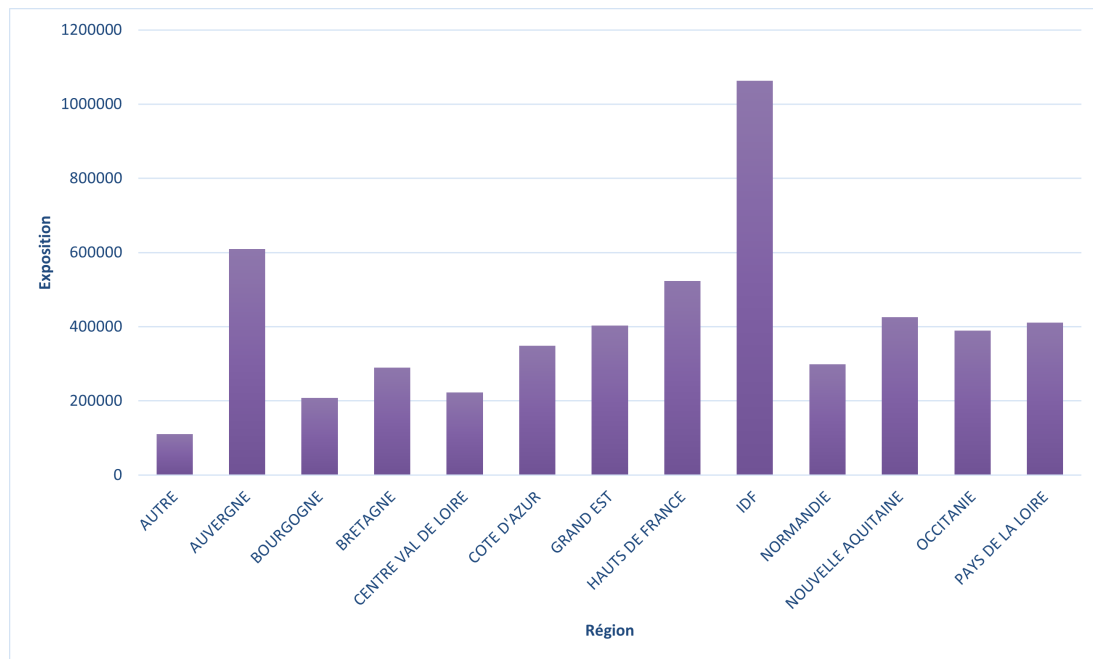


FIGURE 4.13 – Représentation de l'exposition par région

L'exposition semble assez élevée sur la grande majorité des régions. La répartition est très semblable à celle de la population française générale, ce qui indique que le portefeuille est réparti de manière uniforme sur le territoire français. Cependant, la catégorie "AUTRE" est peu représentée, de plus cette modalité regroupe des individus assez différents, ainsi aucune conclusion ne sera faite sur cette modalité, néanmoins elle est tout de même conservée pour l'étude.

Pour finir, la corrélation entre les différentes variables catégorielles est étudiée en utilisant le test du Khi-2 ainsi que le V de Cramer. Le test du Khi-2 permet de tester l'hypothèse nulle affirmant que les deux variables catégorielles sont indépendantes. Si cette hypothèse est vérifiée alors la statistique de test suit une loi du Khi-2 avec 2 degrés de liberté, si le test est effectué avec 2 variables. Le V de Cramer permet quant à lui de mesurer l'association entre les 2 variables. Sa valeur est comprise entre -1 et 1, 0 correspondant à une absence totale de lien et 1 à une dépendance parfaite.

Les variables sexe, tranche d'âge, CSP, niveau de salaire (sur la population générale), secteur d'activité, classe tarifaire et région de résidence sont étudiées. Les résultats des tests du Khi-2 démontrent que toutes les variables testées sont dépendantes deux à deux car toutes les p-values sont inférieures au seuil de 5% ainsi qu'à celui de 1%. Cependant, le V de Cramer permet de mesurer cette dépendance qui n'est pas la même entre toutes les variables. Les V de Cramer lors des tests de dépendance deux à deux sont illustrés dans le tableau suivant.

Ainsi, les dépendances sont relativement faibles entre certaines variables, comme la région de résidence ou encore le niveau de salaire. Une dépendance forte existe entre le sexe et le secteur d'activité ainsi qu'entre la CSP et le secteur d'activité. Par ailleurs, la classe tarifaire est déterminée en fonction du secteur d'activité donc il est normal que ces 2 variables soient extrêmement corrélées. Ces résultats sont cohérents avec ce qui était attendu. La corrélation entre la CSP et le niveau de salaire est forte. Ce résultat est normal car les cadres et les non-cadres n'ont pas les mêmes salaires. Cela renforce la décision prise de calculer des niveaux de salaire par CSP afin de prendre en compte cette disparité entre CSP.

V de Cramer	Sexe	Tranche d'âge	CSP	Niveau de salaire	Secteur d'activité	Classe tarifaire	Région
Sexe	1	0,03	0,21	0,19	0,32	0,14	0,04
Tranche d'âge		1	0,12	0,12	0,09	0,09	0,03
CSP			1	0,42	0,36	0,41	0,13
Niveau de salaire				1	0,23	0,18	0,1
Secteur d'activité					1	0,67	0,1
Classe tarifaire						1	0,12
Région							1

FIGURE 4.14 – Tableau des V de Cramer obtenus

4.2 Analyse de la répartition des décès

Dans cette partie, le temps de déclaration des décès est étudié à l'aide des données internes.

A partir des données internes, des triangles de répartition par année de survenance ont été construits en pourcentage. La distinction est effectuée entre les décès directs et les décès suite à exonération, bien que ces derniers ne soient pas reliés à l'arrêt de travail ayant causé le décès, comme expliqué précédemment. Les triangles sont présentés en annexe (fig. 1 et fig. 2).

Pour les décès directs, plus de 93% des dossiers sont en moyenne saisis avant 2 ans. Plus de 97% le sont avant 3 ans. L'évolution ralentie fortement cependant après la moitié de la deuxième année.

Pour les décès suite à arrêt de travail, 95% des dossiers sont en moyenne saisis avant 2 ans. Plus de 98% le sont avant 3 ans. L'évolution semble assez similaire en moyenne à celle des décès directs.

Ces résultats similaires sont en cohérence avec le fait que ces 2 types de décès ne sont pas différenciés, pour le moment, dans les comptes de résultats.

Ainsi, cela suppose que l'assureur a connaissance de la quasi-totalité des décès un an après la fin de l'exercice comptable. Bien entendu, si les décès faisant suite à arrêt de travail étaient rattachés à l'année de survenance de l'arrêt de travail alors ce délai serait plus long pour ce type de décès.

Cependant, avec les déclarations DSN et les données Insee, les décès peuvent être détectés plus tôt. En effet, seulement quelques mois de latence sont présents dans ces types de déclarations. Aux vues des résultats obtenus, ces nouvelles sources de données permettent donc à l'assureur d'avoir connaissance du décès plus rapidement.

4.3 Analyse de l'évolution des décès sur un modèle simple

Dans cette partie, un exemple de portefeuille est projeté sur plusieurs années avec des changements d'états possibles pour les individus. L'objectif est de faire un modèle assez simple avec le moins de paramètres possible mais permettant d'observer les flux entre les différents états possibles.

L'exemple est projeté sur 10 ans en partant d'une population de 10 000 salariés actifs. Les différents états possibles sont :

- Actif
- En arrêt de travail
 - En incapacité (distinction des 3 années de maintien en incapacité possibles)
 - En invalidité (pas de retour possible à l'état actif)

- Décédé
 - Décès direct
 - Décès suite à exonération
- Départ en retraite à la suite de l'incapacité

Le modèle est fermé, les seules causes de sorties possibles sont le décès ou le départ à la retraite.

La population totale des personnes actives et en incapacité est maintenue à un niveau stable, ainsi les décès (directs, décès suite à incapacité et les décès suite à invalidité dont les années de survenance du décès et de l'invalidité sont les mêmes), les départs à la retraite et les passages en invalidité sont compensés par de nouveaux entrants actifs. Cela permet de supposer que l'âge moyen reste stable au cours des années de développement. Ainsi, cela permet de considérer les taux de passage comme étant constants et par conséquent de simplifier le modèle.

De nombreuses hypothèses ont dû être prises. Notamment en ce qui concerne les taux d'incidence des différents états. Les taux de passage en incapacité et ceux de décès suite à incapacité ont été choisis de manière hypothétique mais réfléchi.

Le taux de passage en invalidité provient des résultats d'une étude menée dans un précédent mémoire [8].

Les taux de passage d'incapacité à invalidité ont été déterminés par rapport aux résultats provenant des tableaux de bord de GGVIÉ.

Les taux de sortie d'incapacité, et donc de retour à l'état actif, ont été calculés à partir de la loi du BCAC sur le maintien en incapacité pour un individu de 40 ans. Les taux de milieu d'année ont été utilisés afin d'éviter les déformations dues à l'entrée dans l'état (nombreuses sorties dans les premiers mois) ou à la sortie légale du risque (la Sécurité Sociale impose le passage en invalidité après 3 ans en incapacité). A noter que pour le taux de sortie de l'année N , une franchise de 3 mois a été prise. La probabilité de décès direct provient de la table réglementaire TH 00-02 pour l'âge de 46 ans. Cet âge est l'âge actuariel des salariés.

La probabilité de décès suite à invalidité provient également de la table TH 00-02 mais pour l'âge de 51 ans ici, cela correspond donc à 5 ans de plus que l'âge actuariel. Cela permet de prendre en compte l'aggravation du risque due à l'état de l'individu.

Enfin, le taux de sortie d'invalidité, pour cause de passage à la retraite, correspond au taux de sortie d'invalidité pour l'âge de 52 ans car il s'agit de l'âge moyen des personnes en invalidité dans le portefeuille de GGVIÉ.

Ces taux sont supposés intervenir en milieu d'année. Le calcul des différents passages entre les états est ordonné, comme le montre le schéma suivant (fig. 4.15). Le nombre de décès direct est calculé en premier sur la population d'actifs, puis dans la population active restante le nombre de passage en incapacité est calculé. Le nombre d'invalides est calculé à partir de la population d'actifs restante. Puis le nombre de décès suite à arrêt de travail est calculé à partir des populations en incapacité et invalidité. Enfin, le retour à l'état actif sur les individus restants en incapacité est calculé et le départ à la retraite des individus restants en invalidité est calculé.

Ce modèle prend donc en compte les multiples changements d'états que peut subir un individu dans une même année. En effet, des individus peuvent tomber en incapacité l'année N et décéder lors de cette même année, ou bien passer en invalidité ou encore redevenir actifs.

Les individus en incapacité depuis 3 ans présents à la fin de l'année sont automatiquement passés en invalidité, conformément à la définition de la Sécurité Sociale. Cela est représenté par l'état 'Incap $N+3$ ' dans la matrice de transition (fig. 4.16).

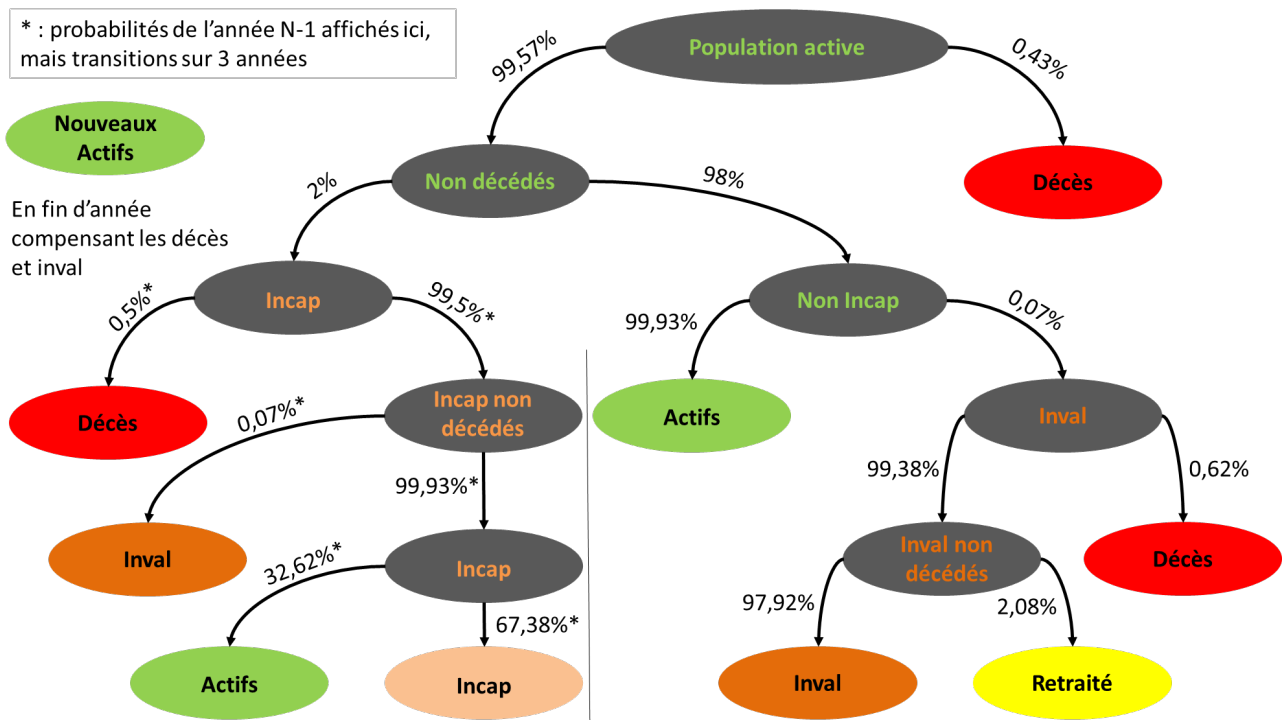


FIGURE 4.15 – Schéma de l'ordre entre les états et des taux de passage entre les différents états

Taux		Etats	Actifs	Incap N	Incap N+1	Incap N+2	Incap N+3	Inval	Décès	Retraite
Population de départ	10 000	Actifs	97,51%	1,99%	0,00%	0,00%	0,00%	0,07%	0,43%	0,00%
Passage en incapacité	2,00%	Incap N	43,11%	0,00%	54,86%	0,00%	0,00%	0,03%	2,00%	0,00%
Passage en invalidité	0,07%	Incap N+1	32,81%	0,00%	0,00%	66,62%	0,00%	0,07%	0,50%	0,00%
Passage d'incap à inval en N	0,03%	Incap N+2	26,90%	0,00%	0,00%	0,00%	72,73%	0,07%	0,30%	0,00%
Passage d'incap à inval en N-1	0,07%	Incap N+3	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%	0,00%	0,00%
Passage d'incap à inval en N-2	0,07%	Inval	0,00%	0,00%	0,00%	0,00%	0,00%	97,31%	0,62%	2,07%
Sortie d'incapacité en N	44,00%	Décès	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%	0,00%
Sortie d'incapacité en N-1	33,00%	Retraite	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	100,00%
Sortie d'incapacité en N-2	27,00%									
Proba de décès direct	0,43%									
Proba de décès si en incap en N	2,00%									
Proba de décès si en incap en N-1	0,50%									
Proba de décès si en incap en N-2	0,30%									
Proba de décès si en invalidité	0,62%									
Sortie d'invalidité	2,08%									

FIGURE 4.16 – Représentation des taux d'incidence et de la matrice de transition entre les différents états

Les taux utilisés en paramètres sont présentés à gauche. Cependant, afin de faciliter la lecture, ces taux sont retranscrits en taux de passage entre les différents états, prenant en compte l'ordre d'application des différents taux et la façon dont sont appliqués les différents taux en hypothèse.

Un résumé des résultats du modèle est présenté ici mais le modèle détaillé peut être retrouvé en annexe (fig. 3).

Vision au 31/12		N	N+1	N+2	N+3	N+4	N+5	N+6	N+7	N+8	N+9	N+10
Actifs		9888,48	9815,01	9762,12	9763,86	9764,64	9764,91	9764,89	9764,88	9764,88	9764,88	9764,88
	AT	111,52	184,99	237,88	236,14	235,36	235,09	235,11	235,12	235,12	235,12	235,12
Incapacité		6,76	13,38	19,87	79,60	137,48	193,78	248,62	302,32	354,92	406,42	456,85
	Décès	43,00	42,52	42,20	41,98	41,98	41,99	41,99	41,99	41,99	41,99	41,99
Direct		4,03	4,58	4,81	5,16	5,52	5,88	6,23	6,57	6,90	7,22	7,54
	Exo	0,10	0,20	0,29	1,18	2,05	2,89	3,70	4,50	5,29	6,06	6,81
Retraite		0,10	0,20	0,29	1,18	2,05	2,89	3,70	4,50	5,29	6,06	6,81
Taille de la nouvelle population		10006,76	10013,38	10019,87	10079,60	10137,48	10193,78	10248,62	10302,32	10354,92	10406,42	10456,85

FIGURE 4.17 – Résumé des résultats du modèle

A noter, l'augmentation importante des invalides entre les années $N + 2$ et $N + 3$ due au passage forcé des individus en incapacité depuis 3 ans en invalidité.

Ainsi, la population en arrêt de travail augmente au fil des années. Par conséquent, les décès suite à arrêt de travail augmentent également dans le temps, une fois que la réserve d'individus en arrêt de travail est assez importante. Le nombre de décès directs diminue légèrement dû à la petite diminution des actifs.

Un point important, est que dans le portefeuille de GGVI, le décès suite à exonération représente plus de 40% des décès. Alors que dans ce modèle après 10 ans d'évolution, les décès suite à exonération ne représentent que 15% des décès. Mais, les décès suite à exonération augmentent alors que les décès directs diminuent, donc la tendance aboutira au résultat observé sur le portefeuille après encore plusieurs années de simulation. Une autre raison est peut-être une surestimation du taux de décès direct. Ce dernier point étant l'objet de ce mémoire il sera développé par la suite.

Des améliorations sont donc envisageables afin de rendre ce modèle plus proche de la réalité, comme une paramétrisation plus fine des taux. Afin de réaliser cela sur le taux de décès direct, la modélisation du décès direct va maintenant être abordée sur les données du portefeuille.

Chapitre 5

Modélisation

5.1 Présentation des modèles

Cette partie a pour but l'explication des principales notions théoriques ainsi que des notations utilisées dans les différents modèles. Pour finir, les différentes méthodes seront expliquées d'un point de vue théorique.

5.1.1 Fonctions liées à la mortalité

Premièrement, les différentes fonctions utilisées afin de représenter la mortalité sont définies et explicitées.

5.1.1.1 Fonction de survie

Notons T une variable aléatoire positive, qui correspond dans ce cadre à la durée de vie résiduelle. La fonction de survie $S(t)$, $t \geq 0$, est définie comme le complémentaire de la fonction de répartition de T , notée $F(t)$:

$$S(t) = 1 - F(t) = 1 - \mathbb{P}(T < t) = \mathbb{P}(T \geq t)$$

La densité de probabilité, $f(t)$ se déduit de cette fonction de survie, il s'agit de la fonction dérivée en un instant dt de la fonction de répartition T :

$$f(t) = -\frac{d}{dt}S(t)$$

5.1.1.2 Force de mortalité

Cette fonction, notée μ_X , également nommée taux de risque instantané, décrit localement l'intensité de mortalité par rapport aux caractéristiques X . Plus précisément, il s'agit de la probabilité de décès dans un petit intervalle de temps rapportée à la longueur de cet intervalle.

$$\mu_X = \frac{1}{h} \lim_{h \rightarrow 0} \mathbb{P}(T_X < h) = f_X(0)$$

La fonction de survie peut être calculée à partir de cette valeur :

$$S_X(t) = \exp\left(\int_{u=0}^t \mu_{X+u} du\right)$$

5.1.1.3 Lien avec la probabilité de décès

La probabilité conditionnelle de décès à t années pour un individu d'âge x est définie comme étant :

$${}_tq_X = \mathbb{P}(0 \leq T_X < t) = F_X(t)$$

La probabilité conditionnelle de décès à un an est notée q_X et correspond à la formule ci-dessus pour $t = 1$:

$$q_X = \mathbb{P}(T_X < 1)$$

Ainsi, la relation suivante relie q_X à la force de mortalité :

$$q_X = 1 - S_X(1) = 1 - \exp\left(-\int_{u=0}^1 \mu_{X+u} du\right)$$

L'hypothèse de constance par morceaux sur des intervalles d'un an entre deux âges entiers de la force de mortalité est prise dans cette étude.

Ainsi, cela se traduit par : $\mu_{X+u} = \mu_X, \forall 0 \leq u < 1$. Cela permet de simplifier la relation précédente, de sorte que :

$$q_X = 1 - \exp(-\mu_X)$$

5.1.2 Censure et Troncature

La censure et la troncature sont présentes dans cette étude étant donné que le phénomène étudié n'est pas directement observé mais qu'une autre variable aléatoire est observée, celle-ci étant liée au phénomène étudié.

Ainsi, dans le cas du décès, le phénomène étudié est le décès et la variable d'intérêt est donc la date de décès. Or, la période d'étude est trop courte pour observer le décès de chaque individu. En effet, les individus sont observés lorsqu'ils sont dans le portefeuille et non pas depuis leur naissance. De plus, une période d'étude doit être choisie, dans ce mémoire elle est choisie en fonction de la disponibilité des données. Ainsi, la variable observée est la date de fin d'observation de l'individu. Mais cette fin d'observation peut ne pas être due au décès de l'individu mais plutôt à son départ des effectifs de l'entreprise par exemple ou à la fin de la période d'étude.

5.1.2.1 Censure

La censure implique que les données sont incomplètes. Deux phénomènes différents peuvent être identifiés.

La censure à droite implique que l'évènement étudié ne s'est pas produit au moment de la fin de l'observation. Dans le cas du décès, si une personne quitte les effectifs de l'entreprise, alors sa date de décès est inconnue mais cet évènement se réalisera forcément à un moment donné dans le futur. L'information connue est alors la date à laquelle l'individu est encore en vie.

Dans le cas de la censure à droite, deux types de censures sont à distinguer : la censure fixe et la censure aléatoire. La première correspond, dans cette étude, à la date d'arrêt des données, qui est le 31 décembre 2021. La deuxième correspond, dans le cadre de cette étude, à la résiliation du contrat par l'entreprise ou la sortie des effectifs de l'individu, cela peut être due à une fin de contrat (pour cause de retraite, licenciement, démission ...).

La censure à gauche apparaît lorsque l'évènement de fin s'est déjà produit. Dans le cas du décès, cela suppose que le décès est survenu mais que la date de décès est inconnue. Ce type de cas n'est pas présent dans cette étude suite aux traitements de fiabilisation des données effectués. En effet, tous les décès recensés ont une date de décès connue.

5.1.2.2 Troncature

La troncature correspond au fait que la variable étudiée n'est observable que conditionnellement à un évènement. Ici encore, deux phénomènes sont identifiables.

La troncature à gauche est la plus répandue car cela implique que la variable étudiée n'est observable que si elle est supérieure à un certain seuil. Ce type de troncature est présent dans cette étude car les individus ne sont pas observés depuis leur naissance. Les données sont donc tronquées au 1er janvier 2018 dans ce mémoire. Ainsi, les décès survenus avant cette date ne sont jamais observés. Aucune information n'est connue sur ces sinistres. En particulier, leur proportion est donc inconnue.

Par opposition, la troncature à droite implique que la variable étudiée n'est observable que si elle est inférieure à un certain seuil. Ce cas n'est pas présent dans ce mémoire.

