

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

Par : Anicia BRAITHWAITE

*La tarification de l'incapacité temporaire sous franchise discontinuée*

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

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

*Membres présents du jury de  
l'Institut des Actuaires*

signature

*Entreprise : AXA France*

Nom :

*Signature :*

*Directeur de mémoire en  
entreprise :*

*Nom : Romain DECAMPS*

*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



Signature du candidat

Anicia BRAITHWAITE



# Remerciements

Dans un premier temps, j'adresse un immense merci à Romain DECAMPS, tuteur d'entreprise, pour sa confiance, sa sympathie, son aide continue, sa disponibilité et sa grande implication dans ce mémoire. Merci également à Anne FAYE, responsable de l'équipe *Normes et Études France* de m'avoir accordé sa confiance en m'accueillant au sein de son équipe et en me confiant la réalisation de ce beau projet. Je remercie également le reste de l'équipe, Anaïs, Laurence et El Hadji, pour votre accueil chaleureux et tous vos conseils durant cette année. Aussi je salue l'ensemble de la *Direction Technique et Innovation* pour sa bienveillance, en particulier Dongdong CHEN pour son aide dans l'analyse des données.

De même, j'adresse ma reconnaissance à Guillaume BIESSY, responsable académique pour son accompagnement consciencieux, sa disponibilité et sa grande contribution au bon déroulement de cette année. Également merci au corps académique de l'ISUP mené par Olivier LOPEZ pour avoir assuré ma formation tout au long de la période d'étude.

Enfin, mes vifs remerciements à ma famille et mes proches, en particulier ma soeur Laure-Anne, qui ont tous été présents pour moi dans ce projet et ont grandement participé à la relecture de ce mémoire.

# Résumé

Actuellement la tarification de la garantie incapacité temporaire est déterminée à partir de nombreux coefficients et correctifs de sexe, âge et franchise établis par le Bureau Commun d'Assurances Collectives (BCAC). Cependant, cette méthode de tarification ne permet pas d'obtenir une vision Best Estimate du coût de l'arrêt de travail comme recommande le deuxième pilier de la directive Solvabilité 2. Cela est notamment dû à la population analysée et à l'ancienneté de la base de données sur laquelle repose les estimations de coût du BCAC.

L'objectif de cette étude est donc l'élaboration d'une méthode de tarification visant à améliorer la tarification aujourd'hui utilisée par une méthode dite "Best Estimate". La réflexion s'articule essentiellement autour des garanties bénéficiant d'une franchise discontinue.

Afin d'obtenir le tarif le plus juste possible, les données analysées sont celles de la Déclaration Sociale Nominative (DSN). Cette base de données permet de récolter les arrêts de travail au premier jour des salariés dont l'établissement est assuré auprès d'AXA. Cela permet de refléter au mieux la sinistralité présente au sein d'un portefeuille propre à AXA. A partir de cette base, il est possible d'établir un tarif pour l'incapacité temporaire proche des coûts réels de ces arrêts au sein d'une entreprise.

Pour cela, il a fallu dans un premier temps établir la fréquence d'occurrence, appelée incidence, de l'évènement d'étude. Ensuite est déterminée la probabilité de maintien en arrêt de travail selon l'âge de l'assuré. A partir de ces deux tables, il est possible d'obtenir une tarification de la garantie incapacité temporaire de la forme "Coût x Fréquence". Enfin, les nouveaux tarifs établis sont comparés avec ce qui est actuellement fait par le BCAC.

**Mots clés :** Incapacité temporaire, franchise discontinue, arrêt de travail, table d'incidence, table de maintien.

# Abstract

Currently the pricing of the Short Term Disability is determined from the BCAC (Bureau Commun d'Assurances Collectives) numerous sex, age and waiting period coefficients and fixes. Unfortunately, this pricing method does not provide a Best Estimate vision of the cost of the work stoppage as recommended in the second pillar of the Solvency 2 Directive.

Therefore, the purpose of this study is to develop a pricing method aimed at improving the pricing currently used by a method known as "Best Estimate". The main focus is on guarantees benefiting from a discontinuous franchise.

In order to obtain the fairest rate, the data analysed is from the DSN (Déclaration Sociale Nominative) database. This database makes it possible to collect work stoppages on the first day of employees whose establishment is insured with AXA. Thus, it is possible to better reflect the claims present within a portfolio which is specific to AXA. From this database, it is also possible to establish a price for short term disability approximately identical to the actual costs of such stoppages within an enterprise.

To do so, it was first necessary to establish the frequency of occurrence, called incidence, of the study event. The probability of continuing to work despite an inability is then determined according to the age of the insured person. From both of these tables, it is possible to obtain a pricing of the disability guarantee in the form "Cost x Frequency". Finally, the new established rates are compared with what is currently done by the BCAC.

**Keywords :** Short term disability, discontinue waiting period, work stoppage, disability incidence table, disability duration table.