Le schéma suivant dépeint plusieurs parcours possibles. L'observation de l'individu 1 est tronquée car il décède avant le début de la période d'observation, ainsi son décès n'est jamais observé.

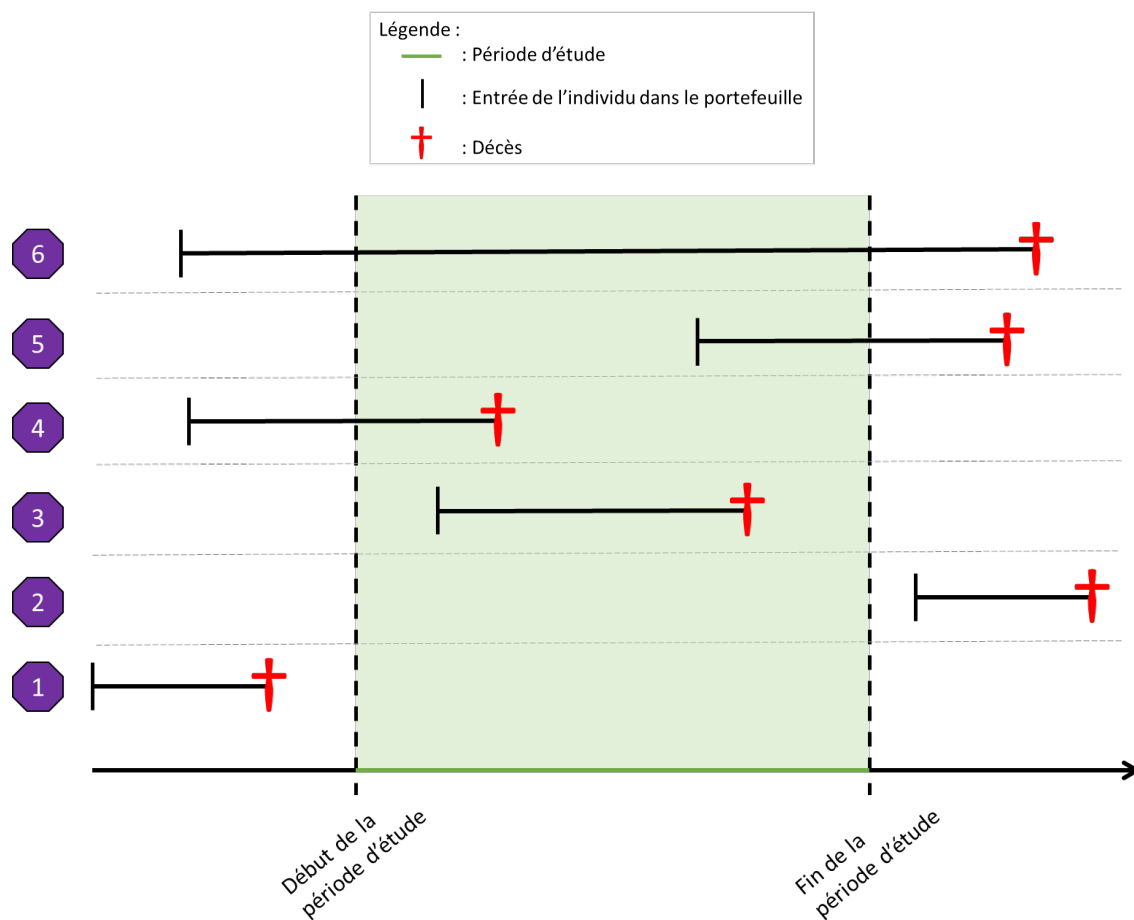


FIGURE 5.1 – Schéma explicatif de la censure et troncature

L'individu 2 n'est également pas observé car il entre dans le portefeuille après la fin de la période d'observation.

L'individu 3 est observé, il rentre dans le portefeuille au cours de la période d'étude et décède durant cette dernière. Son décès est donc observé. Cette observation est complète.

L'observation de l'individu 4 est également complète car son décès a lieu pendant la période d'étude. Cet individu était présent dans le portefeuille avant le début de l'étude.

Les observations des individus 5 et 6 sont incomplètes car leurs décès n'ont pas lieu pendant la période

étudiée, mais ils sont observés dans le portefeuille durant cette période. Ainsi, la seule information connue est qu'ils ont survécu jusqu'à la fin de la période d'étude. Leurs décès sont donc censurés par la fin de la période d'étude.

5.1.3 Taux brut de mortalité

Tout d'abord, la vraisemblance est détaillée afin d'expliquer l'approche et le résultat obtenu.

La vraisemblance associée aux données individuelles présentant de la censure à droite et de la troncature à gauche s'écrit :

$$L(\theta) = \prod_{i=1}^n f_{X_i}(t_i, \theta)^{\delta_i} S_{X_i}(t_i, \theta)^{1-\delta_i}$$

θ étant le vecteur des paramètres et δ correspond au vecteur des décès donc il vaut 1 en cas de décès observé et 0 sinon. Les autres notations ayant déjà été introduites précédemment.

Cette formule est donc séparée en 2 composantes : les individus censurés et ceux non-censurés. Dans le premier cas, la fonction de survie est utilisée pour prendre en compte l'information observée, à savoir que l'individu a survécu jusqu'à la date de fin d'observation, dès lors son décès n'a pas été observé. Dans le second cas, le décès a bien été observé ainsi l'observation est complète, c'est pourquoi la fonction de densité est utilisée.

Rappelons que la force de mortalité est supposée constante par morceaux entre deux âges entiers. La log-vraisemblance s'écrit donc :

$$\begin{aligned} l(\theta) &= \ln \left(\prod_{i=1}^n f_{X_i}(t_i, \theta)^{\delta_i} S_{X_i}(t_i, \theta)^{1-\delta_i} \right) \\ &= \sum_{i=1}^n \ln(f_{X_i}(t_i, \theta)) \times \delta_i + \ln(S_{X_i}(t_i, \theta)) \times (1 - \delta_i) \\ &= \sum_{i=1}^n \delta_i \times \ln \left(\frac{f_{X_i}(t_i, \theta)}{S_{X_i}(t_i, \theta)} \right) + \ln(S_{X_i}(t_i, \theta)) \\ &= \sum_{i=1}^n \delta_i \times \ln(\mu_{X_i, t_i}(\theta)) - \int_{u=0}^{t_i} \mu_{X_i+u}(\theta) du \\ &= \sum_X d_X \ln(\mu_X(\theta)) - \mu_X(\theta) e_X^c \\ &= \ln(\mu(\theta))^t d - \mu(\theta)^t e^c \end{aligned}$$

Avec d_X correspondant au nombre de décès d'individus ayant les caractéristiques X et e_X^c correspond à l'exposition centrale au risque, qui est égale au nombre total d'années d'observation des individus possédant les caractéristiques X . L'écriture avec X permet de passer de la vision individuelle de la base à la vision agrégée par rapport aux caractéristiques, par conséquent l'information est condensée.

L'hypothèse suivante est prise : $\ln(\mu(\theta)) = X\theta$. Cela revient à supposer que la force de mortalité se comporte de façon exponentielle en fonction de l'âge. Afin de minimiser cette log-vraisemblance, la dérivée de cette log-vraisemblance est calculée :

$$\frac{\partial l}{\partial \theta} = X^t (d - e^{X\theta} * e^c)$$

L'estimateur des taux bruts est issu du cas particulier du modèle saturé. Dans ce cas la matrice X est

l'identité et ainsi en cherchant à annuler la dérivée de la log-vraisemblance on obtient :

$$\frac{\partial l}{\partial \theta} = 0 = d - \mu * e^c$$

Ainsi, l'équation suivante est obtenue :

$$\mu = \frac{d}{e^c}$$

L'estimateur des taux bruts est donc $\hat{\mu} = \frac{d}{e^c}$. Cet estimateur est simple à calculer et représente donc une première approche afin de visualiser la forme de la mortalité. Cependant, d'autres méthodes vont être étudiées afin de comparer les résultats.

5.1.4 Kaplan-Meier

Un des estimateurs les plus utilisés afin de modéliser une loi de mortalité est l'estimateur de Kaplan-Meier. Ce dernier est un estimateur non-paramétrique de la fonction de survie. Cette propriété est très appréciable car elle permet de ne pas faire d'hypothèse sur la distribution. Ainsi, dans cette étude, cet estimateur est utilisé comme première approche afin d'observer la distribution générale des données. Cet estimateur a également l'avantage de prendre en compte la censure.

Mais des hypothèses sont nécessaires pour appliquer ce modèle. La principale hypothèse est que la censure doit être indépendante de la variable de durée étudiée. Cependant, cette hypothèse est vérifiée ici car aucun lien ne semble exister entre la date de fin de sortie des effectifs ou la date de fin d'étude et la date de décès des individus.

Les notations suivantes sont utilisées :

- T = variable de durée étudiée (date de décès ici)
- C = variable représentant la censure (date de fin d'observation ici pour une cause autre que le décès)
- δ = vecteur indiquant si l'évènement étudié a été observé (le décès ici)
- $Y = \inf(T, C)$ = variable réellement observée due à la censure (date de fin d'observation de l'individu ici)

La fonction de répartition de Y , notée H , peut s'écrire :

$$H(t) = \mathbb{P}(Y \leq t) = \mathbb{P}(X \leq t; \delta = 0) + \mathbb{P}(X \leq t; \delta = 1) = H_0(t) + H_1(t)$$

Ces différentes fonctions sont facilement estimables :

$$\begin{aligned} \hat{H}(t) &= \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{Y_i \leq t} \\ \hat{H}_0(t) &= \frac{1}{n} \sum_{i=1}^n (1 - \delta_i) \mathbb{1}_{Y_i \leq t} \\ \hat{H}_1(t) &= \frac{1}{n} \sum_{i=1}^n \delta_i \mathbb{1}_{Y_i \leq t} \end{aligned}$$

Le but est d'exprimer F , la fonction de répartition de T , comme une fonction de H , H_0 et H_1 . Cette démarche cherche à estimer F par une loi discrète car la continuité de F est incertaine, de plus une loi discrète est plus facile à manipuler.

Ainsi, dans le cas discret, le taux de risque instantané est défini par :

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

De cette expression, la fonction de répartition peut être déduite du taux de risque instantané par :

$$1 - F(t-) = \prod_{t_i < t} (1 - \mu_F(t_i))$$

Par indépendance entre T et C , on obtient :

$$\mu_F(t) = \frac{dH_1(t)}{1 - H(t-)}$$

Puis par "plug-in", l'estimateur de μ est :

$$\hat{\mu}_F(t) = \frac{d\hat{H}_1(t)}{1 - \hat{H}(t-)}$$

Puis par les formules explicitées précédemment, l'estimateur de Kaplan-Meier est obtenu :

$$1 - \hat{F}(t) = \prod_{Y_i < t} (1 - \hat{\mu}_F(Y_i)) = \prod_{Y_i < t} \left(1 - \frac{d\hat{H}_1(Y_i)}{1 - \hat{H}(Y_i-)} \right)$$

$$\hat{F}(t) = 1 - \prod_{Y_i < t} \left(1 - \frac{\delta_i}{\sum_{j=1}^n \mathbb{1}_{Y_j \geq Y_i}} \right)$$

Cet estimateur ne met donc du poids que sur les observations non censurées.

5.1.5 GLM

Maintenant on s'intéresse aux modèles linéaires généralisés ou GLM, méthode très employée en assurance non-vie. Notons Y la variable aléatoire d'intérêt et X les variables explicatives. Un GLM est défini par 3 composantes :

- $Y|X$, la variable modélisée, supposée suivre une loi appartenant à une famille exponentielle
- g , la fonction de lien qui est bijective
- θ , le vecteur de paramètres qui est estimé

La densité d'une loi exponentielle est de la forme suivante :

$$f_{\beta, \phi}(y) = c_{\phi}(y) \exp\left(\frac{y\beta - a(\beta)}{\phi}\right)$$

- β s'appelle le paramètre canonique
- ϕ le paramètre de dispersion
- $a(\beta)$ est de classe C^2 et convexe
- $c_{\phi}(y)$ ne dépend pas de β

Dans le cas d'un GLM, $Y|X$ suit donc une loi appartenant à une famille exponentielle, ainsi l'espérance est égale à :

$$\mathbb{E}[Y|X] = a'(\beta)$$

La fonction de lien permet de lier l'espérance de $Y|X$ à la partie modélisée :

$$g(\mathbb{E}[Y|X]) = X\theta$$

Par simplification la fonction de lien canonique associée à la loi utilisée est employée. Cette fonction de lien, permet en effet de simplifier l'écriture du modèle. En regroupant les équations ci-dessus on obtient :

$$g(\mathbb{E}[Y|X]) = g(a'(\beta)) = X\theta$$

Ainsi, en prenant g comme étant la fonction inverse, c'est-à-dire $g = (a')^{-1}$, qui correspond donc à la fonction de lien canonique, cela se simplifie par :

$$\beta = X\theta$$

L'estimation du vecteur des paramètres θ se fait par maximum de vraisemblance. En supposant les Y_i indépendants, la log-vraisemblance s'écrit :

$$l(\theta) = \sum_{i=1}^n \log(f(Y_i, \theta, \phi)) = \sum_{i=1}^n \left(\log[c_\phi(Y_i)] + \frac{Y_i \beta_i - a(\beta_i)}{\phi} \right)$$

Afin de trouver le minimum, la dérivée par rapport à θ est calculée, ce qui est égale, après calculs, à :

$$\frac{\partial l}{\partial \theta_j} = \frac{1}{\phi} \sum_{i=1}^n \frac{x_{i,j}(Y_i - \mathbb{E}(Y_i|X_i))}{g'(\mathbb{E}(Y_i|X_i))a''(\beta_i)}$$

La résolution de ce problème peut être approchée par l'algorithme de Newton-Raphson.

5.1.5.1 Modèle de Gompertz

Ce modèle date de 1825 et est encore très utilisé de nos jours. Ce modèle stipule que la mortalité croît exponentiellement avec l'âge :

$$\mu_x = e^{a+bx}$$

où x correspond à l'âge, a varie en fonction du niveau de mortalité et b correspond à l'augmentation du risque de décès avec l'âge.

L'avantage de ce modèle est qu'il possède peu de paramètres. Ce modèle est adapté à la modélisation de la population salariale, car il est généralement utilisé pour modéliser des populations de plus de 30 ans.

Cependant, ce modèle a tendance à sous-estimer les décès avant 40 ans et à surestimer ceux après 80 ans.

5.1.5.2 GLM Poisson

Les notations introduites précédemment dans la partie GLM sont reprises ici.

Il s'agit ici de se placer dans un cas particulier de GLM. En effet, ici $Y|X$ est supposé suivre une loi de Poisson de paramètre λ , avec $y_i \in \{0; 1\}$.

La fonction de densité s'écrit donc :

$$\mathbb{P}(Y_i = y_i | X_i = x_i) = \frac{\lambda^{y_i}}{y_i!} e^{-\lambda}$$

Une propriété notable de la loi de Poisson est l'égalité entre l'espérance et la variance :

$$\mathbb{E}[Y|X] = \text{Var}[Y|X] = \lambda$$

La fonction de lien canonique pour cette loi est la fonction logarithme. Ainsi, le modèle s'écrit :

$$\ln(\lambda) = \ln(\mathbb{E}[Y|X]) = X\theta$$

La log-vraisemblance dans ce cas est égale à :

$$\begin{aligned} l(\theta) &= \sum_{i=1}^n \ln(\mathbb{P}(Y_i = y_i | X_i = x_i)) \\ &= \sum_{i=1}^n \ln\left(\frac{\lambda^{y_i} e^{-\lambda}}{y_i!}\right) \\ l(\theta) &= \sum_{i=1}^n (y_i \ln(\lambda) - \ln(y_i!) - \lambda) \end{aligned}$$

Dans le cadre de cette étude sur le décès, la variable Y correspond au décès, dont le vecteur des observations sera noté D , X correspond aux différentes variables explicatives comme l'âge, le sexe, la CSP ou encore le niveau de salaire. D'après l'approximation vu dans la partie sur le taux brut de décès, le paramètre λ est considéré comme étant $\mu(\theta)e^c$, c'est-à-dire le produit entre la force de mortalité et l'exposition centrale. En effet, l'hypothèse énonçant que $\mu(\theta) = e^{X\theta}$ est prise.

Ainsi, le modèle avec la fonction de lien \ln s'écrit :

$$\ln(\mathbb{E}(D)) = \ln(\mu(\theta)e^c) = \ln(\mu(\theta)) + \ln(e^c) = \ln(\mu(\theta)) + \text{offset} = X\theta + \text{offset}$$

où $\ln(\mu(\theta)) = X\theta$ correspond à la partie modélisée et $\ln(e^c) = \text{offset}$ correspond à un décalage appliqué au vecteur d'observations pour prendre en compte l'exposition qui est différente pour chaque individu.

5.1.5.3 GLM Poisson avec référence

Dans cette partie, la méthode GLM Poisson est reprise avec une modification permettant de transformer cette méthode, en méthode référentielle. En effet, la loi de référence est ajoutée au modèle afin d'obtenir des coefficients qui s'interprètent par rapport à cette dernière. Il s'agit donc d'une méthode de positionnement par rapport aux lois de référence.

Comme précédemment un GLM de Poisson est réalisé avec les mêmes paramètres, c'est-à-dire que le paramètre λ est encore considéré comme étant $\mu(\theta)e^c$ et la fonction de lien est encore la loi logarithmique. Cependant, il a été vu que l'hypothèse énonçant que $\mu(\theta) = e^{X\theta}$ avait été prise. Désormais, cette hypothèse est modifiée, et on suppose que $\mu(\theta) = \mu_{ref} e^{X\theta}$, avec μ_{ref} qui correspond à la force de mortalité de la loi de référence.

Ainsi, le modèle s'écrit :

$$\ln(\mathbb{E}(D)) = \ln(\mu(\theta)e^c) = \ln(\mu_{ref} e^{X\theta}) + \ln(e^c) = \ln(e^{X\theta}) + [\ln(\mu_{ref}) + \ln(e^c)] = X\theta + \text{offset}$$

Ainsi, la partie appelé offset est modifiée par rapport au modèle classique avec l'ajout du logarithme de la force de mortalité de la loi de référence.

L'interprétation des coefficients obtenus est effectuée de la manière suivante. L'exponentielle des coefficients, $e^{X\theta}$, s'interprète comme le ratio entre la force de mortalité observée et celle de référence, $\frac{\mu(\theta)}{\mu_{ref}}$.

A noter que comme la force de mortalité est supposée constante par âge, la relation suivante permet de passer aux probabilités de décès dans l'année :

$$q_x = 1 - e^{-\mu_x}$$

Cette relation permet également d'obtenir la force de mortalité à partir des probabilités de décès, cette relation est notamment utilisée pour les lois de référence :

$$\mu_x = -\log(1 - q_x)$$

Cette méthode permet d'obtenir la significativité des coefficients ainsi que des intervalles de confiance.

5.1.5.4 Sélection des variables et regroupements des modalités

Afin de sélectionner les variables significatives au modèle le test de Wald est utilisé. Ce test considère les deux hypothèses suivantes :

- H_0 : le coefficient θ_i est nul
- H_1 : le coefficient θ_i est non nul

La statistique de Wald est :

$$S_W = \frac{\hat{\theta}_i^2}{\text{Var}(\hat{\theta}_i)}$$

Sous H_0 , S_W suit une loi du χ^2 à 1 degré de liberté.

SAS fournit cette statistique ainsi que la p-value associée. Ainsi, la p-value de ce test permettra de déterminer la significativité ou non du coefficient calculé.

En ce qui concerne la modélisation, dans un premier temps, le modèle est implémenté avec tous les critères retenus pour cette étude. Puis, les variables non significatives selon le test de Wald sont écartées. Pour les variables restantes, si des modalités sont non significatives alors elles sont regroupées avec la modalité significative ayant le coefficient le plus proche.

Le modèle est à nouveau implémenté et les étapes précédentes sont à nouveau réalisées si nécessaire et cela jusqu'à ce que toutes les variables et toutes les modalités soient significatives.

Cette approche se nomme "backward selection".

5.1.5.5 Sélection et validation du modèle

Plusieurs critères et tests existent afin de choisir entre plusieurs modèles. Ces indicateurs permettent de mesurer la qualité de l'ajustement afin de comparer différents modèles. Les tests utilisés dans cette étude sont présentés.

Dans un premier temps, la déviance normalisée Dev^* sera analysée. Elle permet de mesurer la qualité de l'ajustement. Cette statistique compare le modèle implémenté avec le modèle saturé. L'inconvénient est qu'elle ne prend pas en compte le surajustement des données. La déviance est définie par :

$$Dev^* = 2 \times (\ln(L_{sature}) - \ln(L))$$

où, L_{sature} correspond à la vraisemblance du modèle saturé et L à celle du modèle considéré.

Lorsque le modèle est bien ajusté, la déviance normalisée suit une loi du khi-deux à $n - p$ degrés de liberté. n correspond au nombre d'observations et p au nombre de paramètres du modèle. La p-value sera calculée pour chaque modèle afin de déterminer la qualité de l'ajustement.

La déviance normalisée s'écrit également comme :

$$Dev^* = \frac{Dev}{\phi}$$

où, Dev correspond à la déviance et ϕ est un paramètre de dispersion. Ce paramètre peut être estimé par :

$$\hat{\phi} = \frac{Dev}{n - p}$$

Dans ce cas, le test consiste à regarder si cette valeur est proche de 1. Si c'est le cas alors le modèle ajuste bien les données car il n'y a pas de sur-dispersion ou sous-dispersion. Ce test sera réalisé pour chaque modèle afin de vérifier la qualité de l'ajustement.

Dans un deuxième temps, la statistique de Pearson sera étudiée. Comme la déviance, elle permet de juger de la qualité de l'ajustement des données. Elle est définie par :

$$\chi^2 = \sum_{i=1}^n \frac{(Y_i - \hat{\lambda}_i)^2}{\text{Var}(\hat{\lambda}_i)}$$

De même que précédemment, le khi-deux de Pearson normalisée est définie par $\frac{\chi^2}{\phi}$. Et cette statistique suit une loi du khi-deux à $n - p$ degrés de liberté en cas de bon ajustement des données.

De plus, ici encore, le paramètre ϕ peut être estimé par :

$$\hat{\phi} = \frac{\chi^2}{n - p}$$

D'autre part, le critère AIC (Akaike Information Criterion) sera également utilisé. Le meilleur modèle est alors celui avec la valeur la plus faible. Ce critère est défini par :

$$AIC = -2l(\theta) + 2p$$

où $l(\theta)$ est la log-vraisemblance et p correspond au nombre de paramètres du modèle. La première partie de cette formule correspond à la déviance. Cet estimateur permet de prendre en compte le sur-apprentissage afin de conserver un modèle qui peut être généralisé facilement à de nouvelles données.

Pour finir les résidus sont analysés. Ces indicateurs permettent de juger de la qualité du modèle en comparant les valeurs observées et les valeurs prédites. Les résidus de la réponse correspondent simplement à la différence entre les valeurs observées et les valeurs prédites. Cependant, ils sont difficilement interprétables.

C'est pourquoi les résidus de Pearson sont introduits :

$$r_{P_i} = \frac{Y_i - \hat{\lambda}_i}{\sqrt{\text{Var}(\hat{\lambda}_i)}}$$

En cas de bon ajustement, les résidus sont compris entre -2 et 2 .