# Table des matières

<b>Introduction</b>	<b>6</b>
<b>1 Le contexte d'étude</b>	<b>7</b>
1.1 Le groupe AXA . . . . .	7
1.1.1 Activité et organisation du groupe . . . . .	7
1.1.2 Axa France . . . . .	8
1.1.3 Axa Santé & Collectives (ASC) . . . . .	8
1.2 La prévoyance collective . . . . .	9
1.2.1 La sécurité sociale . . . . .	10
1.2.2 Les régimes complémentaires . . . . .	11
1.2.3 Les régimes individuels . . . . .	12
1.3 Le risque incapacité temporaire . . . . .	12
1.3.1 L'indemnisation de l'incapacité . . . . .	12
1.3.2 Les types de franchises . . . . .	14
1.3.3 Le cadre législatif . . . . .	16
<b>2 Les données</b>	<b>18</b>
2.1 La déclaration sociale nominative (DSN) . . . . .	18
2.1.1 Présentation du portefeuille d'étude . . . . .	18
2.1.2 Fiabilité des données . . . . .	19
2.2 La base de données d'étude . . . . .	20
2.2.1 Le risque étudié . . . . .	20
2.2.2 Le périmètre d'observation . . . . .	21
2.3 Statistiques descriptives . . . . .	22
2.3.1 Analyse de la population assurée . . . . .	22
2.3.2 Analyse de la sinistralité . . . . .	25
<b>3 La table d'entrée en incapacité</b>	<b>30</b>
3.1 Notions clés de modèle de durée . . . . .	30
3.1.1 Les fonctions de base . . . . .	31
3.1.2 Les phénomènes d'observations incomplètes . . . . .	32
3.2 Méthodes d'estimation des taux d'incidence . . . . .	34
3.2.1 L'estimateur de Kaplan Meier . . . . .	35
3.2.2 L'estimateur de Hoem . . . . .	37
3.2.3 Résultats et choix . . . . .	38

---

3.3	Méthodes de lissage des taux bruts . . . . .	44
3.3.1	Lissage de Whittaker-Henderson . . . . .	44
3.3.2	Choix et tests de validation . . . . .	51
<b>4</b>	<b>La table de maintien en incapacité</b>	<b>53</b>
4.1	Idée de conception . . . . .	53
4.1.1	Présentation générale . . . . .	53
4.1.2	Les événements observés . . . . .	54
4.2	Elaboration de la table . . . . .	55
4.2.1	Estimation des taux bruts . . . . .	56
4.2.2	Lissage de Whittaker-Henderson . . . . .	59
4.2.3	Validation et résultats . . . . .	61
<b>5</b>	<b>La tarification</b>	<b>63</b>
5.1	Méthode de tarification "Best-Estimate" . . . . .	63
5.1.1	Tarification d'une tête . . . . .	63
5.1.2	Correctif âge actuariel . . . . .	65
5.2	Comparaison au tarif existant . . . . .	70
5.2.1	La tarification du BCAC . . . . .	70
5.2.2	Les autres impacts clés de la tarification "Best Estimate" . . . . .	73
	<b>Conclusion</b>	<b>75</b>
	<b>Bibliographie</b>	<b>77</b>

# Introduction

L'assurance étant confrontée à de nombreuses exigences réglementaires, la législation accorde la possibilité d'avoir recours à des tables d'expériences qui reflètent les probabilités de survenance d'un évènement observé en portefeuille à chaque âge. Il faut cependant noter que cette liberté n'est accordée que dans le cas de la tarification des produits. Le provisionnement reste quant à lui soumis à des tables et exigences très réglementées.

Actuellement, dans le cadre de l'arrêt de travail en assurance collective, les tables et barèmes communément utilisés pour la tarification sont celles du Bureau Commun d'Assurances Collectives (BCAC). Les barèmes du BCAC ont été construits sur la base d'observations menées sur les assurés des principales compagnies françaises d'assurance dans les années 1960. Cependant les caractéristiques de la population ont évolué depuis. Les nombreux correctifs apportés pour actualiser les résultats du BCAC ne cessent de créer des biais, ce qui éloigne les tarifs calculés de la vision "Best-Estimate" requis par la directive Solvabilité II.

L'assureur a donc tout intérêt à construire des tables qui lui sont propres en vue d'évaluer au mieux la sinistralité à laquelle il doit faire face. C'est pourquoi il a été décidé de construire à partir de la DSN (Déclaration Sociale Nominative), une méthode de tarification de l'arrêt de travail basée sur le portefeuille des assurances collectives d'AXA. La DSN est une source de données très intéressante à exploiter. En effet, en assurance collective, la population assurée n'est pas connue dans le détail. Les déclarations permettent donc d'avoir une meilleure vision des effectifs de chaque entité assurée. L'étude portera essentiellement sur la création de tables pour tarifier l'incapacité temporaire de la franchise discontinue qui jusqu'alors n'était déterminée que par un simple correctif porté sur le tarif de la franchise continue.

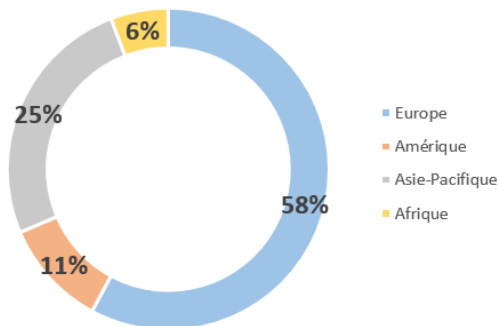
La première partie de ce rapport consiste à élaborer une base de données exploitable à partir de la DSN qui permettra la construction des tables nécessaires à la tarification de la garantie incapacité temporaire en franchise discontinue. Ensuite seront conçus deux tables (une d'incidence et une de maintien) qui une fois combinées permettront d'obtenir un tarif de la garantie d'intérêt. Enfin ce tarif sera réévalué en vue de coller aux particularités des assurances collectives : la tarification d'un groupe pour lequel ne sont connues que les caractéristiques moyennes. Pour finir, cette nouvelle méthode de tarification sera comparée à la méthode actuelle type BCAC.