Un autre type de résidus est introduit, les résidus de la déviance :

$$r_{Dev_i} = \sqrt{Dev_i} \times \text{signe}(Y_i - \hat{\lambda}_i)$$

En cas de bon ajustement, les signes des résidus alternent fréquemment.

Afin de faciliter les analyses, les résidus standardisés seront utilisés.

5.1.6 SMR par critère

Dans un premier temps la définition du SMR est présentée puis la méthode mise en place sera expliquée.

Cette méthode permet d'obtenir des coefficients d'abattelements en segmentant la population itérativement selon plusieurs critères tout en neutralisant les effets des critères précédents à chaque nouvelle itération. Il s'agit donc d'une méthode de positionnement par rapport aux lois de référence.

5.1.6.1 Définition du SMR

Le SMR (Standardized Mortality Rate), ou indicateur standardisé de mortalité¹, est défini comme le ratio entre le nombre de décès observés et le nombre de décès attendus théoriquement.

Ainsi, si ce ratio est inférieur à 1 alors la mortalité observée est inférieure à la mortalité de référence. En revanche, si le ratio est supérieur à 1 alors la mortalité observée est supérieure à la mortalité de référence.

Les notations suivantes vont être utilisées :

- D_x , nombre de décès observé à l'âge x
- e_x^c , exposition de l'âge x
- $q_{x,theorique}$, probabilité de décès de référence

Le SMR est défini comme :

$$SMR = \frac{\sum_x D_x}{\sum_x (e_x^c \times q_{x,theorique})}$$

5.1.6.2 Méthode itérative

La méthode consiste donc à calculer ce ratio critère par critère sur une population corrigée, afin de voir l'influence de chaque critère indépendamment des autres. Les probabilités de mortalité de références sont ainsi réajustées à chaque itération afin de prendre en compte l'effet de la variable précédente. Contrairement à la méthode du GLM Poisson par référence celle-ci permet d'intégrer les variables une par une, cela permet donc d'analyser l'effet de chaque variable indépendamment des autres.

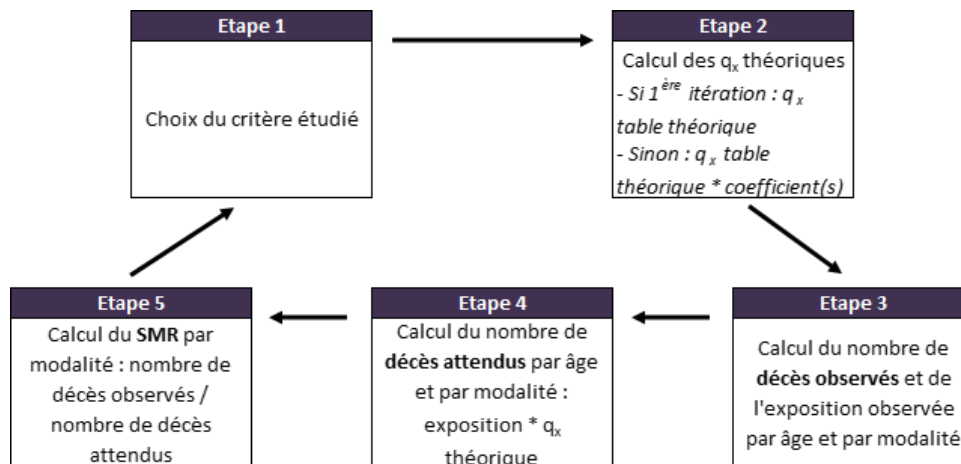


FIGURE 5.2 – Schéma explicatif de la méthode SMR par critère

Dans un premier temps, comme les tables de références utilisées ici sont séparées par sexe, c'est ce critère qui est pris en compte. De plus, il s'agit d'un des critères les plus discriminants pour le décès. Ainsi, le nombre de décès et l'exposition totale par sexe et par âge sont calculés. Cela permet le calcul du nombre de décès

1. <http://onala.free.fr/inegamor.pdf>

théoriques attendus qui est égal au produit entre l'exposition et le q_x théorique associé. Enfin, la somme par sexe du nombre de décès observés et du nombre de décès attendus est calculée, et pour finir le ratio SMR est évalué par sexe.

Ensuite, pour l'ajout d'un nouveau critère la table de données initiale est réutilisée en recalculant chaque q_x théorique en fonction du sexe de l'individu, en multipliant l'ancienne probabilité de décès théorique par le coefficient d'abattement obtenu précédemment. Cela permet donc de corriger l'effet du sexe, afin d'analyser l'effet d'une autre variable par la suite. Ainsi, des nouveaux q_x théoriques sont obtenus. Ensuite, les étapes précédemment décrites sont réitérées avec le nouveau critère choisi.

Afin de faciliter la compréhension de cette méthode, la méthode est implémentée sur un exemple simple. Deux itérations sont présentées afin de montrer en détail l'abattement des tables théoriques. Les probabilités de décès théoriques ont été multipliées par un facteur 100 afin d'avoir des coefficients SMR cohérents, le but étant d'expliquer la méthode et non d'être réaliste. Les étapes sont détaillées pas à pas et les données initiales sont dans un premier temps présentées.

Individu	Sexe	Age	CSP	Décès	Exposition	q_x théorique
1	Homme	45	Cadre	0	0,5	0,39704
2	Homme	45	Ouvrier	1	1	0,39704
3	Homme	45	Cadre	0	1	0,39704
4	Homme	55	Cadre	1	0,5	0,82083
5	Homme	55	Ouvrier	1	1	0,82083
6	Femme	45	Ouvrier	0	1	0,18040
7	Femme	45	Ouvrier	0	0,5	0,18040
8	Femme	45	Cadre	0	0,5	0,18040
9	Femme	55	Cadre	0	1	0,34566
10	Femme	55	Ouvrier	1	1	0,34566

Note : Les q_x théoriques ont été multipliés par 100 afin d'obtenir des coefficients non aberrants au vu du faible nombre d'individus et de décès dans cet exemple

FIGURE 5.3 – Données de départ de l'exemple simplifié de la méthode SMR par critère

1^{ère} itération :

Etape 1 : Sexe

Etape 2 : On reprend les probabilités de décès des tables théoriques ici

Etape 3 :

Sexe	Age	Décès	Exposition
Homme	45	1	2,5
Homme	55	2	1,5
Femme	45	0	2
Femme	55	1	2

Etape 4 :

Sexe	Age	Décès	Exposition	q_x théorique	Décès attendus
Homme	45	1	2,5	0,39704	0,99261
Homme	55	2	1,5	0,82083	1,23125
Femme	45	0	2	0,18040	0,36079
Femme	55	1	2	0,34566	0,69132

Etape 5 :

Sexe	Décès	Exposition	Décès attendus	SMR
Homme	3	4	2,22386	1,35
Femme	1	4	1,05211	0,95

FIGURE 5.4 – Première itération de l'exemple simplifié de la méthode SMR par critère

2^{ème} itération :

Etape 1: CSP

Etape 2 :

Individu	Sexe	Age	CSP	Décès	Exposition	q _x théorique
1	Homme	45	Cadre	0	0,5	0,53561
2	Homme	45	Ouvrier	1	1	0,53561
3	Homme	45	Cadre	0	1	0,53561
4	Homme	55	Cadre	1	0,5	1,10731
5	Homme	55	Ouvrier	1	1	1,10731
6	Femme	45	Ouvrier	0	1	0,17146
7	Femme	45	Ouvrier	0	0,5	0,17146
8	Femme	45	Cadre	0	0,5	0,17146
9	Femme	55	Cadre	0	1	0,32854
10	Femme	55	Ouvrier	1	1	0,32854

Etape 3 :

Sexe	Age	CSP	Décès	Exposition
Homme	45	Cadre	0	1,5
Homme	45	Ouvrier	1	1
Homme	55	Cadre	1	0,5
Homme	55	Ouvrier	1	1
Femme	45	Cadre	0	0,5
Femme	45	Ouvrier	0	1,5
Femme	55	Cadre	0	1
Femme	55	Ouvrier	1	1

Etape 4 :

Sexe	Age	CSP	Décès	Exposition	q _x théorique	Décès attendus
Homme	45	Cadre	0	1,5	0,53561	0,80842
Homme	45	Ouvrier	1	1	0,53561	0,53561
Homme	55	Cadre	1	0,5	1,10731	0,55366
Homme	55	Ouvrier	1	1	1,10731	1,10731
Femme	45	Cadre	0	0,5	0,17146	0,08573
Femme	45	Ouvrier	0	1,5	0,17146	0,25719
Femme	55	Cadre	0	1	0,32854	0,32854
Femme	55	Ouvrier	1	1	0,32854	0,32854

Etape 5 :

CSP	Décès	Exposition	Décès attendus	SMR
Cadre	1	3,5	1,77135	0,56
Ouvrier	3	4,5	2,22865	1,35

FIGURE 5.5 – Deuxième itération de l'exemple simplifié de la méthode SMR par critère

Par conséquent, cette méthode peut être itérée sur plusieurs critères en réajustant les probabilités de décès théoriques pour chaque nouveau critère. Cependant, une limite de cette méthode est que l'exposition de chaque modalité du critère étudié doit être assez élevée, car sinon le SMR est altéré et n'est plus fiable. De plus, cette méthode ne permet pas d'obtenir facilement d'intervalle de confiance ni de test de significativité des coefficients comme c'est le cas avec la méthode du GLM par référence. Ainsi, il est difficile de connaître la fiabilité des coefficients. Cependant, ces tests peuvent tout de même être réalisés. C'est pourquoi ces deux méthodes sont implémentées afin de les comparer et de fiabiliser les résultats obtenus.

L'avantage de cette méthode cependant est qu'elle permet d'étudier l'effet de chaque critère sans prendre en compte les autres critères et ainsi d'obtenir un coefficient d'abattement pour chaque critère.

5.1.6.3 Sélection et validation du modèle

Aucun test de significativité ne sera effectué ici, les critères significatifs étant déduits des modèles précédents.

La déviance sera calculée afin de vérifier le bon ajustement des données. Elle sera calculée à la main ici en supposant que les données suivent une loi de Poisson. Ainsi, la formule en supposant que tous les individus ont un poids de 1 est :

$$Dev_i = 2 \times \left(y_i \ln \left(\frac{y_i}{\hat{\lambda}_i} \right) - (y_i - \hat{\lambda}_i) \right)$$

$$Dev = \sum_i Dev_i$$

Dans le cadre de cette étude, le poids des individus est pris en compte par l'exposition. Le test permettant d'estimer le paramètre de dispersion sera donc effectué afin de vérifier la qualité de l'ajustement. Cette formule est également utilisée dans le calcul des résidus de la déviance.

En effet, afin de valider le modèle les résidus de Pearson et de la déviance sont calculés avec les mêmes formules que vu précédemment dans la section 5.1.5.5.

Finalement, des intervalles de confiance sont calculés. Par simplicité les décès sont supposés suivre une loi binomiale :

$$D_x \sim \mathcal{B}(e_x^c, q_x)$$

Cela entraîne par approximation de Gauss que :

$$\hat{q}_x \sim \mathcal{N} \left(q_x, \frac{q_x(1-q_x)}{e_x^c} \right)$$

où \hat{q}_x représente l'estimateur du taux de mortalité.

Cette approximation permet donc d'obtenir un intervalle de confiance à 95% facile à calculer :

$$q_x \in \left[\hat{q}_x - 1,96 \times \sqrt{\frac{\hat{q}_x(1-\hat{q}_x)}{e_x^c}}; \hat{q}_x + 1,96 \times \sqrt{\frac{\hat{q}_x(1-\hat{q}_x)}{e_x^c}} \right]$$

5.1.7 Lissage P-splines

Dans cette partie, le lissage P-splines est rapidement présenté. Dans un premier temps, la définition des B-splines ainsi que la construction d'une base à partir de ces objets est présentée. Puis l'implémentation de cette méthode afin de réaliser un lissage sera développée.

5.1.7.1 Définition des B-splines

Un B-spline de degré d est une fonction polynomiale par morceaux de degré d dont le support est constitué de $d + 1$ intervalles consécutifs dont les délimitations sont appelées noeuds. Ces noeuds sont supposés régulièrement espacés.

Une propriété intéressante est qu'un B-spline de degré d peut être construit de manière récursive à l'aide de 2 B-splines de degré $d - 1$.

5.1.7.2 Base de B-splines

Une base de B-splines est composée de B-splines qui se chevauchent et dont la somme en tout point du support est égale à 1.

Tout d'abord, afin de construire une base, le support des observations doit être divisé en n_d intervalles délimités par $n_d + 1$ noeuds, nommés noeuds internes. Des noeuds dits externes, placés de chaque côté complètent cet ensemble. Ainsi, il y a $n_d + 2d + 1$ noeuds au total.

Ainsi, au total, $n_d + d$ B-splines composent cette base. Et en tout point du support, $d + 1$ B-splines sont non-nuls.

5.1.7.3 Vraisemblance pénalisée

Pour cette méthode, la vraisemblance pénalisée est utilisée afin d'éviter le surajustement des données et ainsi permettre d'obtenir un lissage satisfaisant.

Par conséquent, la log-vraisemblance pénalisée peut s'écrire :

$$l_p(\theta) = l(\theta) - \frac{\pi}{2} \|D_k \theta\|^2$$

où :

- $l(\theta)$, correspond à la log-vraisemblance
- π , correspond à un paramètre d'échelle
- D_k , correspond à la matrice différence d'ordre k de dimension $(p-k) \times p$ pour $k \in \mathbb{N}$ et où p est la longueur du vecteur θ .

A noter, que cette méthode permet l'extrapolation du lissage, cependant, l'extrapolation n'est pas traitée dans ce mémoire car seuls les décès des salariés ont un intérêt pour l'étude.

5.1.7.4 Implémentation

L'implémentation de cette méthode est réalisée sous SAS en utilisant la procédure *TRANSREG* ainsi que l'option *PBSPLINE*.

Plusieurs paramètres peuvent être adaptés afin d'améliorer la qualité du lissage par rapport aux données disponibles.

Le degré choisi est de 3, ce qui correspond à la valeur par défaut. Le nombre de nœuds peut également être choisi, ce dernier est déterminé afin de conserver entre 4 et 5 observations par subdivisions pour permettre un lissage optimal. Ainsi, 14 nœuds sont retenus. L'espacement des nœuds est également choisi afin qu'il soit uniforme sur le support.

Enfin, le coefficient d'échelle pour la pénalité sera déterminé en minimisant un critère. Plusieurs critères sont disponibles sous SAS, celui retenu est le critère AIC. Le choix du coefficient minimisant ce critère est effectué automatiquement par SAS puis le résultat est directement calculé en utilisant ce coefficient.

5.2 Résultats

Dans cette partie, les résultats des modèles présentés ci-dessus vont être développés. Pour des questions de confidentialité les résultats présentés seront sous forme de graphique avec une échelle appropriée ou sans échelle afin d'illustrer les résultats obtenus sans révéler les résultats exacts.

5.2.1 Taux brut de mortalité

Cette première approche permet d'observer la forme des données. Dans un premier temps, la base comprenant tous les individus est employée, afin d'observer globalement les probabilités de décès. Puis, la méthode est répétée avec une segmentation par sexe afin d'observer les différents comportements dus à ce critère.

Les probabilités de décès dans l'année, q_x , sont représentées ici en échelle logarithmique afin de mieux observer les évolutions notamment aux âges jeunes.

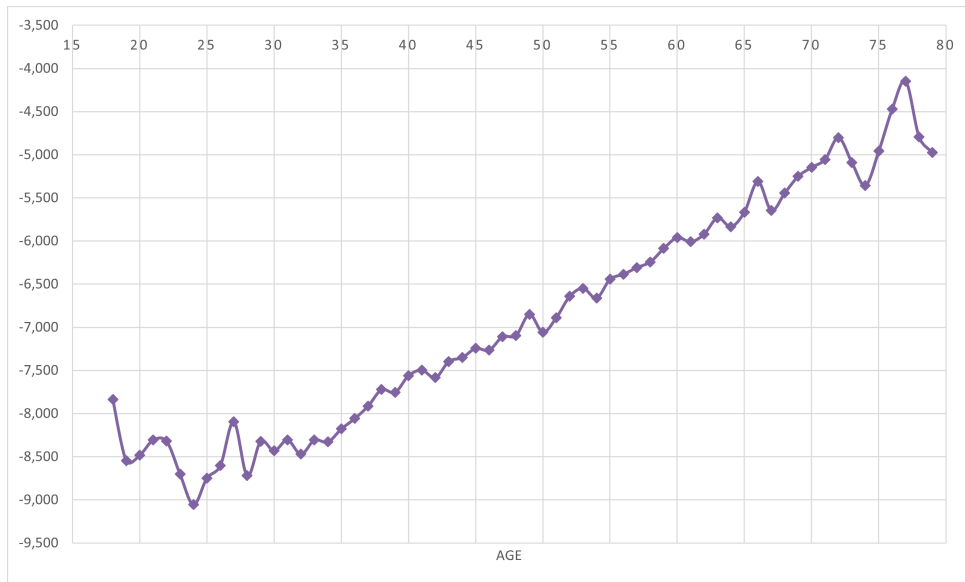


FIGURE 5.6 – Taux brut de mortalité, en échelle logarithmique, en fonction de l'âge

Comme attendu, la probabilité de décès q_x augmente avec l'âge, cette hypothèse étant faite dans le modèle utilisé. Remarquons que pour les âges extrêmes, c'est-à-dire avant 27 ans et après 72 ans, des fluctuations importantes apparaissent. Cela est dû à la faible exposition observée à ces âges et au très faible nombre de décès sur les âges jeunes.

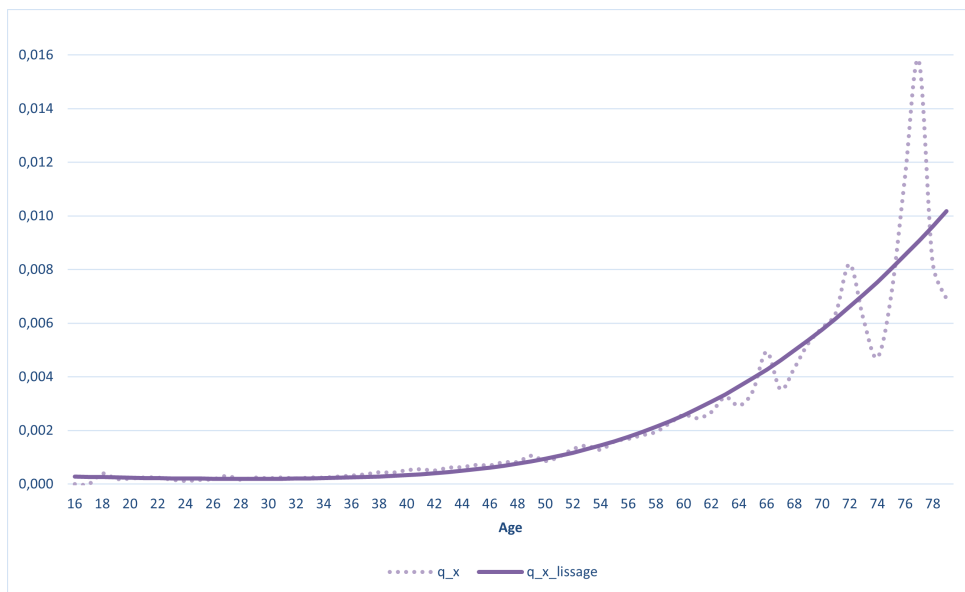


FIGURE 5.7 – Probabilités de décès brutes et lissées en fonction de l'âge

Un lissage a été réalisé avec la méthode des P-splines pénalisées, afin de lisser les fluctuations aux âges extrêmes. La procédure *TRANSREG* a été utilisée sous SAS pour réaliser ce lissage. La courbe lissée est de forme exponentielle, surtout après 35 ans, ce qui est normal car cette hypothèse a été prise dans le modèle.

Comme attendu la plupart des points sont très proches de la courbe même si certains points aux âges élevés sont assez éloignés. On voit bien qu'il s'agit de points outliers qui ne sont pas corrects mais qui sont dus à la faible exposition à ces âges et qui ne sont donc pas fiables.

Les taux bruts par année calendaire sont également calculés.

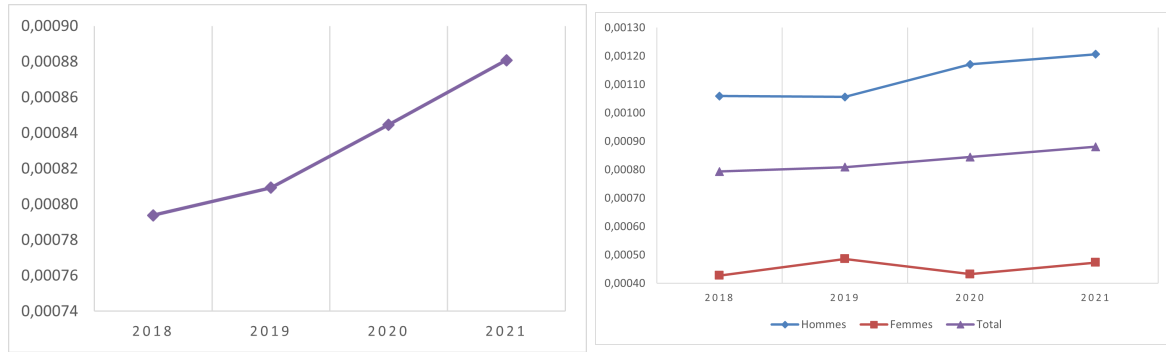


FIGURE 5.8 – Probabilités de décès par rapport à l'année calendaire

Les probabilités de décès augmentent légèrement en fonction de l'année calendaire. Une augmentation plus forte, qui peut être due au COVID-19 ou bien à l'incertitude du risque, est observée en 2020 et 2021. Mais comme on peut voir sur le graphique de droite, on observe la tendance inverse à celle des hommes sur la population féminine sur les 3 premières années, mais le portefeuille étant principalement constitué d'hommes, cela est donc lissé. On remarque également que les q_x des hommes sont toujours supérieurs à ceux des femmes. Cela est cohérent avec les autres lois de mortalité reconnues.

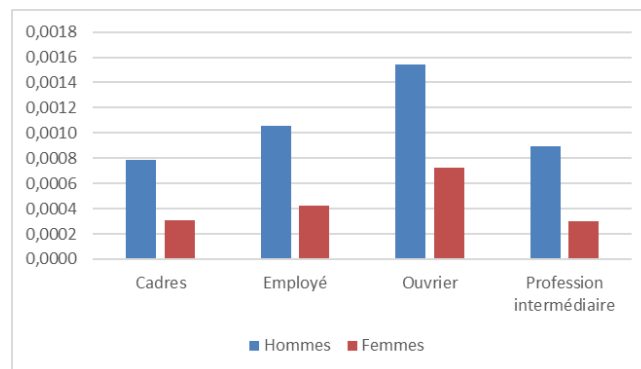


FIGURE 5.9 – Probabilités de décès brutes par rapport au sexe et à la CSP

Sur le graphique représentant les probabilités de décès par sexe et CSP, la différence entre les hommes et les femmes est une nouvelle fois observable. De plus, une vraie différence apparaît entre les différentes CSP, avec des probabilités de décès pour les ouvriers environ 2 fois supérieures à celles des cadres.

Les graphiques des probabilités de décès par âge en fonction du sexe avec lissage sont présentés ci-dessous. Notons que la courbe masculine est très différente de celle des femmes. En effet, les deux courbes se décrochent à partir de 40 ans avec la probabilité de décès des hommes qui croit de manière exponentielle alors que celle des femmes croit plutôt de manière linéaire jusqu'à 60 ans puis de manière exponentielle. La probabilité de décès des hommes avant 24 ans est supérieure à celle des hommes de 30 ans, cela s'explique par un phénomène présent dans les autres tables de référence et appelé la 'bosse des accidents'. Ainsi, cette décroissance locale de la mortalité est normale et attendue.

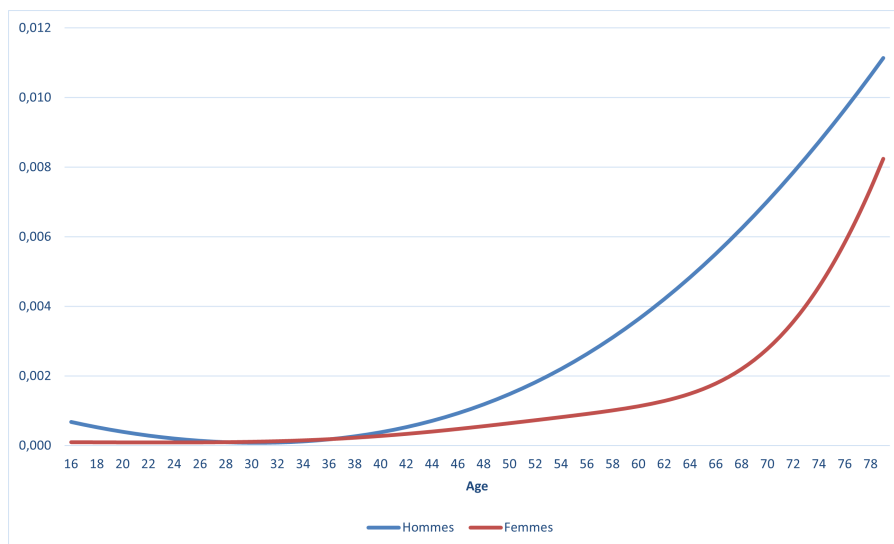


FIGURE 5.10 – Probabilités de décès lissées par rapport au sexe et à l'âge

5.2.2 Kaplan-Meier

Les fonctions de survie obtenues par la méthode de Kaplan-Meier sont divulguées ici. Sous SAS, la procédure *LIFETEST* a été utilisée pour obtenir les fonctions de survie avec la méthode de Kaplan-Meier.

Elles ont été réalisées par année calendaire et âge et segmentées par sexe.

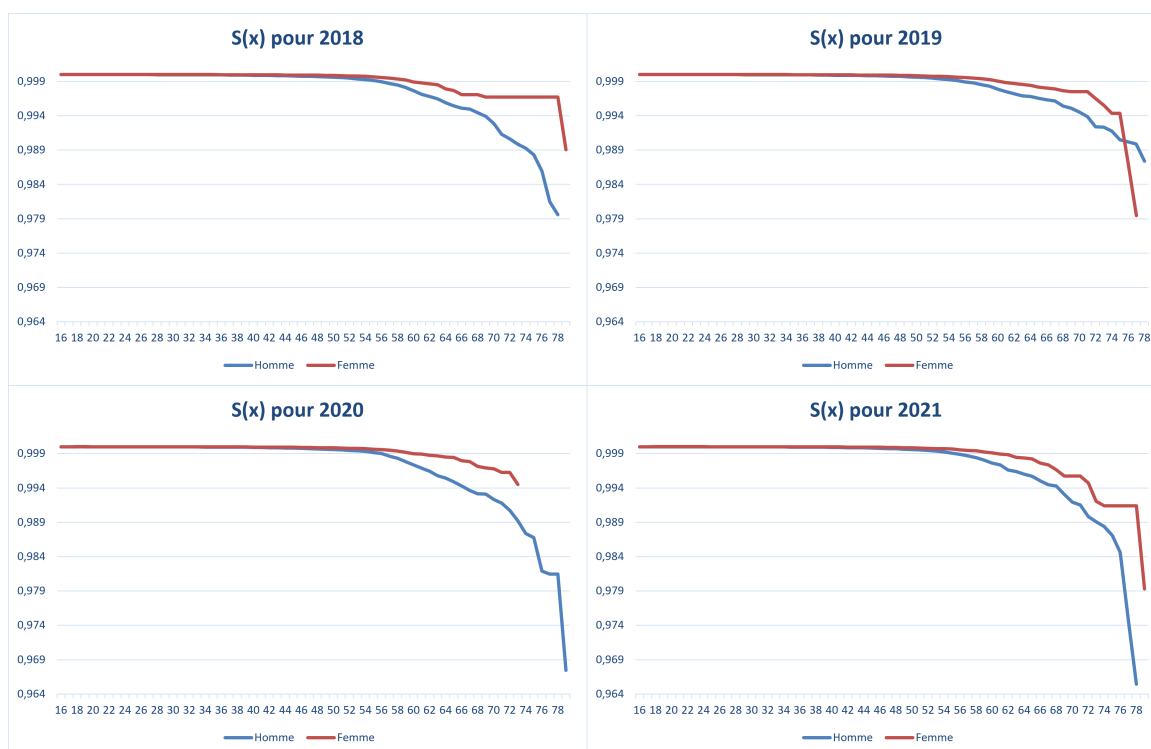


FIGURE 5.11 – Fonctions de survie par année calendaire et sexe

La courbe de survie des hommes est toujours en dessous de celle des femmes, sauf pour les âges élevés en 2019 mais cela est sûrement dû à des fluctuations dues à la faible exposition disponible à ces âges-là.

Les courbes des hommes et des femmes se détachent aux alentours de 55 ans. Les fluctuations importantes des courbes sur les âges élevés démontrent que cette méthode n'est pas robuste sur ces âges-là.

Ces probabilités de survies sont traduites en probabilités de décès et regroupés en faisant la moyenne des 4 q_x obtenus, correspondant aux années calendaires, par âge et sexe. Puis un lissage par P-splines est appliqué. Les résultats sont présentés ci-dessous.

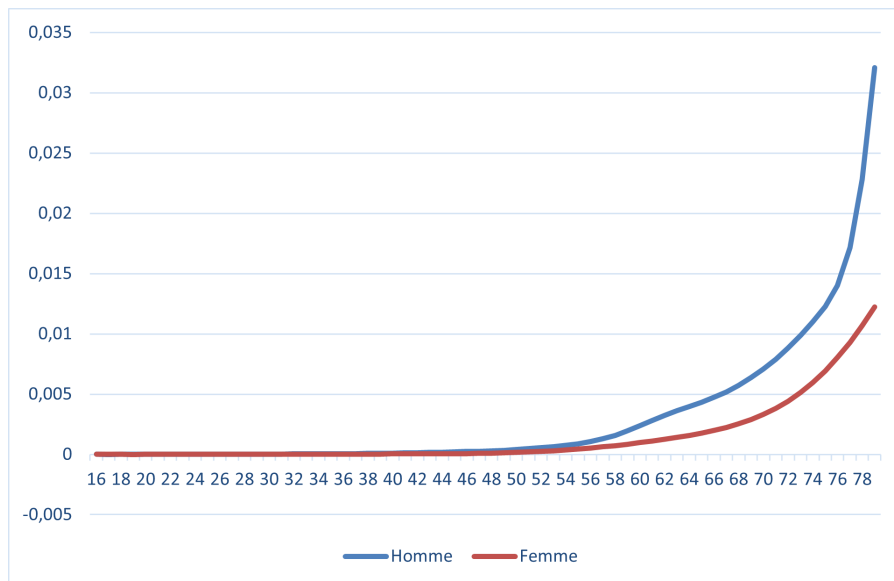


FIGURE 5.12 – Probabilités de décès lissées obtenues par Kaplan-Meier

Cette méthode n'utilise pas d'hypothèse a priori sur la forme des données. Le résultat obtenu confirme les hypothèses prises précédemment concernant le caractère exponentiel des lois de mortalité en fonction de l'âge. De plus, ici encore la courbe masculine est au-dessus de celle des femmes. Bien que le décrochement entre les 2 courbes ait lieu plus tard, aux alentours de 55 ans. La pente de la courbe féminine est moins accentuée que celle de la courbe masculine.

En conclusion, les résultats obtenus par la méthode de Kaplan-Meier confirment les observations constatées précédemment. Mais, cette méthode ne semble pas assez robuste pour supporter une segmentation plus fine. En effet, une segmentation par CSP, par exemple, a été testée et les résultats ne sont pas satisfaisants. Il s'agit donc là d'une limite de la méthode. De plus, le but étant de permettre la tarification des garanties décès, d'autres critères sont à prendre en compte. Ainsi, d'autres méthodes sont développées dans la suite.

5.2.3 GLM Poisson

Les méthodes précédentes n'étant pas satisfaisantes, un GLM Poisson est implémenté. Cette méthode est courante et va donc permettre de vérifier les effets des différents critères ainsi que la cohérence des résultats avec ceux obtenus précédemment.

Dans cette partie la démarche est présentée ainsi que les principaux modèles testés. Sous SAS, la procédure *GENMOD* a été utilisée afin de réaliser les GLM. Une distribution de Poisson et la fonction de lien canonique associée à cette loi, à savoir la fonction logarithme, ont été employées. De plus, le logarithme de l'exposition est pris comme offset. En effet, en supposant que la force de mortalité suit une loi de Poisson de paramètre $\frac{d}{e^c}$, alors en passant au logarithme un offset apparaît comme étant égal au logarithme de l'exposition.

A noter, que la loi de Poisson est utilisée car c'est la loi couramment utilisée dans le cadre des lois de mortalité. Cependant, ici, l'hypothèse fondamentale d'égalité entre l'espérance et la variance n'est pas respectée.

Dans un premier temps, l'évolution de la mortalité est supposée constante avec l'âge. En effet, un seul paramètre sera calculé pour la variable âge, il représente le coefficient de passage entre un âge x et un âge $x + 1$.

A noter que l'utilisation des âges comme variable catégorielle n'a pas été retenue du fait que certains âges ne possèdent pas assez d'exposition, par conséquent les modèles n'auraient pas été assez robustes.

Toutes les variables explicatives ont été utilisées dans un premier temps afin de voir les effets de ces différentes variables sur le modèle ainsi que leurs significativités.

Les niveaux de salaires utilisés ici sont ceux calculés en fonction de la CSP des individus. C'est pourquoi, le niveau de salaire est croisé avec la CSP dans le modèle.

La population de référence est composée des hommes employés, de niveau de salaire 3, travaillant dans le secteur du commerce, réparation d'automobiles et de motocycles et résidants en Ile-de-France.

Cependant, à la suite des premiers résultats, certaines variables ont été écartées. En effet, le secteur d'activité n'était pas significatif et a donc été rapidement écarté des modélisations. Cependant, ce critère est pris en compte dans l'outil actuel de tarification mais sous forme de classes tarifaires. C'est pourquoi, ce critère est rajouté à l'étude afin de voir si ces classes tarifaires ont un impact. Malgré tout, ce critère est lui aussi non significatif. Ainsi, il est également écarté de l'étude.

Plusieurs raisons peuvent être avancées. Premièrement, ces variables sont corrélées avec le sexe, la CSP et le niveau de salaire. Ainsi l'effet de ces variables est déjà pris en compte via ces autres variables. En effet, par exemple, la variable CSP est très significative et très corrélée avec le secteur d'activité. L'effet du secteur d'activité ou des classes tarifaires est non significatif par rapport à celui de la CSP en réalité.

Par ailleurs, les résultats obtenus concernant les niveaux de salaire par CSP étaient absurdes d'un point de vue métier. En effet, les résultats obtenus étaient en contradiction avec les différentes études publiées sur le sujet. L'analyse des résultats a été effectuée afin de déterminer la cause des résultats (problème de méthode, mauvaise qualité des données ...). Cette analyse a été réalisée sur les résultats obtenus par la méthode du SMR par critère. En effet, cette méthode étant réalisée pas à pas et codée à la main, contrairement au GLM où une procédure SAS est utilisée, il était plus facile de tester différentes versions et de décortiquer les différentes étapes de calcul. Néanmoins, comme toutes les méthodes testées aboutissaient aux mêmes résultats il était très vraisemblable que le problème provienne des données. Les conclusions de cette analyse seront donc détaillées dans la section 5.2.5. Mais la conclusion principale est que le problème provenait bien de la qualité de la variable salaire, c'est pourquoi cette variable est écartée dans la suite.

Les variables significatives sont finalement le sexe, l'âge, la CSP et la région. Toutes les modalités des 3 premiers critères sont significatives. Pour la région des regroupements ont du être effectués. Comme expliqué auparavant, les régions non significatives sont regroupées avec la modalité significative ayant le coefficient le plus proche. Les regroupements suivants ont été effectués :

- Région 1 : Ile-de-France ; Autre ; Bourgogne ; Bretagne ; Centre-Val de Loire ; Cote d'Azur ; Nouvelle Aquitaine ; Pays de la Loire
- Région 2 : Occitanie ; Auvergne
- Région 3 : Normandie ; Grand Est ; Hauts-de-France

Finalement, toutes les modalités sont bien significatives au seuil 95%.

L'AIC du modèle est de 4116. Il est bien inférieur à celui obtenu avant regroupement des modalités. La déviance est de 1404 et la p-value du test confirme que la déviance suit une loi du Khi-deux à 1511 degrés de liberté. De plus, le test permettant d'estimer le coefficient de dispersion avec la déviance conduit à une estimation de 0,93 ce qui est proche de 1. Et le test en utilisant le Khi-deux de Pearson est quant à lui, égal à 1,13 ce qui est un peu plus éloigné de 1.

Les intervalles de confiance sont présentés sur quelques exemples :

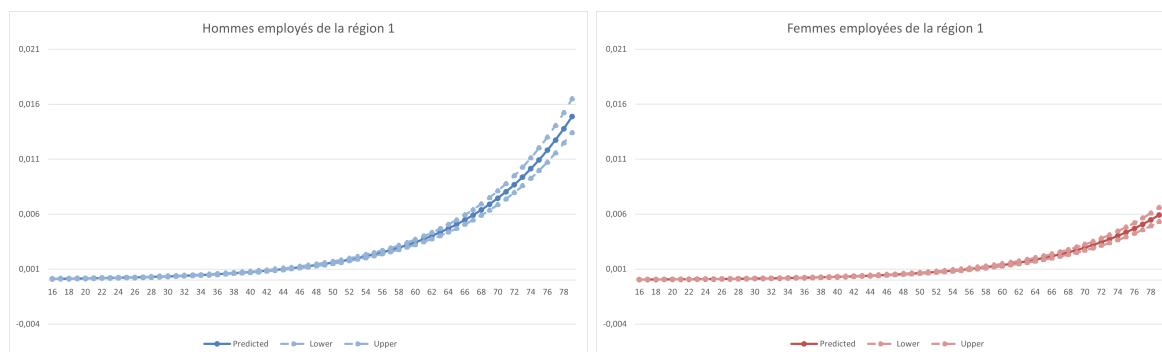


FIGURE 5.13 – Probabilités de décès direct, avec intervalle de confiance, en fonction du sexe, de l'âge, de la CSP et de la région

La population de référence est représentée à gauche. Comme attendu, la probabilité de décès croît avec l'âge. Les hommes ont également, toutes choses égales par ailleurs, une mortalité supérieure à celle des femmes. Les intervalles de confiance sont en cohérence avec les données. En effet, l'exposition disponible pour les âges extrêmes est faible par rapport au reste de la population. Cela est bien reflété par le fait que l'intervalle de confiance est plus large pour ces âges, l'incertitude étant plus grande due au peu de données disponibles.

Les résidus du modèle sont analysés. Les résidus de Pearson standardisés et ceux de la déviance sont étudiés afin de vérifier l'adéquation du modèle :



FIGURE 5.14 – Résidus de Pearson et de la déviance standardisés

Un nombre assez important de résidus de Pearson ne sont pas dans l'intervalle $[-2; 2]$ ce qui suggère que le modèle explique mal les données. En ce qui concerne les résidus de la déviance, les résidus ne semblent pas alterner de signes assez régulièrement. Ainsi, le modèle ne semble pas optimal.

Toutefois, afin de pouvoir comparer réellement avec les lois de référence, la proportion des décès faisant suite à arrêt de travail est rajoutée aux résultats de manière arbitraire.

La région 3 est présentée dans les graphiques suivants, afin de voir le comportement de la population la plus à risque.

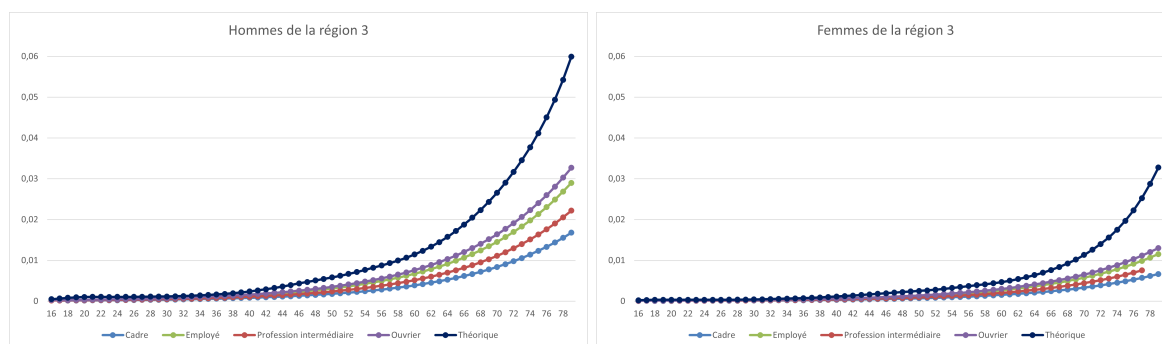


FIGURE 5.15 – Probabilités de décès par CSP en fonction du sexe et de l'âge

La surmortalité masculine est bien représentée, néanmoins, malgré l'ajout arbitraire des décès suite à arrêt de travail, la mortalité obtenue des salariés du portefeuille est bien inférieure à celle de référence. Cela s'explique par plusieurs phénomènes. Premièrement, les lois de référence ont été calculées à partir de la population française de 2000 à 2002. Or, depuis presque 20 ans avec les progrès de la médecine et l'amélioration des conditions de vie, la mortalité a évolué. De plus, comme dit précédemment, ces lois de référence sont calculées sur la population française et pas seulement sur les populations salariées. Or, les inactifs, non présents dans le portefeuille de salariés, ont une surmortalité très importante par rapport à la population générale². Ainsi, cela explique que les résultats obtenus sont bien inférieurs aux lois de référence.

Les ouvriers ont la mortalité la plus élevée tandis que les cadres ont celle la plus faible. Cela semble en cohérence avec les différentes études publiées sur le sujet. La mortalité des employés est plus proche de celle des ouvriers que des cadres, et le phénomène inverse est observé pour les professions intermédiaires.

On remarque également que les hommes employés et ouvriers, de la population étudiée, semblent avoir une mortalité proche de celle de la table de référence des femmes.

L'effet de la région est également présent. Seuls les résultats concernant la région 3 sont présentés ici mais les observations suivantes peuvent être faites sur ce critère. Le regroupement correspondant à la région 1 possède une mortalité intermédiaire tandis que la région 2 possède une mortalité inférieure à l'ensemble de la population et au contraire la région 3 possède une mortalité supérieure à l'ensemble de la population.

Toutefois, le fait d'avoir un seul coefficient ne semble pas idéal. En effet, l'effet exponentiel de la mortalité en fonction de l'âge risque de ne pas bien être pris en compte de cette manière. Par ailleurs, l'analyse des résidus a permis de conclure que les données n'étaient pas bien ajustées par ce modèle.

C'est pourquoi, dans un second temps, les tranches d'âge sont utilisées. De cette manière, l'évolution entre les différentes tranches d'âge pourra ne pas être constante. De plus, l'exposition sera assez conséquente dans chaque tranche d'âge.

La population de référence est alors composée des hommes entre 40 et 50 ans, employés, de niveau de salaire 3, travaillant dans le secteur du commerce, réparation d'automobiles et de motocycles et résidants en Ile-de-France.

La même méthodologie que précédemment est reproduite ici. Les mêmes commentaires sont encore valides ici en ce qui concerne le salaire et le secteur d'activité.

2. <http://onala.free.fr/inegamor.pdf>

Les variables retenues sont finalement le sexe, la tranche d'âge, la catégorie socio-professionnelle et la région de résidence.

Cependant, en considérant les tests de Wald, toutes les modalités ne sont pas significatives. En effet, la catégorie *profession intermédiaire* n'est pas significative au seuil 95%, elle est donc regroupée avec la catégorie la plus proche qui est en l'occurrence la catégorie *employé*. De même, pour la région de résidence seules 3 régions sont significatives ainsi les mêmes regroupements que pour le modèle précédent sont effectués.

L'AIC est de 549, et la déviance de 81,9. La p-value du test est bien supérieure à 0,05 ce qui indique que la déviance suit bien une loi du Khi-deux à 80 degrés de liberté. En effectuant le test pour estimer le paramètre de dispersion avec la déviance, le résultat est de 1,02 ce qui est très proche de 1 et indique donc que le modèle explique bien les données. Le même test effectué avec le Khi-deux de Pearson résulte à une estimation de 0,98 ce qui est également très proche de 1.

En ce qui concerne les lois de référence, ces dernières étant exprimées en fonction de l'âge, la moyenne des q_x est prise comme référence pour chaque tranche d'âge.

Dans un premier temps, quelques courbes avec les intervalles de confiance associés sont présentées ici.

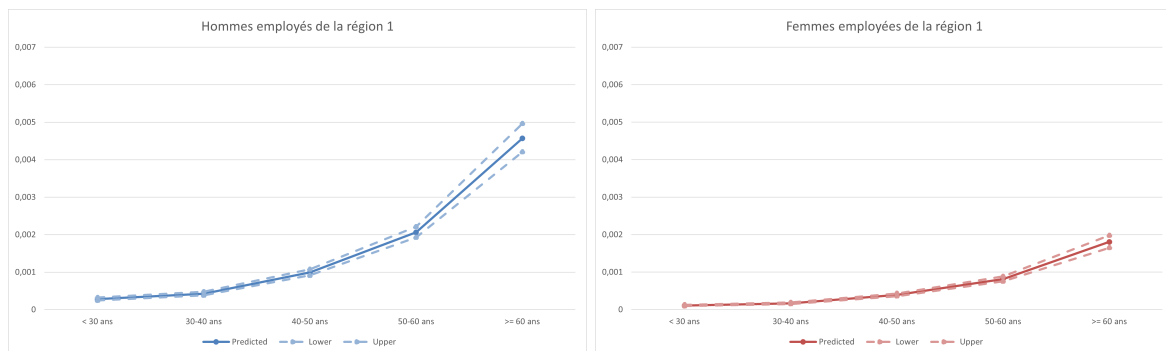


FIGURE 5.16 – Probabilités de décès direct, avec intervalle de confiance, en fonction du sexe, de la tranche d'âge, de la CSP et de la région

La population de référence est représentée à gauche. Comme attendu, la probabilité de décès croît avec l'âge. Les hommes ont également, toutes choses égales par ailleurs, une mortalité supérieure à celle des femmes. Les intervalles de confiance sont en cohérence avec les données. En effet, l'exposition disponible pour la tranche des plus de 60 ans est bien inférieure à celle des autres tranches. Cela est bien reflété par le fait que l'intervalle de confiance est plus large pour cette tranche d'âge, l'incertitude étant plus grande due au peu de données disponibles.