# 1 | Le contexte d'étude

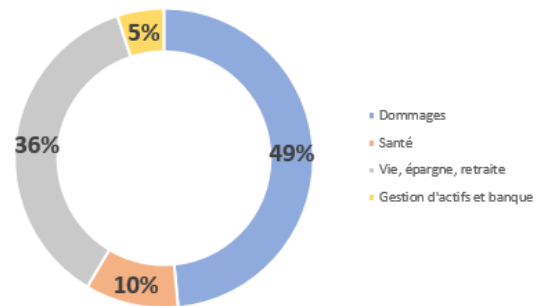
## 1.1 Le groupe AXA

### 1.1.1 Activité et organisation du groupe

Le groupe AXA figure parmi les leaders mondiaux de l'assurance et des services financiers. Le groupe s'engage auprès d'environ 108 millions de clients autour du monde. Ses activités sont soit domestiques, soit internationales (les grands risques internationaux, l'assistance et la réassurance). Le groupe compte environ 160 000 collaborateurs dont 121 000 personnels salariés répartis sur 57 pays.



(a) Répartition géographique des effectifs salariés d'Axa



(b) Répartition du chiffre d'affaire par activité

Le groupe offre son expertise et ses services dans les trois grands domaines d'activité suivant.

- l'assurance dommage : elle regroupe les dommages aux biens (automobile, habitation, ...) ainsi que la responsabilité (civile et professionnelle) ;
- l'assurance vie, épargne, retraite et santé : elle concerne les contrats individuels et collectifs des produits d'épargne et de retraite d'un côté et ceux de prévoyance et santé de l'autre ;
- la gestion d'actifs : elle place et gère les investissements des sociétés d'assurance du groupe et de leur clients, ainsi que ceux de tiers, particuliers ou entreprises.



### 1.1.2 Axa France

AXA France est composée de plusieurs entités comme l'indique le graphique ci-dessous. Le chiffre d'affaire de leurs activités représente près de 26 milliards d'euros.

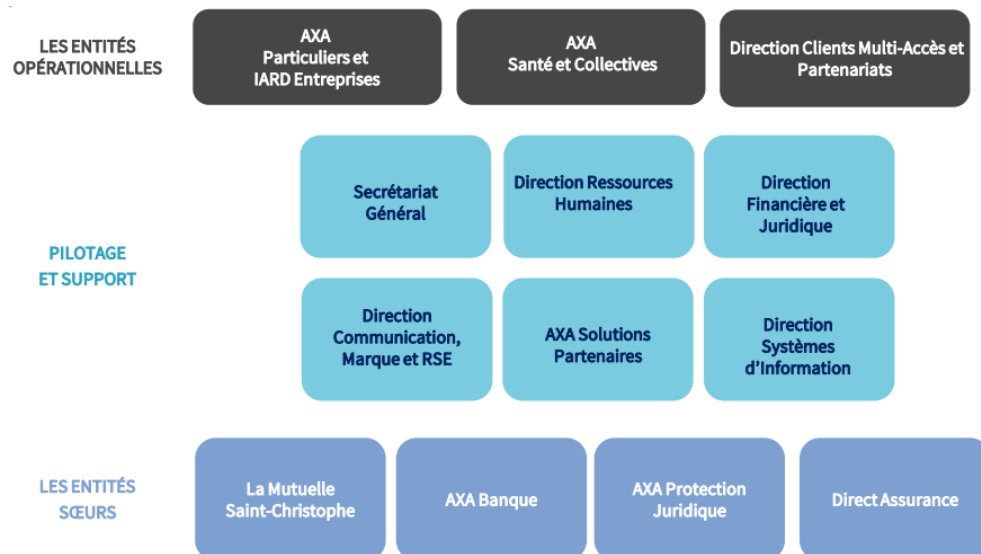


FIGURE 1.2 – La gouvernance d'Axa France

C'est au sein de l'entité AXA Santé Collectives que le sujet de cette étude est traité. Cette dernière, présentée dans le paragraphe suivant, regroupe l'ensemble des activités collectives (Santé, prévoyance et Retraite en France et à l'international) ainsi que la santé individuelle.

### 1.1.3 Axa Santé & Collectives (ASC)

AXA Santé et Collectives, un des leader du marché des assurances collectives, propose une gamme complète de solutions aux entreprises clientes et aux partenaires dans le domaine de la protection sociale complémentaire (prévoyance, dépendance, santé, retraite, épargne salariale, emprunteur, réassurance) en France et à l'international afin de protéger les salariés et leur famille.

Avec ses 9,8 milliards d'euros de chiffre d'affaire, l'entité contribue à 47% du chiffre d'affaire d'Axa France vie et 34% de celui d'Axa France. Ce résultat se partage entre les quatre principales directions qui compose Axa Santé & Collectives.

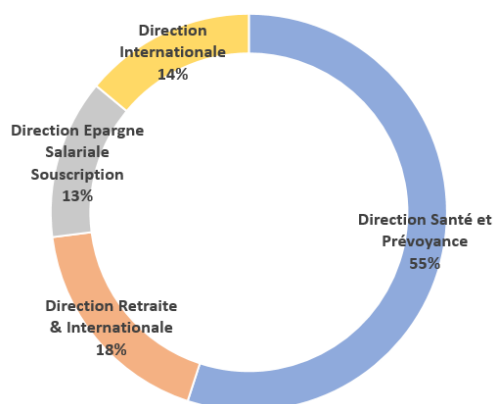


FIGURE 1.3 – Répartition du chiffre d'affaire entre les directions d'ASC

Le poids de la Direction Santé et Prévoyance dans le chiffre d'affaire global montre la part importante qu'occupe les activités de santé et de prévoyance au sein d'Axa Santé & Collectives.