Les résidus du modèle sont analysés. Les résidus de Pearson standardisés et ceux de la déviance sont étudiés afin de vérifier l'adéquation du modèle :

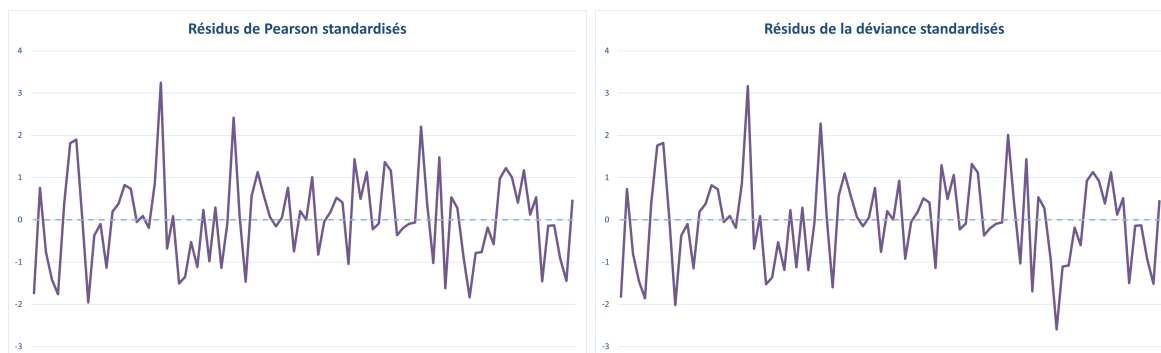


FIGURE 5.17 – Résidus de Pearson et de la déviance standardisés

Seuls 3 résidus de Pearson ne sont pas dans l'intervalle $[-2; 2]$ ce qui suggère que le modèle explique bien les données. En ce qui concerne les résidus de la déviance, les résidus alternent bien de signes sauf vers la fin où plusieurs résidus successifs sont du même signe. Ainsi, le modèle semble correct mais pas optimal.

Toutefois, afin de pouvoir comparer les résultats avec les lois de référence, la proportion des décès faisant suite à arrêt de travail est rajoutée de manière arbitraire.

Les 3 regroupements de régions sont présentés dans les graphiques suivants, afin de voir l'effet de cette variable.

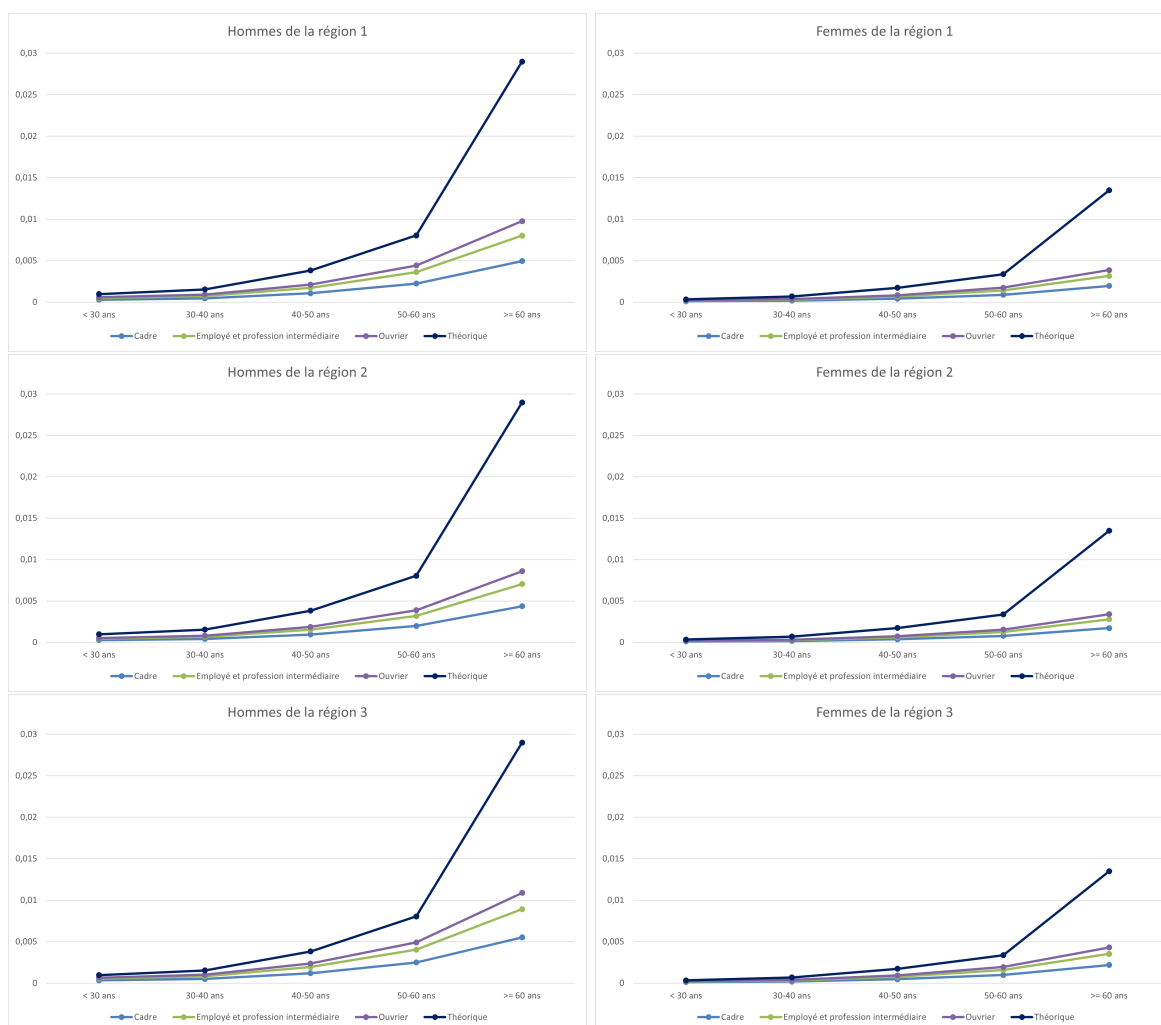


FIGURE 5.18 – Probabilités de décès par CSP en fonction du sexe, de la tranche d'âge et de la région

La surmortalité masculine est bien représentée, néanmoins, malgré l'ajout arbitraire des décès suite à arrêt de travail, la mortalité obtenue des salariés du portefeuille est, une nouvelle fois, bien inférieure à celle de référence.

L'ordre des courbes selon la CSP et le sexe est le même que précédemment.

Un effet de l'âge semble exister car les coefficients ne sont pas proportionnels entre les tranches d'âge. La classe des plus de 60 ans ayant un coefficient très élevé, qui confirme que la mortalité est liée à l'âge et qu'elle s'accélère avec l'âge et notamment après 60 ans.

L'effet de la région est également visible. Le regroupement correspondant à la région 1 possède une mortalité intermédiaire tandis que la région 2 possède une mortalité inférieure à l'ensemble de la population et au contraire la région 3 possède une mortalité supérieure à l'ensemble de la population.

A noter que la plupart des modèles ont été appliqués aux populations masculines et féminines de manière séparée afin de vérifier les différents effets des variables explicatives par rapport au sexe. Les niveaux de significativité des variables n'étaient cependant pas modifiés. Mais certains coefficients présentent des écarts importants entre les 2 populations, à modèle équivalent.

Pour conclure ce modèle semble mieux adapté aux données.

Cependant, le fait de ne pas avoir une loi par âge peut être déroutant et manquer de précision. De plus, l'approximation prise sur les lois de référence peut être remise en question.

Ainsi, des pistes d'amélioration sont envisageables, comme l'augmentation du nombre de tranches d'âges ou bien des catégories socio-professionnelles plus fines comme la séparation des ouvriers qualifiés ou non par exemple.

Ces premiers modèles permettent de mettre en évidence les variables significatives. Ce sont donc ces variables qui seront utilisées dans le modèle suivant. Rappelons que le but est de tarifier le décès direct ainsi, une loi de mortalité n'est pas nécessaire, seuls des coefficients d'abattement par rapport aux lois de référence sont recherchés. C'est pourquoi ces méthodes constituent une première approche afin de mieux appréhender les données et de mieux choisir les critères de segmentation et leur ordre dans les parties suivantes. De plus, une comparaison des différents résultats obtenus permettra de valider et fiabiliser les conclusions de l'étude. La méthode suivante permet d'obtenir des coefficients d'abattements en modifiant légèrement le modèle du GLM Poisson.

5.2.4 GLM Poisson avec référence

Dans cette section, les logarithmes des lois de référence sont ajoutés en offset afin d'obtenir des coefficients de passage entre les forces de mortalité prédites et celles de référence. Néanmoins, le modèle est ici encore un GLM Poisson. Les lois de référence utilisées ici sont encore les lois TH-00-02 et TF-00-02.

Dans un premier temps, la variable âge est utilisée. Cette variable est considérée comme numérique ainsi ici encore un seul coefficient de passage entre chaque âge est calculé.

La population de référence est la même que celle du précédent modèle utilisant la variable âge. Cependant, comme pour les modèles précédents, le secteur d'activité n'est pas significatif et les classes tarifaires de l'outil de tarification ne le sont pas non plus. On rappelle que les niveaux de salaire ne sont plus utilisés car la qualité de données est très mauvaise.

Les variables conservées sont donc le sexe, l'âge, la CSP, la région ainsi que le croisement entre l'âge et la CSP.

Un résultat suprenant aux premiers abords, est obtenu. En effet, les coefficients concernant le sexe et l'âge ne sont pas significatifs. Cependant, comme ces coefficients s'interprètent par rapport aux lois de référence, cela indique que les effets du sexe et de l'âge sur les données sont très similaires à ceux des lois de référence. Ainsi, cela confirme que nos données sont en accord avec les lois certifiées et renforce donc la raison de leur utilisation.

Ainsi, le modèle est de nouveau implémenté en supprimant le sexe et l'âge aux vues de leur non significativité. Les autres variables sont bien significatives au seuil de 95%, mais pas toutes les modalités de la variable région. Ainsi, les mêmes regroupements que précédemment sont effectués. Toutes les modalités sont alors bien significatives.

L'AIC du modèle est de 4090. Il est bien inférieur à celui obtenu avant regroupement des modalités. La déviance est de 1376 et la p-value du test confirme que la déviance suit une loi du Khi-deux à 1510 degrés de liberté. De plus, le test permettant d'estimer le coefficient de dispersion avec la déviance conduit à une estimation de 0,91 ce qui est proche de 1. Et le test en utilisant le Khi-deux de Pearson est quant à lui, égal à 0,99 ce qui est encore plus proche de 1.

Les intervalles de confiance sont présentés sur quelques exemples :

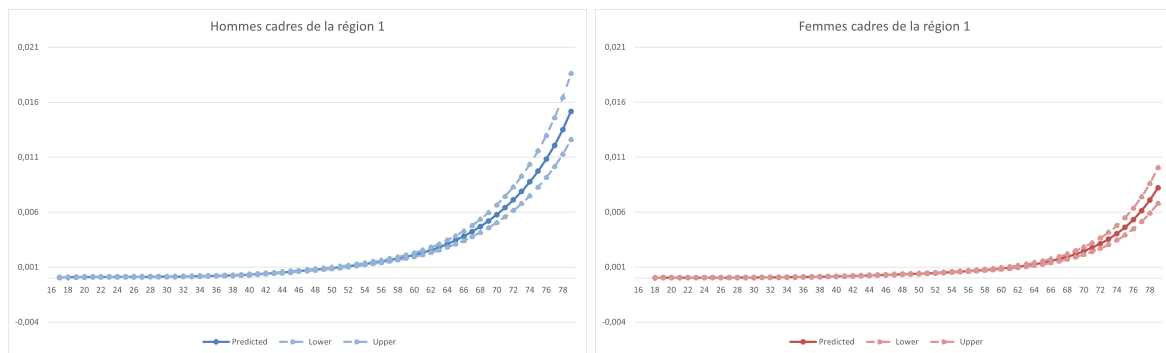


FIGURE 5.19 – Probabilités de décès direct, avec intervalle de confiance, en fonction du sexe, de l'âge, de la CSP et de la région

La population de référence est représentée à gauche. Les intervalles de confiance sont ici encore croissants avec l'âge et l'exposition observée.

Les résidus du modèle sont analysés. Les résidus de Pearson standardisés et ceux de la déviance sont étudiés afin de vérifier l'adéquation du modèle.

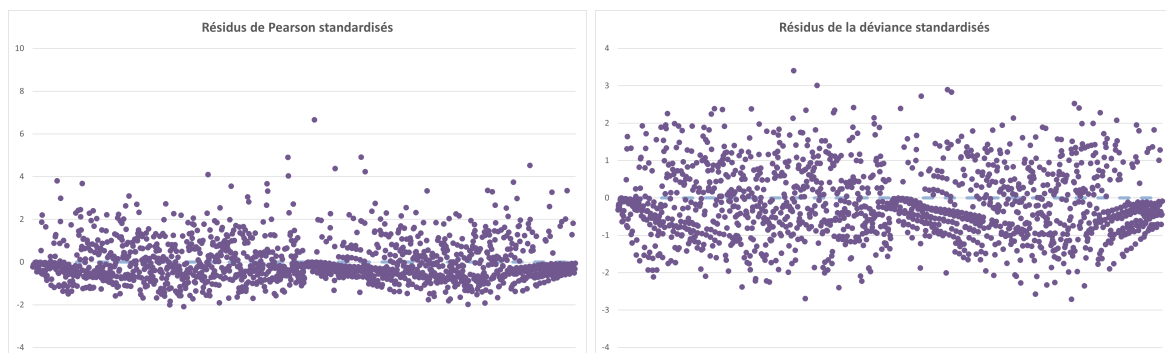


FIGURE 5.20 – Résidus de Pearson et de la déviance standardisés

Un nombre assez important de résidus de Pearson ne sont pas dans l'intervalle $[-2; 2]$ ce qui suggère que le modèle explique mal les données. Cependant, par rapport au grand nombre d'observations, la proportion de points en dehors de cet intervalle reste correcte.

En ce qui concerne les résidus de la déviance, les résidus ne semblent pas alterner de signes assez régulièrement. Cela semble être le cas sur les observations correspondants aux âges élevés. Cela peut être du aux données et à l'exposition plus faible sur ces âges.

Ainsi, le modèle ne semble pas optimal mais les catégories sur lesquels le modèle est le moins adapté correspondent aux observations où l'incertitude est élevée. Ainsi, le fait de se baser sur les lois de référence permet de consolider cette partie même si cela ne convient pas aux données observées.

Afin de pouvoir comparer avec les lois de référence, la proportion des décès faisant suite à arrêt de travail est rajoutée aux résultats de manière arbitraire.

La région 1 et 3 sont présentées dans les graphiques suivants, afin de voir le comportement de la population par rapport à ce critère.

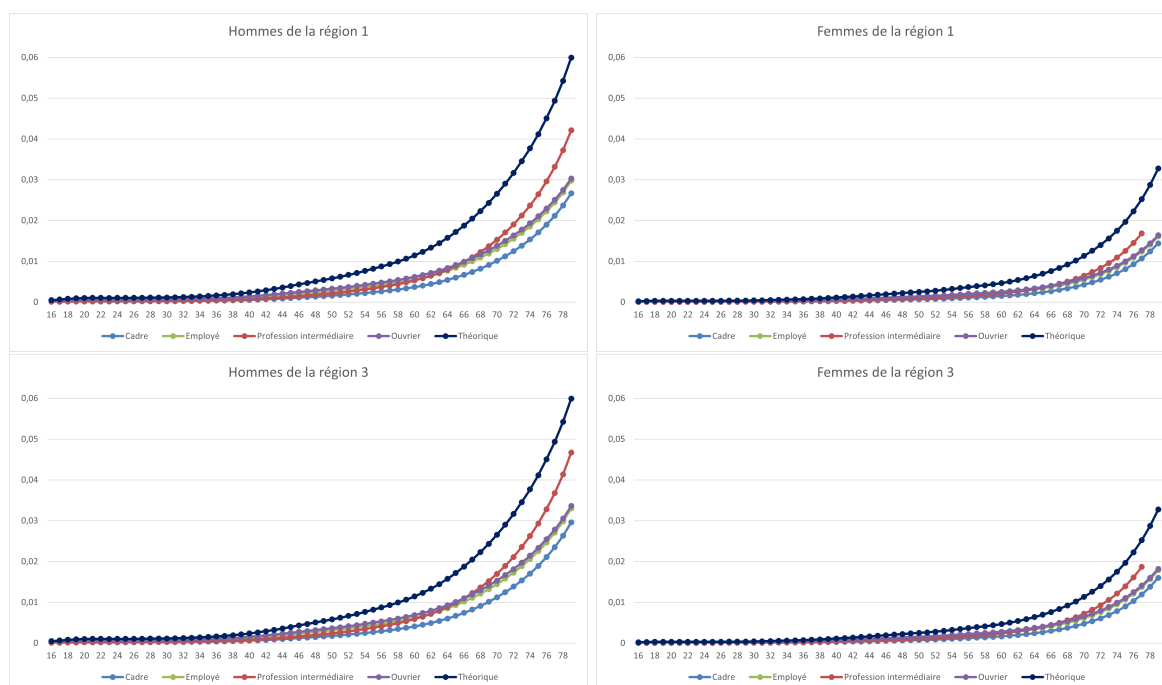


FIGURE 5.21 – Probabilités de décès par CSP en fonction du sexe, de l'âge et de la région

Les ouvriers ont la mortalité la plus élevée aux âges jeunes mais aux âges élevés ce sont les professions intermédiaires qui ont la mortalité la plus élevée. Cependant il faut rester prudent avec cette observation car l'incertitude est très élevée car l'exposition des professions intermédiaires est inférieure à celle des autres CSP. La mortalité des ouvriers et des employés est très similaire aux âges élevés.

L'effet de la région est également présent. La différence entre la région 1 et la région 3 est visible sur ces graphiques. La région 2, quant à elle, possède une mortalité inférieure à l'ensemble de la population.

Ce modèle semble donc très satisfaisant car les lois de référence permettent d'assurer que les potentiels problèmes d'exposition sont corrigés.

Cependant afin de voir si sans la région les résidus sont meilleurs, le même modèle sans ce critère est implémenté.

Toutes les modalités sont significatives. L'AIC est de 1999. Il est normal qu'il soit inférieur à l'autre modèle

car il y a un critère en moins. La déviance est de 549 et la p-value du test confirme que la déviance suit une loi du Khi-deux avec 505 degrés de liberté. De plus, le test permettant d'estimer le coefficient de dispersion vaut 1,09 avec la déviance et avec le Khi-deux de Pearson.

Les résidus obtenus sont affichés afin de voir si ils sont mieux que ceux du modèle précédent.

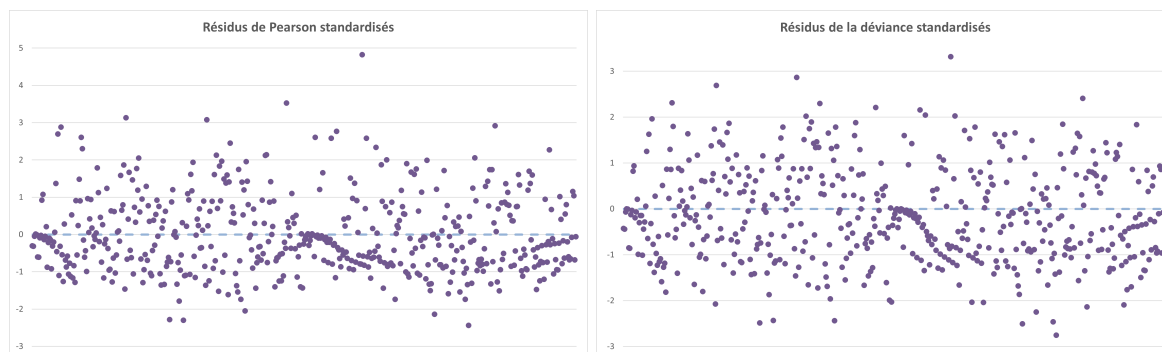


FIGURE 5.22 – Résidus de Pearson et de la déviance standardisés

De même que précédemment, une petite partie des résidus de Pearson est en dehors de l'intervalle $[-2; 2]$. Les résidus de la déviance semblent alterner assez souvent.

Ce modèle ayant des résultats similaire au modèle précédent mais étant plus simple il est donc retenu.

La CSP et le croisement entre la CSP et l'âge sont bien significatifs. Toutes les modalités le sont également. Les résultats indiquent que les cadres et les professions intermédiaires ont une mortalité inférieure à celle des employés. Les ouvriers quand à eux meurent plus que les employés.

Le fait que le croisement entre l'âge et la CSP soit significatif signifie que la mortalité avec l'âge n'évolue pas de la même manière en fonction de la CSP. Ainsi, malgré le fait que la mortalité des ouvriers soit la plus élevée elle a tendance à diminuer légèrement avec l'âge par rapport à l'évolution des lois de référence. Ainsi les 2 courbes s'éloignent de plus en plus avec l'âge. Tandis que le phénomène inverse à lieu chez les cadres et les professions intermédiaires.

En ce qui concerne les coefficients pour les décès directs, une analyse des coefficients correspondants au rapport entre la force de mortalité estimée et celle de référence est effectuée. Ces derniers correspondent également à l'exponentielle de la matrice des caractéristiques avec celles des coefficients obtenus, i.e. $X\theta$, en utilisant les mêmes notations que dans la section 5.1.5.3. Les impacts de ces ratios étant équivalents, et quasiment égaux, à ceux correspondants au rapport entre la probabilité conditionnelle de décès à un an estimée et celle de référence.

Ici, les ratios dépendent seulement de l'âge et de la CSP. Néanmoins, le rapport entre les forces de mortalité pour les employés ne dépend pas de l'âge car il s'agit de la population de référence. Ce dernier est d'environ 0,28 ce qui indique que la force de mortalité de référence est presque 3,6 fois plus importante que celle observée. En rajoutant la proportion des décès suite à arrêt de travail cela correspond à un coefficient d'environ 0,49 entre la force de mortalité estimée et celle de référence.

Pour les cadres, les ratios croissent avec l'âge et varient d'environ 0,09 à 0,25, pour les décès directs. Donc, avec l'âge la force de mortalité observé se rapproche de plus en plus de celle de référence mais aussi de celle des employés. Ainsi, dans la population des cadres du portefeuille la sous-mortalité aux jeunes âges est accentuée par rapport à celle aux âges élevés.

Pour les professions intermédiaires, le constat est similaire à celui des cadres avec des ratios variants de 0,1 à 0,4. Par conséquent, l'accroissement des ratios avec l'âge est plus rapide que chez les cadres. Ainsi, la mortalité des professions intermédiaire finit par dépasser celle des employés. Cependant, la fiabilité des résultats sur les âges extrêmes est très faible pour cette population donc les résultats sont à prendre avec prudence.

Pour les ouvriers, les ratios varient de 0,38 à 0,28. Par conséquent, le phénomène inverse a donc lieu pour les ouvriers. L'écart entre la force de mortalité observée et celle de référence croît avec l'âge. En outre, avec l'âge la force de mortalité observé se rapproche de plus en plus de celle des employés.