L'étude réalisée ici est effectuée au sein d'une des directions supports aux quatre fonctions précédentes : la *Direction Technique et Innovation (DTI)*.

Les missions de la DTI, décrites ci-dessous, portent sur des dimensions en transversal pour les activités en France et à l'international :

- Normes Techniques : élaboration et contrôle du respect des normes ;
- Support aux Marchés : aide à la souscription, pilotage réglementaire, marketing stratégique ;
- Modèles, rentabilité et big data.

## 1.2 La prévoyance collective

La prévoyance est une activité propre à l'assurance de personnes. L'assurance de personnes couvre des risques liés à la personne humaine. Ce type d'assurance est à distinguer des assurances de dommages, qui ont pour objet la protection du patrimoine. Au sein de l'assurance de personnes, en plus de l'activité prévoyance/santé se retrouve également l'activité épargne/retraite. Les opérations de prévoyance servent donc à couvrir les risques liés à la personne. Ces opérations regroupent l'ensemble des protections contre le décès, les incapacités temporaires, les invalidités permanentes. Cela permet aux ménages de faire face plus facilement à une baisse de revenus liée aux risques d'accidents ou maladie via le versement d'un capital ou d'une rente.

La prévoyance peut être proposée par l'employeur (prévoyance collective), ou être souscrite de façon individuelle (prévoyance individuelle). L'objet de cette étude porte uniquement sur la prévoyance collective. Elle regroupe deux grandes branches d'assurances :

- Les assurances en cas de décès ;
- Les assurances en cas de maladie ou d'accident corporel.

En cas de décès, la prestation consiste au versement d'un capital ou d'une rente au conjoint et aux enfants survivants. Les garanties en cas de maladie ou d'accident corporel permettent le remboursement des soins en complément des régimes obligatoires de protection sociale, et/ou le versement d'indemnités en cas d'incapacité de travail, d'invalidité ou de dépendance. En France, les couvertures de prévoyance se scindent en trois niveaux.

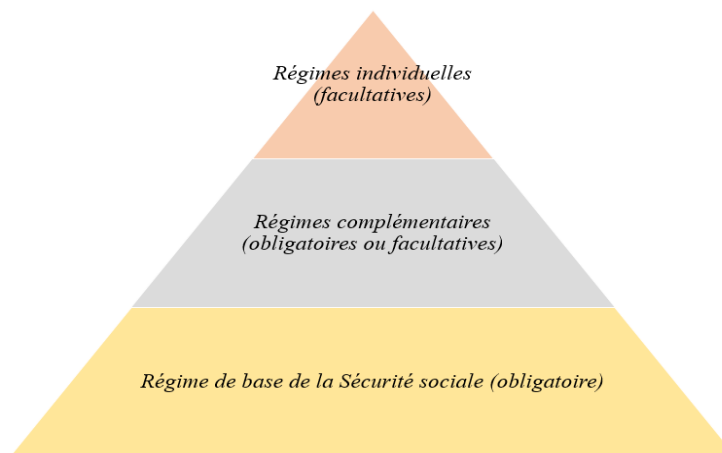


FIGURE 1.4 – Système de la protection sociale en France

### 1.2.1 La sécurité sociale

Une part des couvertures prévoyance est prise en charge par le régime général de la sécurité sociale. Depuis 1945, date de création de la Sécurité Sociale (SS), celle-ci assure la protection des individus contre les conséquences financières liées à la maladie, à l'accident, la maternité, le veuvage.

L'organisation de la sécurité sociale résulte de l'ordonnance de 1967 qui instaure la séparation de la sécurité sociale en 4 branches autonomes. Le code de la sécurité sociale définit ces branches couvrant chacune une catégorie de risques ainsi que les modes de couverture et prestations prévues pour les ayants droit concernés :

- la branche maladie (maladie, maternité, invalidité, décès) ;
- la branche accidents du travail et maladies professionnelles (AT/MP) ;
- la branche vieillesse et veuvage (retraite) ;
- la branche famille (dont handicap, logement ...)

Les graphiques ci-dessous synthétisent la répartition des recettes et des dépenses par branche pour l'ensemble des régimes obligatoires de base.

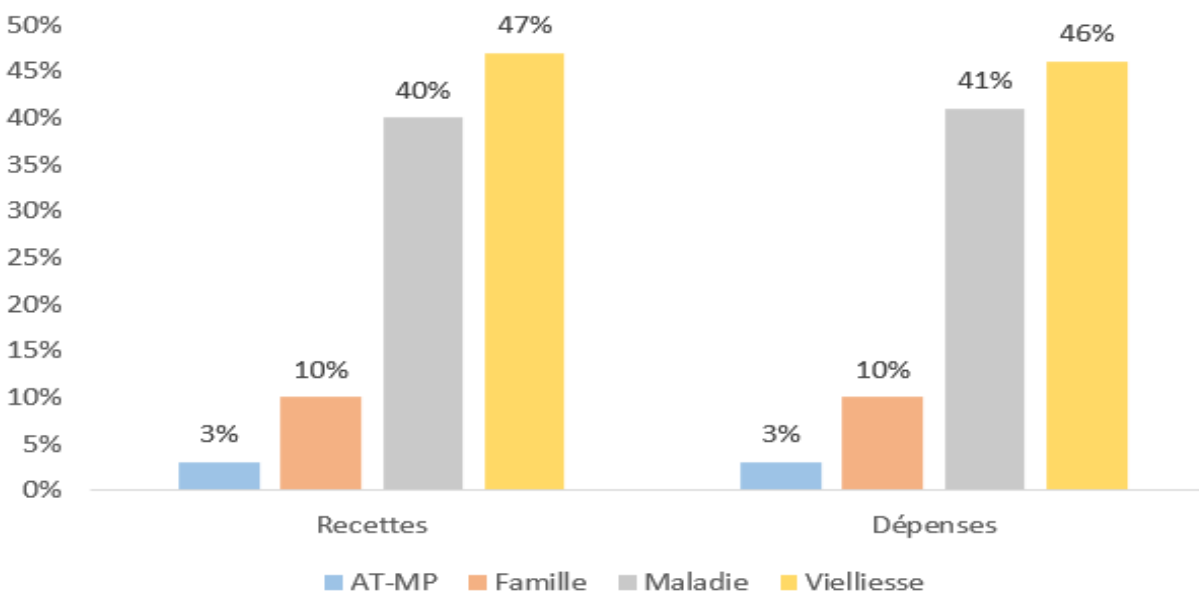


FIGURE 1.5 – Ventilation des dépenses et recettes de la sécurité sociale (2019)

### 1.2.2 Les régimes complémentaires

Comme décrit ci-dessus, la sécurité sociale n'assure pas le remboursement de la totalité des pertes de chacun. L'autre part de la couverture prévoyance est alors assurée par des régimes complémentaires. Ces régimes viennent en complément de la sécurité sociale, soit pour répondre à des obligations légales et conventionnelles des entreprises, soit pour palier l'insuffisance des prestations servies par la Sécurité sociale, notamment pour les assurés aux revenus élevés en arrêt de travail. D'après la loi Evin, les organismes agréés pour proposer de tels produits d'assurance sont :

- Les sociétés d'assurance régies par le Code des Assurances
- Les institutions de prévoyance régies par le Code de la Sécurité Sociale
- Les mutuelles régies par le Code de la mutualité
- Les institutions régies par le Code Rural

Les garanties en prévoyance offertes aux assurés par ces organismes sont définies par la loi n° 89 - 1009 du 31 Décembre 1989. Ces garanties concernent « les opérations ayant pour objet la prévention et la couverture du risque décès, des risques portant atteinte à l'intégrité physique de la personne ou liés à la maternité ou des risques d'incapacité de travail ou d'invalidité ou du risque chômage ». Dans ce sens, les garanties prévoyance communément proposées par les organismes d'assurance sont :

- La garantie Incapacité Temporaire / Invalidité Permanente (IT/IP) : La garantie IT, objet de cette étude, permet à l'assuré de compenser sa perte de revenus en cas d'arrêt de travail dû à une atteinte à son intégrité physique : maladie ou accident. La garantie invalidité permanente permet à l'assuré de compenser la diminution durable de sa capacité de travail et de gain à la suite d'une maladie ou d'un accident. En cas d'arrêt de travail, la garantie IT/IP consiste en un versement d'indemnités journalières à l'assuré ;

- La garantie décès : Cette garantie permet aux proches de l'assuré de faire face aux conséquences de son décès : frais immédiats (frais d'obsèques et de dernière maladie) et compensation de la perte de revenus pour la famille. Elle peut consister en un versement d'un pourcentage du capital en fonction du salaire ou alors en une rente.
- La rente éducation : Il s'agit d'une garantie qui vient souvent en complément d'une garantie décès. En cas de décès de l'assuré, cette garantie prévoit le versement d'une rente à chaque enfant à charge au moment du décès de l'assuré ;
- La rente de conjoint : Elle a pour objet en cas de décès de l'assuré de servir au conjoint survivant une rente viagère ou une rente temporaire dans certaines situations.

### 1.2.3 Les régimes individuels

Les contrats de prévoyance individuelle sont souscrits individuellement et protègent l'assuré, son conjoint et ses enfants. L'objectif est de compléter les garanties proposées par le contrat de prévoyance collective de l'entreprise en vue de minimiser les conséquences financières que peuvent provoquer un accident, une maladie ou un décès.

## 1.3 Le risque incapacité temporaire

### 1.3.1 L'indemnisation de l'incapacité

Le risque incapacité temporaire concerne toutes les interruptions de travail ayant pour cause la maladie professionnelle ou non, ainsi que les accidents du travail.

Selon l'article L321-1 du Code de la Sécurité Sociale, « *L'assurance maladie comporte (...) l'octroi d'indemnités journalières à l'assuré qui se trouve dans l'incapacité physique constatée par le médecin traitant, (...) de continuer ou de reprendre le travail* ».

L'article L441-1 du Code de la Sécurité Sociale définit l'accident du travail comme un accident survenu par le fait ou à l'occasion du travail ou encore un accident survenu à un travailleur pendant le trajet d'aller et de retour qui sépare le lieu de travail du lieu de résidence ou de déjeuner. La maladie est d'origine professionnelle lorsqu'elle figure dans un tableau de maladies professionnelles et est contractée dans les conditions mentionnées dans ce tableau.

#### a) Les prestations de la sécurité sociale

L'ouverture des droits varie en fonction de la durée et du contexte de l'arrêt de travail. L'indemnité journalière est due au titre de chaque jour ouvrable ou non à la suite de l'expiration d'un délai de carence de 3 jours. A noter que le délai de carence ne s'applique pas lorsque la reprise d'activité entre deux prescriptions d'arrêt de travail ne dépasse pas 48 heures ou encore lorsque l'assuré est en affectation de longue durée.