On rajoute arbitrairement la proportion des décès suite à arrêt de travail pour pouvoir comparer les courbes obtenues à celles de référence. Une transformation logarithmique est appliquée afin de mieux observer les différentes courbes.

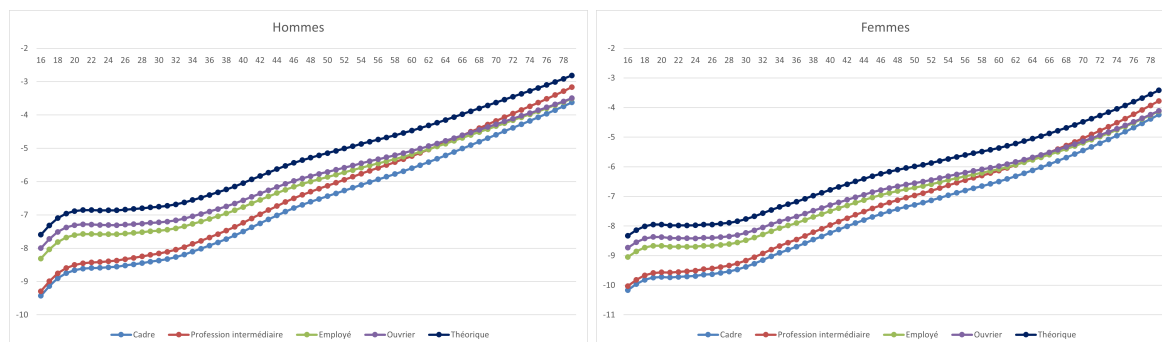


FIGURE 5.23 – Probabilités de décès en échelle logarithmique, en fonction du sexe, de l'âge et de la CSP

Les effets décrits précédemment sont bien visibles sur ces courbes.

Toutefois, afin de vérifier qu'il n'y a pas d'effet lié au sexe, la modélisation est séparée par sexe. Ainsi, les modèles sont implémentés sur la population masculine puis sur la population féminine.

Les hommes sont étudiés en premier. L'âge, la CSP, le croisement entre la CSP et l'âge et la région sont analysés ici. L'âge est encore une fois non significatif, ainsi que la région. Finalement, l'effet des coefficients masculins est similaire à l'effet global, hommes et femmes inclus.

En ce qui concerne les femmes seule la CSP est significative. En effet, ici au seuil 95% le croisement entre la CSP et l'âge n'est pas significatif. Il ne l'est pas non plus au seuil 90% mais presque. Ainsi, le fait qu'il y ait plus d'hommes dans les données masque cet effet.

Les coefficients du rapport des forces de mortalité dépendent seulement de la CSP ici. Celui des employés, seulement pour les décès directs, est d'environ 0,26 alors qu'il est approximativement de 0,30 pour les ouvriers. Celui des cadres est d'environ 0,17 alors qu'il est approximativement de 0,19 pour les professions intermédiaires. Par conséquent, la hiérarchie entre les catégories socio-professionnelles reste intact par rapport à celle observée aux âges jeunes précédemment. Mais pour les femmes, l'effet de l'âge semble uniforme sur les différentes catégories socio-professionnelles. Les employés et les ouvriers ont une mortalité similaire tandis que les cadres et les professions intermédiaires ont également une mortalité similaire.

La proportion des décès suite à arrêt de travail est rajouté afin de produire le graphique suivant :

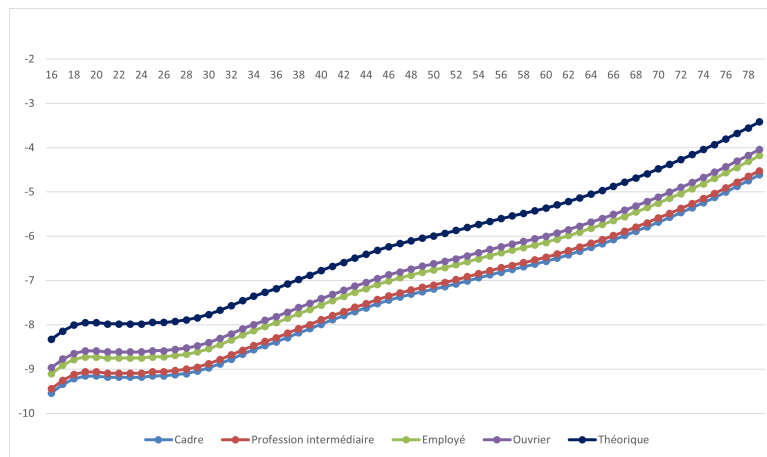


FIGURE 5.24 – Probabilités de décès des femmes, en échelle logarithmique, en fonction de l'âge et de la CSP

Malgré tout, l'effet de l'âge étant non significatif, les tranches d'âge sont utilisées afin de voir si un effet de génération apparait. La tranche des 40-50 ans est utilisée comme référence. Les autres variables analysées précédemment sont encore étudiées ici.

La moyenne des lois de référence selon les tranches d'âge est utilisée ici comme offset.

Cette fois les variables significatives sont le sexe, les tranches d'âge, la CSP, le croisement entre la CSP et les tranches d'âge ainsi que la région. Cependant, seule la modalité des plus de 60 ans est significative au seuil 95% pour la variable tranche d'âge par rapport à la modalité de référence. Ainsi cela nécessiterait de regrouper tous les individus de moins de 60 ans dans une même tranche d'âge. Cela n'est pas intéressant pour cette étude car ce sont justement les moins de 60 ans qui en sont l'objet principal.

Néanmoins, les regroupements ont été effectués pour la tranche d'âge, la CSP et la région afin de tester le modèle. Pour ces 2 dernières variables, il s'agit des mêmes regroupements que précédemment.

La variable sexe devient alors non significative et est donc retirée du modèle.

Finalement, les tests concernant la qualité de l'ajustement du modèle sont effectués. Les tests d'approximation du paramètre de dispersion via la déviance et via le Khi-deux de Pearson sont éloignés de 1 (entre 1,33 et 1,35), de ce fait, le modèle explique mal les données. Cependant, la p-value du test d'adéquation à la loi du Khi-deux, confirme que la déviance suit une loi du Khi-deux à 30 degrés de liberté.

L'analyse des résidus démontre que certains résidus de Pearson ne sont pas dans l'intervalle $[-2; 2]$ et pour ceux de la déviance, l'alternance de signe n'a pas lieu à plusieurs reprises.

Pour toutes ces raisons ce modèle n'est pas retenu car il ne semble pas bien adapté aux données, ni à l'utilité finale.

Cette méthode a donc permis de mettre en lumière différents effets et notamment de comparer les effets par rapport aux lois de référence. Les résultats du second modèle, discrétisé par âge, sont très intéressants et indiquent que les lois de référence sont bien paramétrées en fonction du sexe et de l'âge mais sont trop pessimistes par rapport à notre portefeuille. De plus, des effets comme la CSP et la région ne sont pas pris en compte dans ces lois alors qu'ils sont significatifs sur la mortalité.

Cependant, ce type de modèle impose des hypothèses assez fortes comme le fait de supposer que les données suivent une loi de Poisson. De plus, toutes les variables sont traitées ensemble dans le GLM, il est ainsi assez

difficile d'observer les effets des différentes variables. C'est pourquoi, la méthode suivante, le SMR par critère, est utilisée car elle permet de corriger les données pas à pas avec peu d'hypothèses pré requises.

5.2.5 SMR par critère

Dans cette partie, les critères sont étudiés un par un, tout en corrigeant les lois de référence à chaque étape afin de neutraliser les effets du critère analysé.

Par soucis de confidentialité les coefficients ne sont pas explicités. Seules les tendances sont illustrées par des courbes.

Le premier critère pris en compte sera le sexe car les tables de référence utilisées sont divisées selon ce critère. En effet, il s'agit d'un des critères les plus discriminants pour la mortalité.

Ainsi, deux SMR sont calculés, un pour les hommes et un pour les femmes.

Les deux coefficients obtenus sont presque égaux, environ 0,25, ce qui suggère que les lois de référence sont bien paramétrées par rapport au sexe. Les coefficients paraissent toutefois très faibles. Une première raison est que seuls les décès directs sont étudiés ici. Cependant même en rajoutant les 43% de décès faisant suite à un arrêt de travail, les coefficients obtenus restent en dessous de 0,5. Cependant, le fait que les coefficients trouvés soient inférieurs à 1 semble cohérent car cela indique que la mortalité des salariés est inférieure à celle de la population générale. Or cela est attendu car les inactifs ont une mortalité supérieure à celle de la population générale³. Un autre argument est que les tables de référence TH 00-02 et TF 00-02 sont basées sur la population française de 2000 à 2002. Or, la mortalité a évolué en 2 décennies à l'aide notamment des progrès en médecine. Ainsi, les lois de référence utilisées ne sont plus nécessairement adaptées à la population actuelle.

C'est pourquoi les coefficients ont été recalculés en changeant les lois de référence. L'Ined publie des lois de mortalité chaque année mais elles ne sont pas certifiées. L'Ined a récemment publié une table de mortalité basée sur la population française, Mayotte inclus, de 2018 à 2020⁴.

Les coefficients obtenus sont supérieurs à ceux observés précédemment ce qui confirme l'évolution de la mortalité par génération. De plus, une légère différence existe ici entre les 2 coefficients ce qui laisse supposer que le sexe est moins bien paramétré sur ces 2 tables.

Par ailleurs le fait que les 2 coefficients soient similaires, est cohérent avec les résultats trouvés dans la partie précédente. En effet, en utilisant les âges, comme ce qui a été effectué ici, le sexe n'était pas significatif dans le GLM par référence. Ainsi, si le coefficient d'abattement est le même pour les 2 sexes il est normal d'avoir trouvé précédemment que le sexe n'était pas significatif. Les lois de référence étant déjà séparés par sexe, ce critère n'est pas significatif car les données se comportent de la même manière mais par rapport à des références différentes.

Le deuxième critère pris en compte sera la tranche d'âge. L'âge, et donc la génération, sont très importantes dans les lois de mortalité. En outre, il s'agit d'une variable d'intérêt pour cette étude.

Chaque $q_{x,theorique}$ est d'abord recalculé comme étant le produit entre le $q_{x,theorique}$ initial et le ratio SMR correspondant. Cela permet de neutraliser les effets de la variable sexe.

La courbe des coefficients obtenus est présentée ci-dessous.

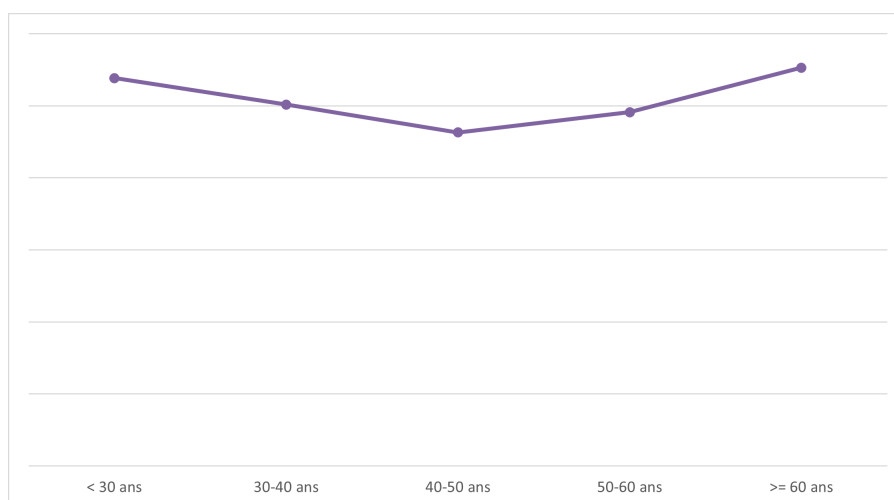


FIGURE 5.25 – Coefficients d'abattement par tranche d'âge par rapport aux tables TH 00-02 et TF 00-02 abattues

Les coefficients obtenus sont assez proches du niveau d'équilibre qui est égale à 1. Une variation maximale de 11% peut être observée. Cela indique que nos données sont, ici encore, en accord avec la forme des lois TH 00-02 et TF 00-02.

A noter que la distinction par âge est préservée ce qui permet de conserver une loi finale par âge. En effet, les coefficients de chaque tranche d'âge seront appliqués à tous les q_x différents de la tranche d'âge associée. La distinction par sexe et âge est maintenue.

Ici encore les lois de l'Ined de 2020 sont utilisées afin de comparer les résultats. Les tables de l'Ined de 2020 ayant été abattues des coefficients trouvés pour le critère sexe en utilisant ces tables à l'étape 1.

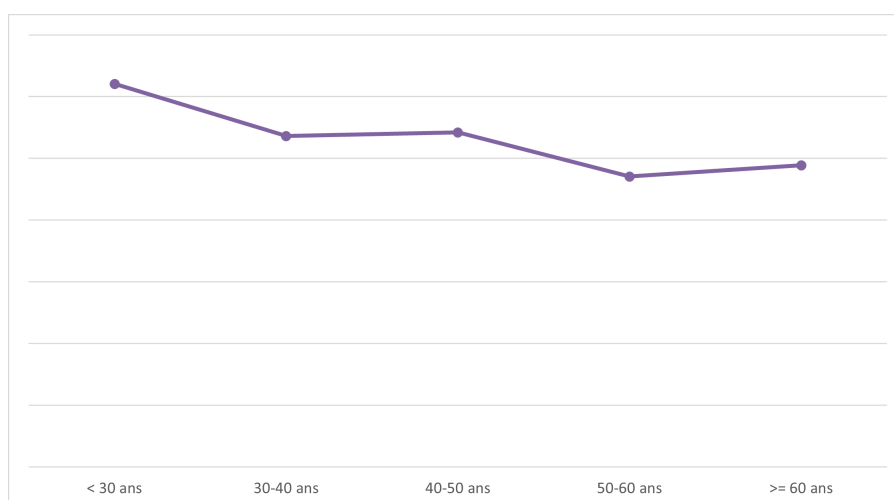


FIGURE 5.26 – Coefficients d'abattement par tranche d'âge par rapport aux tables Ined de 2020 abattues

Dans ce cas les résultats sont différents. En effet, une plus grande variation apparaît, avec presque 25% de variation pour la tranche des moins de 30 ans. Ainsi, ces tables, bien qu'ayant des coefficients de mortalité plus adaptés à la population actuelle car moins élevés, semblent moins bien adaptées au niveau du sexe et de l'âge.

3. <http://onala.free.fr/inegamor.pdf>

4. <https://www.ined.fr/fr/tout-savoir-population/chiffres/france/mortalite-cause-deces/table-mortalite/>

Le prochain critère étudié est celui de la CSP. En effet, après le sexe et l'âge cette variable est apparue comme étant très discriminante dans les modèles précédents.

Toutefois dans un premier temps, la CSP croisée avec le niveau de salaire calculé selon la CSP avait été étudié. Cependant, les résultats obtenus n'étaient pas cohérents et en accord avec les connaissances métier. En effet, les individus les plus riches avaient une mortalité plus élevée que celle des plus pauvres. Plusieurs pistes ont été envisagées afin de vérifier s'il s'agissait d'un problème de méthode, de données ou s'il s'agissait bien des véritables conclusions de l'étude. Ainsi, différentes versions ont été testées mais ces dernières conduisaient aux mêmes résultats.

Néanmoins, des incohérences dans la qualité des données avaient été observées dès le début de l'étude. Par exemple une correction avait été apportée aux temps partiels afin de ramener leurs salaires à des équivalents temps plein. C'est pourquoi, la répartition par niveau de salaire avait été mise en place afin de lisser ces incohérences. Cependant, après une étude précise des résultats, il s'agit bien d'un problème de qualité des données de la variable salaire. En effet, une analyse par CSP, niveau de salaire et tranche d'âge a été réalisée sur les résultats obtenus et des résultats aberrants ont été observés. Sur le niveau de salaire le plus élevé des résultats incohérents ont été observés avec des q_x qui n'était pas du tout dans l'intervalle de confiance par rapport au q_x moyen de la tranche d'âge et de la CSP en question. Certains q_x obtenus pour le niveau de salaire 5 étaient en effet plus de 4 fois supérieurs au q_x moyen de la CSP et de la tranche d'âge en question. En outre, ces résultats erronés ont été observés à plusieurs reprises sur ce niveau de salaire ce qui conduisait aux résultats aberrants obtenus. Le problème provient donc de la qualité de la variable. C'est pourquoi cette variable a été finalement écartée de l'étude. De plus, la variable CSP est corrélée au salaire et capture donc déjà cet effet dans les modèles. Cela peut s'expliquer par le fait que les professions des cadres sont similaires au niveau des conditions de travail, c'est-à-dire le plus souvent ce sont des métiers bureautiques, donc les causes de décès sont les mêmes quelques soit le salaire. Alors que les ouvriers peuvent être qualifiés ou non, et avoir des conditions de travail plus ou moins dangereuses, ce qui peut expliquer le fait que le salaire soit une variable plus impactante dans ce cas.

Finalement, le troisième critère pris en compte sera donc la CSP et le salaire quant à lui est écarté de l'étude.

Le $q_{x,theorique}$ est recalculé comme étant le produit entre le $q_{x,theorique}$ abattu précédent et le ratio SMR correspondant. Cela permet de neutraliser les effets de la variable âge en fonction de la tranche associée.

Les coefficients obtenus selon la CSP sont présentés ci-dessous.

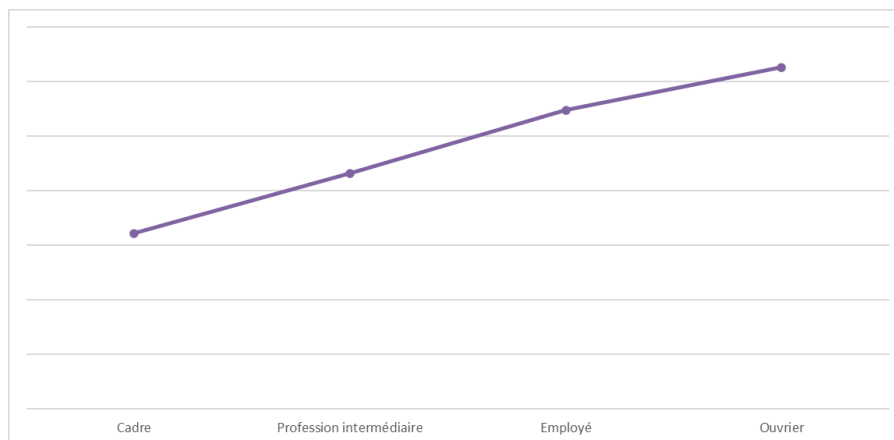


FIGURE 5.27 – Coefficients d'abattement par CSP par rapport aux tables TH 00-02 et TF 00-02 abattues

Le résultat obtenu est cohérent avec les méthodes précédentes ainsi qu'avec ce qui était attendu par rapport aux connaissances métier. Toutefois, on peut remarquer une différence importante entre les employés et les ouvriers qui peut surprendre. En effet, une différence d'environ 15% est observée entre les 2 coefficients. Une

telle différence n'avait pas été observée avec les modèles précédents. Ainsi, une séparation plus précise, par exemple des ouvriers qualifiés ou non, pourrait être envisagée afin de mieux observer certains effets qui sont peut-être masqués ici.

Par ailleurs, la mortalité des cadres est bien inférieure à celle des ouvriers. Cela semble juste par rapport aux connaissances métier ainsi qu'aux résultats précédents.

A savoir que la région de résidence a été étudiée à la suite de ces variables. Cependant, et cela était attendu par rapport aux résultats des modèles précédents, beaucoup de régions n'ont pas des coefficients très démarqués, comme on peut le voir ci-dessous. Ainsi, si la région devait être retenue alors les mêmes regroupements que précédemment devraient être effectués. Ces regroupements ont donc été testés. Cependant, après étude des résidus, du R^2 , et de l'estimation du coefficient de dispersion via la déviance, la région même avec les regroupements n'a pas été retenue car l'ajustement obtenu est moins bon que sans. En effet, certains résidus de Pearson ne sont pas dans l'intervalle $[-2; 2]$ alors que pour le modèle sans la région ils le sont tous. De même, les alternances de signes des résidus de la déviance sont moins présentes quand la région est prise en compte. De plus, le R^2 est plus faible pour le modèle avec la région et l'estimation du paramètre de dispersion est très éloignée de 1 alors qu'elle est très proche de 1 dans le modèle sans la région.

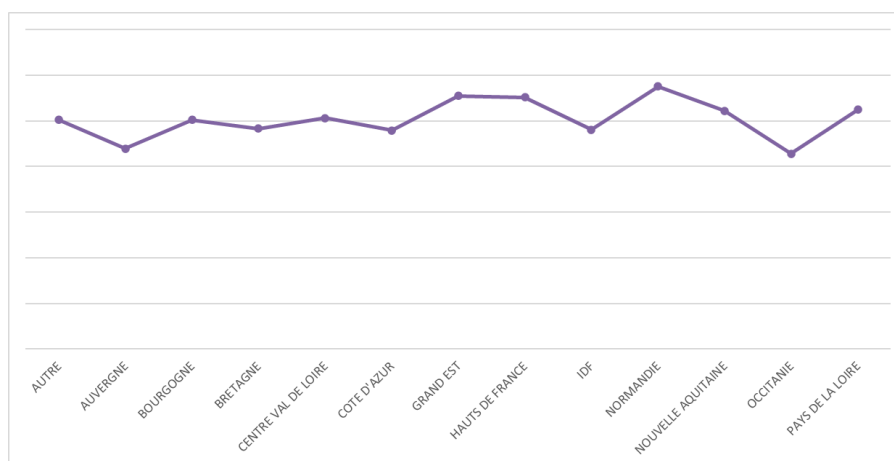


FIGURE 5.28 – Coefficients d'abattement par région par rapport aux tables TH 00-02 et TF 00-02 abattues

Le modèle retenu est donc celui corrigeant le sexe puis la tranche d'âge et enfin la CSP. Les résidus obtenus sont présentés afin d'illustrer les conclusions explicitées précédemment.

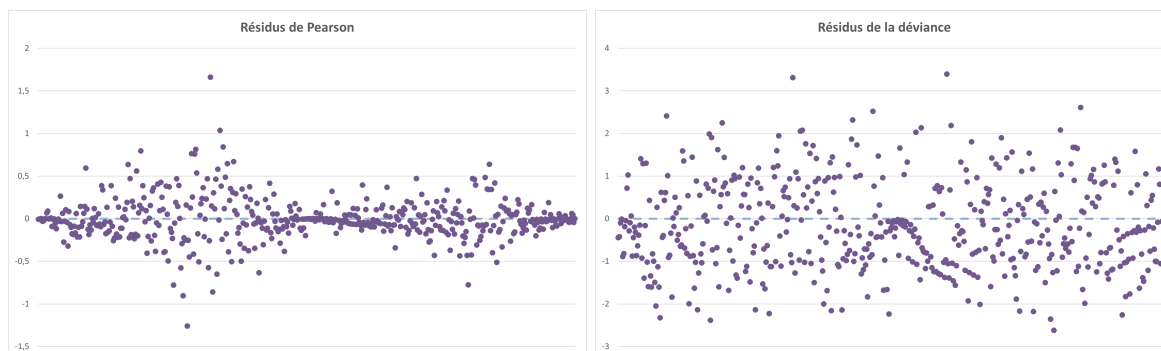


FIGURE 5.29 – Résidus de Pearson et de la déviance

Comme expliqué auparavant, des intervalles de confiance à titre indicatif ont été calculés afin de permettre de se rendre compte de la fiabilité des estimations. Ils sont présentés ci-dessous.