Il existe trois causes possibles de sorties de l'état d'incapacité : le rétablissement, le décès et le passage en invalidité. Le graphique suivant montre les liens envisageables entre les différents états. Les liens en pointillés ne rentrent pas dans le cadre de cette étude.

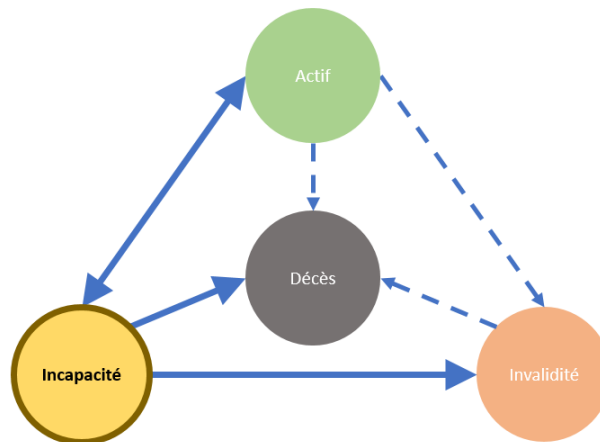


FIGURE 1.6 – Les différents états de l'arrêt de travail

Le niveau de l'indemnité journalière versée par la Sécurité sociale s'élève à 50% du salaire journalier de base (porté à 66,66% à compter du 31ème jour d'arrêt lorsque l'assuré a 3 enfants à charge). Dans le cas d'un arrêt d'origine professionnelle, aucun délai de carence n'est applicable et le montant de l'indemnisation est de 60% du salaire journalier de base jusqu'au 28ème jour puis 80% à partir du 29ème jour. Le salaire journalier de base est calculé sur la moyenne des salaires bruts des trois derniers mois travaillés précédant l'arrêt, plafonné à  $1,8 \times SMIC$ . Ce salaire de base ne prend en compte que le dernier salaire dans le cas de l'arrêt d'origine professionnelle. L'indemnité peut également être revalorisée au-delà de trois mois d'arrêt en cas d'augmentation générale des salaires dans l'entreprise.

## b) Les prestations complémentaires

La garantie incapacité temporaire de travail permet au salarié en arrêt de travail de percevoir des indemnités journalières afin de compenser sa perte de salaire. Ces indemnités viennent compléter celles de la Sécurité sociale et le complément de revenus versé par l'employeur (obligation mise en place par la loi de mensualisation de 1987). Il faut distinguer les prestations versées au titre d'un contrat collectif obligatoire de celle versées sur une assurance individuelle facultative. En effet, le caractère obligatoire de l'assurance collective et la volonté législative de protection de l'assuré en matière de prévoyance imposent certaines règles à l'assureur. La loi Evin qui sera détaillée ultérieurement précise les modalités à respecter pour la couverture (obligatoire ou facultative) de risques sociaux.

Le montant des indemnités journalières versées par l'organisme assureur et la date de début de versement des indemnités dépendent du contrat mis en place dans l'entreprise. Les quatre principaux paramètres définissant une garantie incapacité sont la franchise (type et durée), la durée de prise en charge, le montant en pourcentage du salaire et l'assiette de calcul.

### 1.3.2 Les types de franchises

En assurance de biens, une franchise est une somme restant à la charge de l'assuré autrement dit non indemnisée par l'assureur. Dans le cas de l'arrêt de travail, la franchise représente un délai de carence précédant la couverture de l'incapacité temporaire. Elle s'exprime en nombre de jours et est décomptée à partir du premier jour d'arrêt hors congés de maternité, de paternité ou d'adoption qui bénéficient de garanties particulières. Un contrat sans franchise, permet l'indemnisation de la période d'absence dès le premier jour d'arrêt.

Les franchises peuvent être très variables d'un contrat à un autre et ont différents rôles. Suivant les cas, elles permettent à l'assureur :

- d'éviter l'indemnisation des arrêts très courts, souvent frauduleux et plus chers en frais de gestion qu'en indemnisation,
- de diminuer l'aléa moral, l'assuré continuant à subir une perte en cas de sinistres.

A noter que plus la franchise est courte, plus la prime à payer est élevée. Il existe différents types de franchises. Cette étude met en avant la franchise dite discontinue.

#### a) La franchise discontinue

Dans le cas d'une franchise discontinue, tous les arrêts de l'assuré sont comptabilisés et leurs durées additionnées. L'indemnisation intervient une fois que le nombre de jours d'arrêt de travail cumulés au cours d'une année glissante atteint la limite définie contractuellement. L'année glissante se définit comme les 365 jours précédant le début de la dernière incapacité. Au delà de cette limite, toute nouvelle incapacité est indemnisée au premier jour. Ce principe est applicable tant que le cumul de jours d'arrêt sur les 365 derniers jours dépasse la durée de la franchise. Il est également important de noter deux particularités de la franchise discontinue :

- les jours indemnisés ne sont pas comptabilisés dans le cumul des jours d'absence ;
- la couverture intervient sur les arrêts de travail débutant durant la période indemnisable et non pas uniquement sur les jours inclus dans cette période. Ainsi un arrêt débutant dans une période indemnisable et se terminant au-delà de cette période sera entièrement couvert par la garantie en franchise discontinue.

Le schéma ci-dessous explique le processus de mise en place de la couverture incapacité temporaire sous de la franchise discontinue. Dans cet exemple la franchise discontinue utilisée est de durée 90 jours.

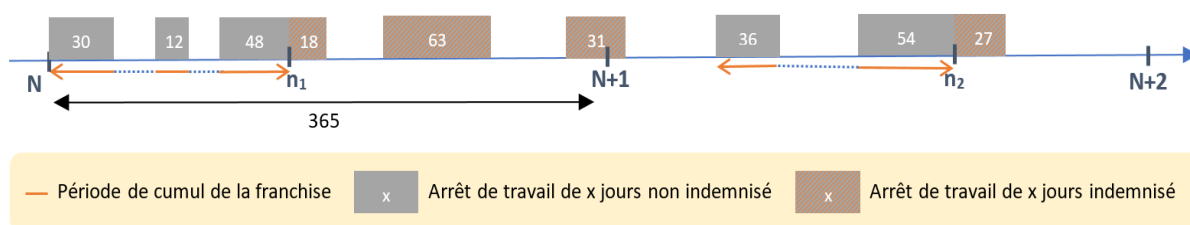


FIGURE 1.7 – Schématisation de l'indemnisation en franchise discontinue - Cas n° 1

Ici, l'assuré étudié présente 7 périodes d'absence entre les dates  $N$  et  $N + 2$ . Il est supposé qu'aucun arrêt de travail n'est survenu entre le début du contrat de l'assuré et la date  $N$ . La franchise de 90 jours est atteinte la première fois à la date  $n_1$ . La période indemnisable est donc le laps de temps entre  $n_1$  et  $N + 1 = N_1 + 365$ . L'ensemble des arrêts débutant entre  $n_1$  et  $N_1 + 1$  seront donc indemnisés. Il s'agit dans ce cas des 18 derniers jours du 3ème arrêt, du 4ème arrêt ainsi que le 5ème arrêt. Les arrêts débutant au delà de la date  $N + 1$  sont de nouveau soumis au cumul de la franchise.

**b) Les autres franchises**

La **franchise continue** est définie comme la différence de jours entre la date de première période couverte et la date de survenance de l'arrêt. L'indemnisation intervient dès lors que le nombre de jours d'un même arrêt dépasse la durée limite fixée contractuellement. A la différence de la franchise discontinue, la franchise continue est appliquée pour chaque nouvelle incapacité à l'issue d'une période continue d'arrêt de travail.

$$Franchise = \text{date de première période couverte} - \text{date de survenance de l'arrêt}$$

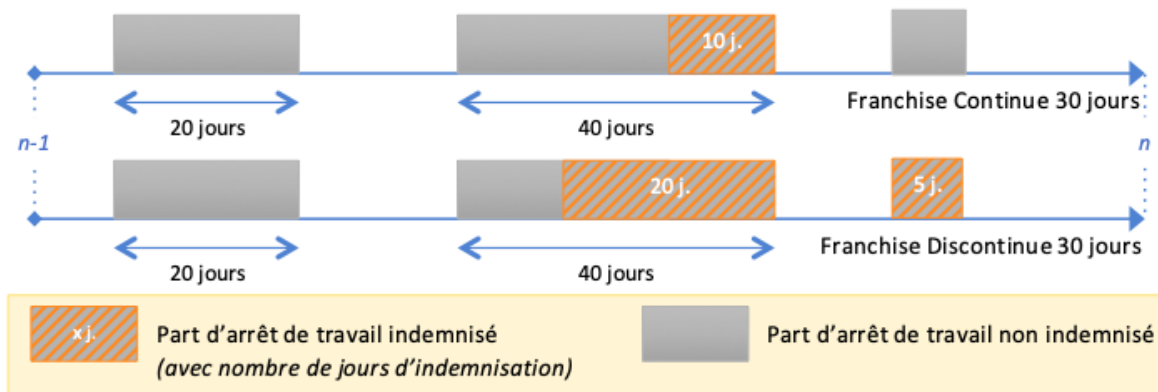


FIGURE 1.8 – Comparaison de l'indemnisation de l'arrêt de travail suivant la franchise continue ou discontinue (à durée égale)

Cet exemple présente une garantie sous franchise continue et une sous franchise discontinue toutes deux de 30 jours. Dans le cas continu, le cumul des jours d'absence se renouvelle à chaque arrêt de travail. Ainsi, seuls les 10 derniers jours de l'arrêt seront indemnisés contre 20 jours dans le cas discontinu.

Il est possible qu'une clause de "rechute" existe au contrat. Dans ce cas, si l'assuré tombe en incapacité avant un certain délai et pour les mêmes causes, aucune franchise ne lui sera appliquée.



La loi de mensualisation de 1978 oblige tout employeur, sous certaines conditions et selon la Convention Collective Nationale (CCN) dont relève l'entreprise, à maintenir à un certain niveau le salaire de ses salariés en arrêt de travail. L'Employeur peut financer ce complément d'indemnisation sur sa propre trésorerie, confier à un Organisme Assureur la gestion de cet engagement en souscrivant un contrat de mensualisation ou alors inclure cette prestation dans les garanties de son contrat de Prévoyance. En choisissant cette dernière option, l'assureur est tenu de verser des indemnités afin de maintenir voire dépasser le niveau de salaire prévu par la CCN. Il s'agit de **la franchise dite en relais de la Convention Collective Nationale** (relais CCN).

Dans le cas d'une franchise continue, lorsque la durée d'arrêt de travail est supérieure à la franchise, il peut être prévu que l'assureur verse rétroactivement les indemnités journalières. La période de franchise est alors indemnisée totalement ou partiellement. La rétroactivité peut s'appliquer quelle que soit la cause de l'arrêt ou uniquement en cas d'arrêt consécutif à un accident ou une hospitalisation. Il s'agit alors d'une **franchise rétroactive**.