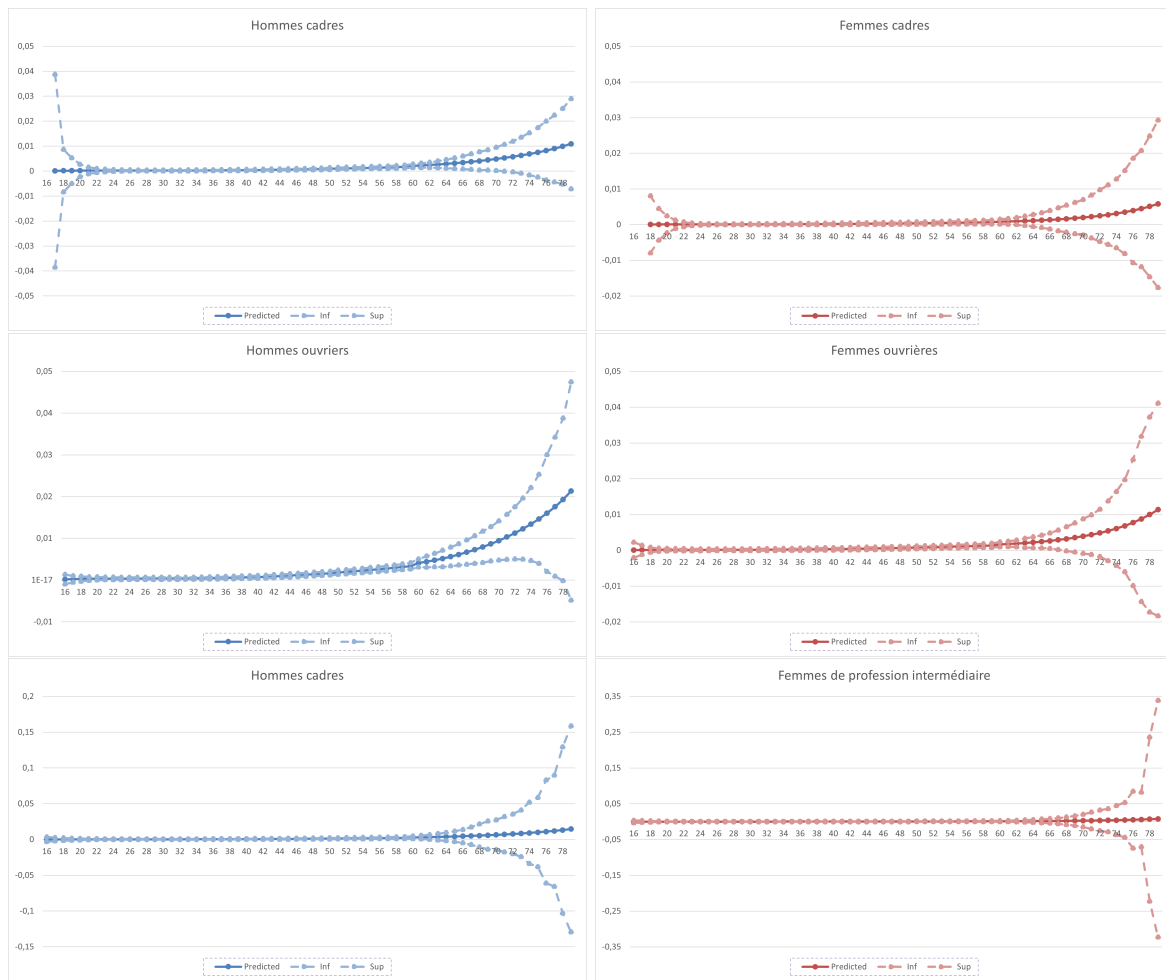


FIGURE 5.30 – Probabilités de décès, avec intervalle de confiance, en fonction du sexe, de l'âge et de la CSP

Ces résultats sont cohérents avec les remarques effectuées auparavant. Les cadres aux âges extrêmes n'étant pas très présents dans les données, ces estimations sont très incertaines. Cependant, le fait d'avoir positionné nos coefficients par rapport aux tables de référence permet de garder une certaine confiance par rapport aux résultats obtenus. Cela est très adapté aux âges jeunes. Pour les âges les plus élevés, la même observation est présente quelque soit la CSP. L'exposition étant faible les intervalles de confiance s'élargissent.

Par ailleurs, aux niveaux des tranches d'âges les jonctions ne sont pas lisses, un saut peut être observé. Cela peut éventuellement être corrigé en effectuant un lissage sur les courbes obtenues. Cela ne sera cependant pas présenté dans cette étude.

Quelques graphiques sont présentés ci-dessous afin de voir les courbes abattues obtenues. La proportion des décès suite à arrêt de travail est rajoutée de manière arbitraire afin de pouvoir comparer avec les lois de référence.

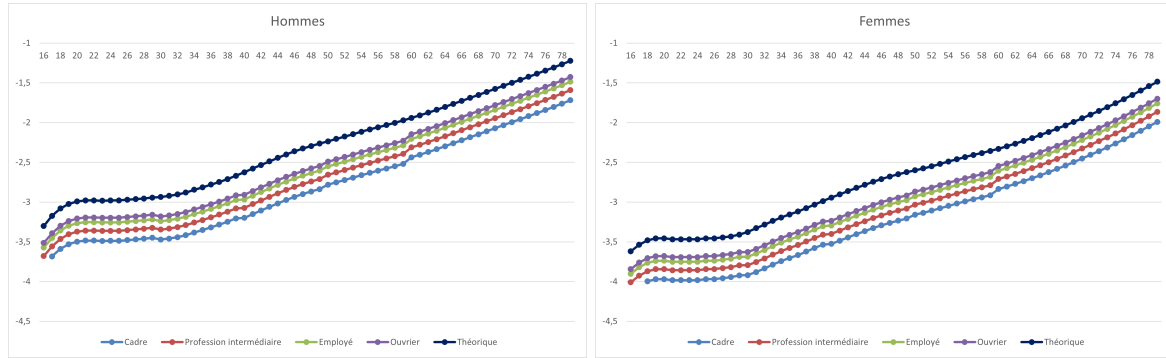


FIGURE 5.31 – Probabilités de décès, en échelle logarithmique, en fonction du sexe, de l'âge et de la CSP

Une étude intéressante est celle des classes tarifaires de l'outil de tarification actuel. En effet, afin de voir les impacts de ce critère celui-ci a tout de même été étudié. Dans un premier temps le critère a été appliqué après que le sexe, l'âge et la CSP aient été corrigés. La non significativité du critère est bien visible car les 3 catégories ont des coefficients très proches de 1, avec moins de 3% d'écart avec l'équilibre.

Ainsi, le critère est dans un second temps étudié après avoir corrigé le sexe et l'âge. Les résultats obtenus sont très intéressants. En effet, l'effet est plus prononcé ici avec un écart de 17% avec l'équilibre au maximum. De plus, les coefficients augmentent en fonction de la classe tarifaire. Ainsi, cela démontre que ces classes sont bien paramétrées car la classe 1 correspond bien à la classe avec la mortalité la plus faible et la troisième à l'inverse à celle avec la mortalité la plus élevée. Cependant, le critère de la CSP est le plus impactant des 2 et seul ce dernier est donc significatif. En effet, en appliquant par la suite la correction sur le critère CSP les coefficients varient encore de manière importante par rapport à l'équilibre ce qui démontre que les classes tarifaires non pas corrigés entièrement cet effet.

Afin de voir si cet effet était donc masqué par une partie de la population, le modèle est réappliqué aux non-cadres seulement puis dans un second temps aux ouvriers seulement. Le sexe, la tranche d'âge et la CSP sont corrigés dans un premier temps puis la classe tarifaire. Malgré cela, les résultats ne sont toujours pas très distincts selon les classes tarifaires avec moins de 6% de différence avec l'équilibre au maximum.

Pour conclure, cette méthode permet d'interpréter les coefficients en ayant corrigé les effets des autres critères. Elle permet donc d'interpréter facilement les résultats et d'observer les effets des différents critères. Ces coefficients peuvent être facilement utilisés afin de tarifier le décès direct.

Des améliorations sont tout de même envisageables. Le coefficient peut être calculé à partir d'une autre méthode que le SMR. Par exemple une régression peut être effectuée pour calculer le coefficient d'abattement. En outre, un lissage des courbes obtenues peut être effectué, ou des tranches d'âge plus restreintes peuvent être mises en place.

5.2.6 Comparaison des résultats obtenus avec les différentes méthodes

Finalement, 4 modèles principaux ont été retenus lors de l'étude après validation de l'adéquation des modèles. Le tableau suivant permet de résumer les variables d'intérêt dans chacun d'eux. A noter, que la couleur orange signifie que le critère n'était pas présent dans le modèle mais que les résultats sont discrétisés à cette maille. En effet, pour les méthodes par référence, la discrétisation des résultats est basée au minimum sur celle des lois de référence utilisées. Par exemple, pour le GLM Poisson avec référence en utilisant la variable âge, le sexe et l'âge n'étaient pas significatifs mais comme ce modèle se base sur les lois TH 00-02 et TF 00-02, qui sont discrétisés

par sexe et âge, alors les résultats finaux le sont également.

De plus, il est évident que si les résultats sont discrétisés à la maille âge alors ils le sont également par tranche d'âge.

En ce qui concerne, les éléments en bleu, ils symbolisent le fait que des regroupements ont été effectués afin de considérer toutes les modalités comme significatives.

Variables	GLM Poisson 1	GLM Poisson 2	GLM Poisson avec référence avec âge	SMR par critère
Sexe	1	1	1	1
Age	1	0	1	1
Tranche d'âge	0	1	0	1
CSP	1	1	1	1
Région	1	1	0	0
CSP x Age	0	0	1	0

FIGURE 5.32 – Tableau résumant les variables utilisées dans les différents modèles

Ainsi, une seule méthode n'est pas discrétisée par âge mais seulement par tranche d'âge. Les 2 dernières méthodes ne prennent pas du tout en compte la région. Elles sont celles avec les meilleurs résultats.

En effet, le fait de garder la précision par âge permet plus de précision dans les calculs. De plus, ces 2 méthodes sont basées sur les tables de références. Ainsi, ces 2 méthodes s'appuient sur des tables certifiées et proposent seulement des abattements en fonction des critères spécifiques de l'individu.

Finalement, ces 2 méthodes sont complémentaires car chacune d'entre elles ont des avantages et des inconvénients. En effet, le GLM Poisson avec référence permet de connaître le poids de chaque variable ainsi que l'incertitude et la significativité des critères et des modalités. Cependant, une hypothèse fondamentale afin d'utiliser une telle méthode est le fait que les variables sont supposées indépendantes entre elles. Or, comme vu précédemment il a été démontré qu'elles sont toutes corrélées entre elles et que la corrélation est plus ou moins intense selon le V de Cramer obtenu. Ainsi, les résultats obtenus doivent être pris avec précaution afin de ne pas tirer de conclusions erronées.

Par ailleurs, le SMR par critère est une méthode très intéressante car elle permet de choisir l'ordre des critères et d'étudier ces derniers un par un. Ainsi, chaque nouveau critère permet d'expliquer le bruit résiduel dans les données. En outre, très peu d'hypothèses sont nécessaires afin d'appliquer cette méthode. Cependant, cette méthode ne permet pas explicitement de connaître la significativité des critères bien que suivant les critères obtenus il peut être évident que ce n'est pas le cas. De plus, l'incertitude n'est également pas quantifiée ce qui peut mener à une mauvaise interprétation des résultats si des précautions ne sont pas prises. Néanmoins, des tests peuvent être mis en place, comme ce fut le cas ici, afin de mesurer ces incertitudes.

De plus, les tranches d'âge ayant été utilisées ici pour le SMR par critère, les courbes ne sont pas lissées. Mais cela peut être facilement réalisé à l'aide d'un lissage par splines, par exemple.

En conclusion, ces 2 méthodes sont complémentaires. Afin de savoir si les résultats obtenus sont significativement différents, ils sont comparés. Les taux d'évolutions par rapport aux probabilités de référence ont été calculés dans un premier temps pour chacune des méthodes puis la différence entre ces taux a été calculée afin de comparer les 2 méthodes entre elles.

Ainsi, les écarts sont assez faibles. Ils varient de -13% à 16% . Ces valeurs extrêmes sont observées pour les professions intermédiaires. Cela est normal que cette CSP corresponde à celle avec les écarts les plus importants car c'est celle avec la plus petite exposition et par conséquent celle avec l'incertitude la plus élevée. Pour les

cadres les écarts varient de -8% à 7% . Pour les employés ils varient extrêmement peu, de -3% à 2% . Pour les ouvriers les écarts varient de -7% à 5% .

Les écarts les plus importants étant observés sur les âges extrêmes, qui correspondent à ceux avec l'exposition la plus faible et donc l'incertitude la plus élevée.

Cela permet donc de valider les résultats obtenus par ces 2 méthodes, avec une incertitude maximale d'environ 16% sur les résultats. Ainsi, cela permet de fiabiliser les résultats et de les valider à l'aide de méthodologies différentes.

La méthode retenue pour la tarification sera la méthode SMR par critère car cette méthode à l'avantage de corriger dans l'ordre souhaité et pas à pas l'effet des différents critères. De plus, la méthode est explicite et ne suppose pas d'hypothèses préliminaires au traitement.

Chapitre 6

Tarification

6.1 Outil actuel

Pour rappel, l’outil de tarification actuel est basé sur les lois de mortalité 96-98, publiées par l’Insee. Des abattements en fonction du sexe et de la CSP sont effectués. Puis des coefficients, permettant de majorer le tarif en fonction de classes tarifaires qui dépendent du secteur d’activité, sont appliqués. Pour finir, un abattement général de 10% est effectué car ces normes datent de quelques années et cela permet donc de prendre en compte l’évolution de la mortalité.

Afin de comparer les lois de mortalité utilisées dans la tarification avec celles obtenues dans cette étude, un abattement supplémentaire correspondant à la rentabilité moyenne observée est réalisé sur les lois utilisées dans l’outil actuel, afin de comparer des lois équivalentes. En effet, les lois obtenues à travers cette étude n’intègrent pas de marge de rentabilité car le but de la tarification est d’égaliser les valeurs actuelles probables des engagements de l’assureur avec ceux de l’assuré. Alors qu’avec celles utilisées dans l’outil actuel, une marge est observée.

Pour finir, les probabilités de décès sont comparées entre 20 et 70 ans inclus car l’outil de tarification actuel ne permet pas de tarifier sur des âges plus jeunes ou plus élevés. Cela permet, en outre, de réaliser la comparaison sur un intervalle où l’incertitude, quant aux résultats obtenus, est faible et donc d’améliorer la qualité de l’analyse.

6.2 Comparaison des lois de mortalité

En ce qui concerne les probabilités de décès prédites à travers cette étude, ce sont celles obtenues via la méthode de positionnement à l’aide du SMR qui seront utilisées ici. Afin de comparer des périmètres équivalents, la proportion des décès suite à arrêt de travail est rajoutée arbitrairement. Cependant, cela est un point d’attention et d’amélioration de l’étude car ces décès suite à arrêt de travail ne possèdent pas nécessairement la même évolution que ceux sans arrêt de travail. Par exemple, l’abattement par tranche d’âge ou CSP n’est pas nécessairement identique à celui obtenu dans le cadre des décès sans arrêt de travail.

D’autre part, il faut noter que les lois ne sont pas basées sur les mêmes lois de références. En effet, les probabilités de décès de cette étude sont ajustées à partir des lois TH 00-02 et TF 00-02 alors que celles de l’outil actuel le sont à partir des lois 96-98 de l’Insee.

Les ratios entre les valeurs prédites par cette étude et celles utilisées par l’outil de tarification, abattues par rapport à la rentabilité observée, sont calculés afin de mesurer les écarts. Les écarts avec les 3 différentes classes tarifaires de l’outil de tarification sont calculés mais la classe 1 est celle de référence car elle correspond à un risque neutre. Les 2 autres classes permettent de majorer le tarif en cas de secteur d’activité plus risqué.

A noter que cette répartition est corrélée avec la CSP, par exemple moins de 6,5% des ouvriers font partis de la première classe tarifaire dans cette étude alors que 65% des ouvriers font partis de la troisième classe tarifaire. Ainsi, afin de comparer les tarifs des ouvriers, la première classe ne sera pas étudiée et une attention plus particulière sera portée à la troisième classe.

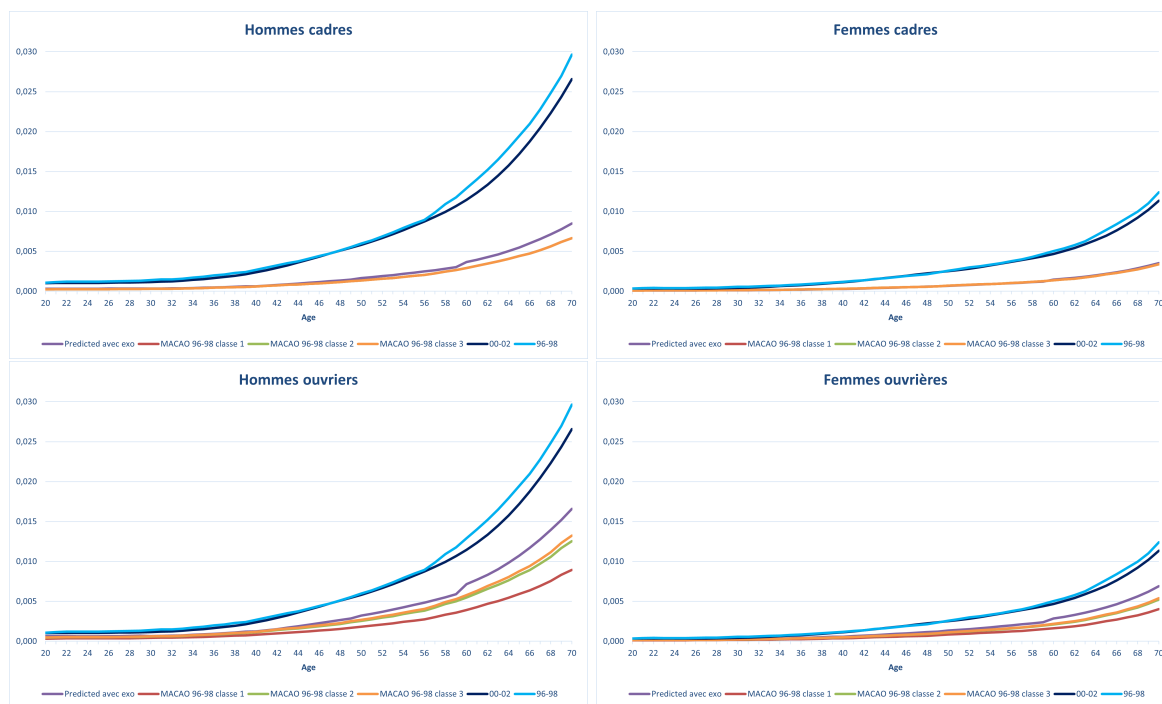


FIGURE 6.1 – Comparaison entre les différentes lois de mortalités selon le sexe et la CSP

Premièrement, pour les cadres, l’outil de tarification ne comporte qu’une seule classe tarifaire. Cela est cohérent avec le fait que les conditions de travail des cadres sont uniformes quelque soit le secteur d’activité dans lequel la personne exerce.

Pour les hommes, les probabilités de décès prédites sont supérieures à celles de l’outil de tarification. L’écart le plus important étant au niveau des âges extrêmes avec un écart maximal de 30%. En revanche, entre 30 et 44 ans les écarts sont plus faibles, ils fluctuent entre 5% et 13% au maximum.

Pour les femmes cadres, les valeurs prédites sont pour la plupart inférieures à celles de l’outil de tarification. L’écart le plus important étant de -21% pour les femmes âgées de 30 ans. En revanche, à 20 ans et à plus de 60 ans les probabilités de décès prédites sont supérieures à celles de l’outil de tarification. Cependant les écarts sont extrêmement faibles, moins de 7%. Ainsi, les courbes semblent très similaires.

D’autre part, pour les ouvriers, l’outil de tarification comprend bien 3 classes tarifaires. Néanmoins, comme expliqué auparavant très peu d’ouvriers de l’étude font partie de la première classe tarifaire ainsi il n’est pas possible d’utiliser cette classe dans l’étude des similitudes.

Pour les hommes ouvriers, les écarts sont plus importants. Pour tous les âges, les probabilités de décès prédites sont supérieures à celles de l’outil de tarification. Pour la seconde classe, les écarts sont entre 9 et 35%. Pour la troisième classe, la plus risquée mais aussi celle dont font partie la majorité des ouvriers de l’étude, les écarts se réduisent et vont de 3% à 29%. Ici encore les écarts les plus importants se situent sur les âges extrêmes.

Pour les femmes ouvrières, les écarts sont dans le même ordre de grandeur que pour les hommes. Pour la quasi majorité des cas, les probabilités de décès prédites sont supérieures à celles de l’outil de tarification. Pour la seconde classe, les écarts vont de 0 à 36%. Pour la troisième classe, les écarts se réduisent se situent entre -4% et 32%. Ainsi, les écarts sont parfois très faibles et notamment sur les âges médians.

Finalement, pour les employés et les professions intermédiaires, l'outil regroupe ces 2 CSP sous la notion ETAM. Ici encore les 3 classes tarifaires sont bien présentes.

Pour les hommes employés, toutes les probabilités de décès prédites sont supérieures à celles de l'outil de tarification. Pour la première classe, les écarts sont très importants, entre 33% et 68%. Les écarts se réduisent légèrement avec les autres classes tarifaires. Pour la seconde classe, les écarts vont de 29 à 63%. Pour la troisième classe, la plus risquée, les écarts se réduisent encore mais restent très importants et vont de 25% à 55%.

Pour les hommes de professions intermédiaires les écarts sont plus faibles. Mais les probabilités de décès prédites sont dans la grande majorité des cas, supérieures à celles de l'outil de tarification. Pour la première classe, les écarts sont compris entre 4% et 32%. Les écarts se réduisent, ici encore, avec les autres classes tarifaires. Pour la seconde classe, les écarts vont de 1 à 28%. Pour la troisième classe, les écarts se situent entre -2% à 22%.

Pour les femmes employées, toutes les probabilités de décès prédites sont supérieures à celles de l'outil de tarification. Pour la première classe, les écarts sont importants, entre 15% et 54%. Les écarts se réduisent légèrement cependant avec les autres classes tarifaires. Pour la seconde classe, les écarts vont de 9 à 49%. Pour la troisième classe, les écarts se réduisent encore mais restent importants et vont de 7% à 45%.

Pour les femmes de professions intermédiaires, les probabilités de décès calculées et celles de l'outil sont en accord avec des écarts plus faibles. Pour la première classe, les écarts sont compris entre -10% et 21%. Pour la seconde classe, les écarts vont de -14% à 17%. Pour la troisième classe, les écarts vont de -16% à 14%.

Ainsi, cela démontre que la modalité ETAM n'est pas assez précise et regroupe un ensemble de métiers avec des risques différents et notamment que les employés et les professions intermédiaires n'ont pas le même risque. En effet, la mortalité des employés semble sous-estimée à l'heure actuelle alors que celle des professions intermédiaires semble plus en adéquation avec la mortalité prédite pour les ETAM. En outre, moins de 10% des employés étudiés se trouvent dans la troisième classe tarifaire alors que c'est celle avec les plus faibles écarts pour cette CSP.

Une remarque générale est que les plus de 60 ans sont les âges sur lesquels les écarts sont les plus importants quelque soit la CSP. Ainsi, la mortalité de ces individus est sous-estimée à l'heure actuelle. Cependant, la tarification se fait par rapport à l'âge actuariel des individus de l'entreprise. Ainsi, ces âges ne sont donc que très rarement utilisés dans la pratique.

Finalement, il est donc intéressant de voir que les mailles actuelles de l'outil de tarification ne sont pas nécessairement les plus adaptées au risque réel. Le secteur d'activité ne semble pas significatif dans les analyses réalisées car trop corrélé à la CSP. Par conséquent, des CSP plus spécifiques, tout en gardant assez d'exposition, semblent plus adaptées à la tarification plutôt que l'utilisation de CSP trop globales ainsi que des secteurs d'activité.

Cependant, cette étude concerne seulement les décès directs. Ainsi, une étude complémentaire sur les décès faisant suite à un arrêt de travail devra être réalisée afin de démontrer les critères significatifs dans ce cas, qui seront probablement différents, et les abattements à réaliser par rapport aux lois réglementaires.

Ainsi, il s'agit d'une des limites les plus importantes dans cette étude. En effet, afin de réaliser des comparaisons avec des lois de mortalité comprenant l'ensemble des décès, la proportion moyenne observée de décès suite à arrêt de travail a été rajoutée à chaque fois. Cependant, cette remise à niveau est trop générale et ne prend pas en compte le fait que la répartition des décès suite à un arrêt de travail n'est très certainement pas la même que celle étudiée ici. Il avait été observé, par exemple, que l'âge moyen des décès suite à arrêt de travail était supérieur à celui des décès directs. De ce fait, c'est en combinant les résultats obtenus dans ces deux études que de vraies conclusions pourront être faites.

Conclusion

L'objectif de ce mémoire est de proposer une méthode de tarification du décès ne faisant pas suite à arrêt de travail, en prévoyance collective.

En effet, le décès suite à exonération étant rattaché à un arrêt de travail, il se traite de manière particulière. De plus, les causes de décès ne sont pas nécessairement les mêmes, ainsi une tarification spécifique est envisagée pour ce risque. Une méthode par positionnement par rapport aux lois de référence est l'approche privilégiée.

La problématique était donc la suivante : comment se comporte le portefeuille par rapport aux lois de référence ?

Dans un premier temps, un long travail sur les données a été réalisé afin d'obtenir les périodes d'observations des individus en retraitant les arrêts de travail. Des hypothèses comme la durée de la franchise, ont été choisies. Ensuite, une étape de fiabilisation a été effectuée avec les données de l'Insee. Cela a permis de révéler que les déclarations des décès en DSN sont fiables.

Les statistiques sur les données internes ont permis de comparer le décès direct et le décès suite à arrêt de travail. Ainsi, l'âge moyen des individus décédés suite à arrêt de travail est supérieur à celui des individus décédés sans arrêt de travail.

Les statistiques sur les données montrent que la population étudiée est composée de plus de 56% d'hommes mais qu'ils représentent plus de 76% des décès observés. La surmortalité des hommes est donc très marquée ici.

Les différentes modélisations ont démontré que les variables les plus discriminantes sont le sexe et l'âge (ou tranche d'âge). Cela est cohérent avec les lois de référence qui sont seulement segmentées par rapport à l'âge et au sexe. Un résultat très intéressant de cette étude est l'importance de la CSP. En effet, il s'agit d'un critère retenu dans tous les modèles et comportant de grandes disparités entre les différentes modalités. De plus, il a été vu que ce critère a tellement d'impact, et comme il est corrélé avec d'autres variables, comme le salaire ou le secteur d'activité, qu'il ne permet pas aux autres critères d'avoir un effet significatif sur les modèles. La région est également un critère significatif, mais ce critère n'est pas nécessairement retenu car il peut dans certains cas détériorer l'ajustement du modèle aux données.

De plus, tous les modèles ont démontré que la mortalité du portefeuille étudié est bien inférieure à celle des lois de référence. Ces dernières datent de quelques années et ne sont plus adaptées à la mortalité actuelle. De plus, elles sont réalisées sur la population française et sont donc trop prudentes pour un portefeuille de salariés.

En outre, la méthode calculant un coefficient d'abattement de manière itérative selon différents critères, à l'aide du SMR, semble la plus appropriée pour la tarification. Elle permet de choisir l'ordre des critères étudiés et de corriger les effets de chaque critère au fur et à mesure. Cependant, les modélisations utilisant des GLM ont permis de déterminer rapidement les variables significatives. Ainsi, ces méthodes sont complémentaires. Il est intéressant de toutes les implémenter afin de confirmer les résultats obtenus dans chaque modèle. Les écarts observés, entre les différents résultats, étaient assez faibles et permettent donc de valider les résultats obtenus.

De plus, la comparaison avec les lois utilisées actuellement pour la tarification ont permis de conclure que sur certaines CSP la mortalité, en prenant en compte la marge moyenne observée actuellement, est sous-estimée.

Toutefois afin d'améliorer les résultats, des segmentations plus poussées sur certaines sous-populations sont envisageables avec par exemple des tranches d'âge différentes ou une segmentation des ouvriers en fonction de leur qualification ou non. Une étude plus poussée sur le salaire pourrait également permettre de fiabiliser cette variable afin qu'elle puisse être intégrée aux modèles.

Enfin, une étude est en cours, afin d'analyser les décès suite à arrêt de travail. Ces travaux sont donc complémentaires à ceux effectués dans cette étude et permettront l'obtention d'une vision complète du risque décès en prévoyance collective. Il sera alors intéressant de comparer les résultats avec ceux obtenus lors de cette étude afin de voir les différences entre les deux catégories de décès. Les critères n'étant pas nécessairement les mêmes entre ces deux types de décès. La combinaison de ses résultats permettra une tarification optimale du décès.

Annexes

1 Analyse de la répartition des décès

Nombre de dossiers échus par rapport à la première date d'opération		Survenance						
		2013	2014	2015	2016	2017	2018	2019
N0	janvier	0,9%	1,2%	0,7%	0,5%	0,7%	0,1%	0,4%
	février	4,8%	3,8%	4,2%	2,0%	3,1%	1,5%	0,9%
	mars	10,2%	10,2%	9,1%	7,6%	5,8%	4,2%	2,7%
	avril	17,3%	15,5%	13,6%	13,7%	10,8%	7,6%	5,9%
	mai	22,6%	21,5%	17,4%	19,6%	16,2%	10,5%	10,4%
	juin	29,3%	26,5%	26,6%	26,0%	24,1%	14,7%	17,4%
	juillet	37,9%	34,4%	33,2%	32,0%	29,4%	21,1%	23,4%
	août	43,3%	41,8%	40,3%	38,5%	35,9%	24,4%	27,9%
	septembre	50,9%	49,1%	48,7%	45,3%	42,7%	27,3%	34,3%
	octobre	59,2%	55,8%	54,6%	52,3%	49,6%	33,7%	43,2%
	novembre	69,2%	62,8%	60,8%	61,7%	57,5%	41,9%	49,0%
	décembre	77,0%	70,7%	70,3%	69,2%	62,7%	46,3%	57,3%
N1	janvier	84,7%	78,2%	76,6%	74,4%	68,0%	53,2%	62,2%
	février	88,6%	81,7%	80,4%	79,2%	72,9%	58,4%	67,5%
	mars	90,9%	86,0%	84,1%	82,7%	77,0%	65,5%	73,6%
	avril	92,4%	88,0%	86,8%	85,7%	80,7%	71,8%	76,7%
	mai	93,4%	89,2%	88,2%	87,3%	83,2%	77,9%	79,1%
	juin	94,1%	91,7%	90,5%	89,6%	85,1%	81,1%	81,2%
	juillet	95,5%	92,1%	91,9%	90,3%	86,6%	82,6%	82,9%
	août	95,8%	92,6%	92,3%	91,6%	87,6%	83,4%	84,6%
	septembre	96,2%	93,8%	93,5%	92,9%	88,5%	85,3%	86,1%
	octobre	96,6%	94,5%	93,6%	94,0%	89,3%	87,8%	87,6%
	novembre	96,6%	95,7%	94,4%	94,8%	90,3%	88,6%	89,7%
	décembre	97,1%	96,1%	95,9%	95,2%	91,0%	89,4%	90,7%
N2	janvier	97,5%	96,2%	95,9%	95,9%	91,7%	91,5%	91,1%
	février	97,5%	96,4%	96,0%	96,0%	92,2%	92,8%	91,9%
	mars	97,7%	96,6%	96,2%	96,1%	93,2%	93,6%	93,4%
	avril	97,8%	96,8%	96,3%	96,4%	94,2%	93,9%	94,1%
	mai	97,9%	97,2%	96,4%	96,5%	95,4%	94,1%	94,3%
	juin	98,0%	97,3%	96,6%	96,7%	95,6%	94,5%	94,9%
	juillet	98,1%	97,4%	96,8%	96,7%	95,8%	95,0%	95,5%
	août	98,4%	97,8%	97,0%	96,8%	96,0%	95,3%	95,6%
	septembre	98,4%	98,0%	97,1%	96,8%	96,2%	95,8%	96,2%
	octobre	98,5%	98,1%	97,2%	96,9%	96,9%	96,3%	96,6%
	novembre	98,7%	98,3%	97,5%	97,0%	97,0%	96,8%	96,7%
	décembre	98,7%	98,3%	97,5%	97,1%	97,0%	97,1%	96,9%
N3	janvier	98,7%	98,3%	97,6%	97,1%	97,3%	97,1%	97,1%
	février	98,9%	98,3%	97,6%	97,3%	97,4%	97,5%	97,8%
	mars	98,9%	98,3%	97,7%	97,5%	97,5%	98,0%	98,9%
	avril	98,9%	98,4%	97,9%	97,5%	97,5%	98,0%	99,3%
	mai	98,9%	98,5%	97,9%	98,0%	97,6%	98,1%	99,7%
	juin	99,0%	98,5%	97,9%	98,3%	97,6%	98,5%	99,9%
	juillet	99,0%	98,5%	97,9%	98,6%	97,6%	98,7%	100,0%
	août	99,1%	98,5%	97,9%	98,7%	97,7%	98,7%	99,9%
	septembre	99,1%	98,7%	97,9%	98,7%	98,0%	98,7%	99,9%
	octobre	99,1%	98,7%	97,9%	98,7%	98,0%	98,9%	99,9%
	novembre	99,2%	98,9%	98,1%	98,7%	98,3%	98,9%	99,9%
	décembre	99,2%	99,3%	98,1%	98,8%	98,6%	99,2%	99,9%

FIGURE 1 – Répartition par année des dossiers pour décès direct en pourcentage

Nombre de dossiers échus par rapport à la première date d'opération		Survenance						
		2013	2014	2015	2016	2017	2018	2019
N0	janvier	0,1%	0,8%	1,2%	0,5%	0,5%	0,5%	0,5%
	février	1,5%	3,8%	6,3%	1,9%	3,0%	1,3%	1,1%
	mars	5,9%	8,4%	12,4%	6,8%	7,2%	4,4%	3,1%
	avril	10,1%	12,3%	18,7%	12,4%	11,6%	9,1%	7,3%
	mai	14,6%	18,9%	24,6%	17,4%	16,2%	12,5%	13,1%
	juin	22,5%	25,1%	32,6%	24,3%	24,9%	18,4%	19,9%
	juillet	30,9%	35,8%	39,7%	31,5%	31,0%	25,0%	26,7%
	août	34,5%	41,0%	47,0%	40,8%	37,1%	28,7%	30,9%
	septembre	41,7%	48,5%	53,7%	49,2%	43,5%	32,2%	36,9%
	octobre	49,8%	59,9%	60,4%	55,3%	50,4%	37,2%	46,3%
	novembre	66,7%	67,6%	66,4%	63,2%	59,3%	43,1%	52,8%
	décembre	75,1%	75,0%	75,6%	70,3%	64,5%	47,7%	62,2%
N1	janvier	84,5%	81,8%	84,7%	78,4%	72,9%	53,7%	67,7%
	février	87,7%	86,1%	89,3%	82,5%	76,7%	58,2%	71,7%
	mars	91,0%	88,5%	91,8%	87,2%	80,8%	66,3%	77,6%
	avril	91,9%	89,5%	93,3%	89,6%	83,7%	73,4%	82,1%
	mai	93,2%	90,4%	93,9%	90,9%	86,7%	80,6%	83,6%
	juin	94,6%	92,5%	95,4%	92,5%	88,5%	84,3%	86,6%
	juillet	95,6%	93,0%	95,4%	93,5%	90,1%	86,5%	88,1%
	août	95,9%	93,4%	96,3%	93,9%	90,1%	87,4%	89,2%
	septembre	96,5%	93,8%	96,3%	94,0%	90,7%	88,3%	91,1%
	octobre	96,7%	94,5%	96,4%	95,3%	91,6%	90,1%	91,7%
	novembre	96,9%	95,0%	96,6%	95,8%	92,3%	91,0%	93,3%
	décembre	97,1%	95,5%	96,9%	96,1%	93,6%	91,9%	94,2%
N2	janvier	97,2%	96,8%	96,9%	96,9%	93,8%	93,8%	94,6%
	février	98,0%	96,1%	96,9%	96,9%	94,1%	94,0%	95,4%
	mars	98,5%	96,9%	97,2%	97,5%	95,0%	95,2%	95,9%
	avril	98,5%	97,1%	97,2%	97,7%	95,3%	95,4%	96,4%
	mai	98,6%	97,1%	97,2%	97,9%	95,9%	96,0%	96,7%
	juin	98,7%	97,4%	97,3%	98,1%	96,2%	96,2%	97,2%
	juillet	98,7%	97,6%	97,3%	98,1%	96,3%	96,5%	97,4%
	août	98,7%	97,7%	97,3%	98,1%	96,7%	96,7%	97,6%
	septembre	98,7%	97,7%	97,3%	98,1%	97,0%	96,9%	97,9%
	octobre	98,7%	97,8%	97,4%	98,1%	97,1%	97,4%	98,2%
	novembre	98,7%	98,1%	97,7%	98,1%	97,2%	97,7%	98,2%
	décembre	98,9%	98,1%	97,9%	98,2%	97,6%	97,9%	98,6%
N3	janvier	98,9%	98,1%	98,2%	98,2%	97,6%	97,9%	98,9%
	février	98,9%	98,1%	98,2%	98,2%	97,6%	98,1%	99,3%
	mars	99,1%	98,1%	98,2%	98,4%	98,0%	98,6%	99,7%
	avril	99,1%	98,3%	98,2%	98,4%	98,0%	98,8%	99,8%
	mai	99,2%	98,4%	98,2%	98,6%	98,0%	99,0%	99,8%
	juin	99,2%	98,4%	98,3%	98,6%	98,2%	99,2%	99,9%
	juillet	99,4%	98,4%	98,7%	98,6%	98,4%	99,3%	100,0%
	août	99,4%	98,4%	98,7%	98,7%	98,5%	99,3%	99,9%
	septembre	99,4%	98,4%	98,7%	98,8%	98,5%	99,3%	99,9%
	octobre	99,4%	98,4%	98,7%	99,0%	98,6%	99,4%	99,9%
	novembre	99,4%	98,4%	98,7%	99,0%	98,8%	99,4%	99,9%
	décembre	99,4%	98,4%	98,7%	99,0%	99,0%	99,4%	99,9%

FIGURE 2 – Répartition par année des dossiers pour décès suite à exonération en pourcentage

2 Analyse du décès sur un exemple

			N	N+1	N+2	N+3	N+4	N+5	N+6	N+7	N+8	N+9	N+10
Actifs	Sans changement de situation		9751,0	9642,3	9570,6	9519,1	9520,8	9521,5	9521,8	9521,8	9521,8	9521,8	9521,8
	Sortie d'incapacité	dont N-2			19,9	19,7	19,5	19,4	19,4	19,4	19,4	19,4	19,4
		dont N-1		36,2	35,8	35,5	35,3	35,3	35,3	35,3	35,3	35,3	35,3
		dont N	83,6	82,6	82,0	81,6	81,6	81,6	81,6	81,6	81,6	81,6	81,6
	Total		83,6	118,8	137,7	136,8	136,4	136,3	136,3	136,3	136,4	136,4	136,4
Nouveaux entrants		53,9	53,9	53,8	108,0	107,4	107,0	106,8	106,8	106,8	106,8	106,8	
Total			9888,5	9815,0	9762,1	9763,9	9764,6	9764,9	9764,9	9764,9	9764,9	9764,9	9764,9
AT/exo	Incapacité temporaire	dont N-2			74,7	73,9	73,3	72,9	73,0	73,0	73,0	73,0	73,0
		dont N-1		111,5	110,3	109,5	108,9	108,9	108,9	108,9	108,9	108,9	108,9
		dont N	199,1	196,9	195,5	194,4	194,4	194,5	194,5	194,5	194,5	194,5	194,5
	Total		199,1	308,4	380,4	377,7	376,6	376,3	376,3	376,3	376,3	376,3	376,3
	Passage d'incap à inval	dont N-2			0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1
		dont N-1		0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1
		dont N	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1
	Total		0,1	0,1	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2
	Invalidité	Ancienne		6,8	13,4	19,9	79,6	137,5	193,8	248,6	302,3	354,9	406,4
		Nouvelle	6,8	6,8	6,7	61,2	60,6	60,2	59,9	59,9	59,9	59,9	
		Total	6,8	13,5	26,1	81,1	140,2	197,7	253,7	308,6	362,3	414,9	466,4
	Incapacité finale	dont N-2			54,5	53,9	53,5	53,2	53,3	53,3	53,3	53,3	53,3
		dont N-1		74,7	73,9	73,3	72,9	73,0	73,0	73,0	73,0	73,0	73,0
		dont N	111,5	110,3	109,5	108,9	108,9	108,9	108,9	108,9	108,9	108,9	108,9
	Total		111,5	185,0	237,9	236,1	235,4	235,1	235,1	235,1	235,1	235,1	235,1
Invalidité finale	Ancienne		6,6	13,1	19,5	77,9	134,6	189,7	243,4	296,0	347,5	398,0	
	Nouvelle	6,8	6,8	6,8	60,1	59,5	59,2	58,9	58,9	58,9	58,9		
	Total	6,8	13,4	19,9	79,6	137,5	193,8	248,6	302,3	354,9	406,4	456,9	
Total			118,3	198,4	257,8	315,7	372,8	428,9	483,7	537,4	590,0	641,5	692,0
DC	Direct		43,0	42,5	42,2	42,0	42,0	42,0	42,0	42,0	42,0	42,0	42,0
	Exo	Suite à incap	dont N-2			0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2
			dont N-1		0,6	0,6	0,5	0,5	0,5	0,5	0,5	0,5	0,5
			dont N	4,0	3,9	3,9	3,9	3,9	3,9	3,9	3,9	3,9	3,9
	Total		4,0	4,5	4,7	4,7	4,7	4,7	4,7	4,7	4,7	4,7	
	Suite à invalidité	Ancienne		0,0	0,1	0,1	0,5	0,9	1,2	1,5	1,9	2,2	2,5
		Nouvelle	0,0	0,0	0,0	0,4	0,4	0,4	0,4	0,4	0,4	0,4	
		Total	0,0	0,1	0,1	0,5	0,9	1,2	1,6	1,9	2,2	2,6	2,9
	Total		4,0	4,6	4,8	5,2	5,5	5,9	6,2	6,6	6,9	7,2	7,5
	Total			47,0	47,1	47,0	47,1	47,5	47,9	48,2	48,6	48,9	49,2
Retraite	Sortie d'invalidité	Ancienne		0,1	0,2	0,3	1,2	2,0	2,8	3,6	4,4	5,2	5,9
		Nouvelle	0,1	0,1	0,1	0,9	0,9	0,9	0,9	0,9	0,9	0,9	
		Total	0,1	0,2	0,3	1,2	2,0	2,9	3,7	4,5	5,3	6,1	6,8
Taille de la nouvelle population (actifs et AT)			10 006,8	10 013,4	10 019,9	10 079,6	10 137,5	10 193,8	10 248,6	10 302,3	10 354,9	10 406,4	10 456,9

FIGURE 3 – Modèle entier d'un exemple d'évolution d'une population salarié

Bibliographie

- [1] Guillaume BIESSY. *Cours Dépendance*. 2022.
- [2] Guillaume BLETIO. Impact des lois de mortalité dans des représentations solvabilité II et IFRS 17. *Mémoire Actuariat*, 2018.
- [3] Maxence BOCCIARELLI. Impacts de la déliaison sur l'équilibre des contrats emprunteurs. *Mémoire Actuariat*, 2014.
- [4] INSEE Christian Monteil et Isabelle Robert-Bobée. Les différences sociales de mortalité : en augmentation chez les hommes, stables chez les femmes. <http://onala.free.fr/inegamor.pdf>, 2005.
- [5] Cindy CORNUAILLE. *Cours Prévoyance collective*. 2021.
- [6] L'argus de l'assurance. Prévoyance : Provisionner le décès des personnes en incapacité-invalidité. <https://www.argusdelassurance.com/social/prevoyanceprovisionner-le-deces-des-personnes-en-incapacite-invaliditeun-recent-article-insere-dans-la-loi-evin-concernant-les-personnes-en-incapacite-invalidite-vient-ajouter-une-strate.11448>, 2002.
- [7] Julien TOMAS et Frédéric PLANCHET. Méthodes de positionnement : aspects méthodologiques. https://www.institutdesactuaire.com/global/gene/link.php?doc_id=1175&fg=1.
- [8] Maylis HELIOT. Tarification du risque invalidité en prévoyance collective. *Mémoire Actuariat*, 2021.
- [9] INSEE. Temps et conditions de travail. <https://www.insee.fr/fr/statistiques/fichier/2891765/ecrt17j-f4-temps-travail.pdf>, 2017.
- [10] Michel LEJEUNE et Gilbert SAPORTA Jean-Jacques DROESBEKE. *Modèles statistiques pour données qualitatives*. 2005.
- [11] Adel KOB TAN. Construction d'une table de maintien en incapacité à partir des données dsn et impact sur le pilotage. *Mémoire Actuariat*, 2020.
- [12] Véronique LAMBLÉ. Provisions en prévoyance collective. <https://www.aprecialis.fr/wp-content/uploads/2020/10/Provisions-en-Prvoyance-collective.pdf>, 2008.
- [13] Olivier LOPEZ. *Cours de modèle de durée*. 2021.
- [14] Bertrand MOURELOT. Provision d'égalisation et solvabilité, étude d'un portefeuille de prévoyance collective décès. *Mémoire Actuariat*, 2009.
- [15] Caisse nationale d'assurance vieillesse. Cahier technique dsn. <https://www.net-entreprises.fr/media/documentation/dsn-cahier-technique-2022.1.pdf>, 2022.
- [16] Frédéric PLANCHET. *Cours modèles de durée - Statistique des modèles non paramétriques*. 2021.
- [17] Manon DAL PONT. Construction d'une table de mortalité d'expérience en assurance emprunteur. *Mémoire Actuariat*, 2020.
- [18] Maud THOMAS. *Cours d'économétrie dans l'assurance non-vie*. 2021.