De la même manière, le contrat peut prévoir que la franchise soit d'une durée plus courte dans certains cas : arrêt de travail consécutif à un accident et/ou ayant donné lieu à une hospitalisation. Il s'agit d'une **franchise réduite**.

### 1.3.3 Le cadre législatif

Le risque arrêt de travail peut s'avérer très lourd à porter en terme de prestation. En effet, l'assureur doit pouvoir honorer ses engagements envers l'assuré en cas d'incapacité ou d'invalidité et maintenir ses prestations même en cas de rupture du contrat d'assurance (loi de mensualisation de 1978). De plus, il doit prendre en considération la possibilité qu'un assuré en incapacité temporaire puisse définitivement passer en invalidité sous indication de la Sécurité sociale. L'assurance collective étant à caractère obligatoire, dans un souci de protection de l'assuré en matière de prévoyance et de limitation du risque de faillite, la législation impose certaines règles à l'assureur concernant les opérations de prévoyance complémentaire.

#### a) Loi Evin du 31 Decembre 1989

Avant l'année 1990, la seule obligation des organismes d'assureurs en matière de provisions imposait uniquement que les sinistres de l'année soient couverts par les primes de cette même année. Pour provisionner l'incapacité, les organismes pouvaient, sans obligations légales, se baser sur les méthodes proposées par le guide de l'assurance collective du BCAC et réalisées à partir de travaux de M. Wetzl datant des années 1960. Cependant ces méthodes présentaient des limites.

Dans le but de protéger et sécuriser les opérations de prévoyance, la loi n° 89-1009 du 31 décembre 1989 dite Evin est le premier ensemble de règles portant sur la couverture (obligatoire ou facultative) de risques sociaux, applicables à toutes les familles d'assureurs autorisées à pratiquer la prévoyance collective complémentaire (mutuelles, instituts de prévoyance et compagnies d'assurance). La liste suivante décrit ses principales dispositions :

- Interdiction de sélectionner les risques pour les contrats de groupe ;
- Prise en charge des suites de pathologies antérieures ;
- Obligation des assureurs à proposer la poursuite de la couverture santé aux anciens salariés (retraités, licenciés, ayants droit d'un assuré décédé . . .) pour une augmentation tarifaire limitée (Art.7-1) ;
- Maintien du paiement des rentes au niveau atteint ou plus en cas de résiliation (Art.7) ;
- Interdiction à l'employeur d'imposer unilatéralement de cotiser à un régime de prévoyance ;
- Remise d'une notice résumant les garanties ;
- Consultation du Comité d'entreprise ;
- Remise d'un rapport sur les résultats du contrat (Art.15)

La loi du 08 Août 1994 renforce les mesures de la loi Evin en imposant à l'entreprise souscriptrice la revalorisation des prestations en cours de service en cas de résiliation du contrat et changement d'organisme assureur. Pour ce faire, deux solutions sont possibles : soit ces prestations sont prises en charges par l'ancien assureur, soit elles sont transférées vers le nouvel assureur.

#### **b) Arrêté du 28 Mars 1996**

L'arrêté du 28 mars 1996 fixe les règles et réactualise les méthodes jusqu'alors utilisées concernant le provisionnement de prestations d'incapacité et d'invalidité. Contrairement à la loi Evin, ce texte s'applique aussi bien aux opérations individuelles que collectives. Il permet la mise en oeuvre des articles A331-22 du Code des Assurances, A212-9 du Code de la Mutualité et 931-10-9 du Code de la Sécurité sociale en homologuant :

- Des lois de maintien en incapacité/invalidité ;
- Des lois de passage d'incapacité en invalidité ;
- Des coefficients de provisionnement ;
- Des taux d'actualisation à appliquer.

Établis pour la première fois en 1993, ces lois et barèmes étaient basés sur des portefeuilles d'assurances collectives des principales compagnies françaises d'assurance : AGF, AXA, GAN et UAP. Dans le cadre de la réforme des retraites de 2010, l'arrêté du 24 décembre 2010 a fixé les tables de maintien que les Organismes Assureurs devront utiliser pour le provisionnement à compter de 2011. A noter également que de nouvelles tables ont été construites en 2013 par le BCAC, mais ne sont pas homologuées.

Par ailleurs, les tables réglementaires relatives aux prestations d'arrêt de travail ne doivent être obligatoirement utilisées que pour le provisionnement. La tarification du risque Arrêt de travail n'est pas encadrée réglementairement.

## 2 | Les données

### 2.1 La déclaration sociale nominative (DSN)

#### 2.1.1 Présentation du portefeuille d'étude

La Déclaration Sociale Nominative (DSN) est un fichier mensuellement établi par l'employeur et destiné à communiquer les informations nécessaires à la gestion de la protection sociale des salariés aux organismes et administrations concernées (CPAM, Urssaf, AGIRC ARRCO, Organismes complémentaires, Pôle emploi, Centre des impôts, Caisses régimes spéciaux ...). Il s'agit d'un échange de données informatisées qui a pour objectif de remplacer l'ensemble des déclarations périodiques ou événementielles et les diverses formalités administratives adressées jusqu'alors par les employeurs. Le but est de réduire les risques d'erreur et la charge de travail déclarative.

La DSN repose sur la transmission unique, mensuelle et dématérialisée des données issues de la paie et sur des signalements d'événements. Elle permet donc de refléter la situation d'un salarié au moment où la paie a été réalisée ainsi que les événements survenus (maladie, maternité, changement d'un élément du contrat de travail, fin de contrat de travail...) dans le mois ayant eu un impact sur la paie. La DSN substitue depuis 2017 de nombreuses déclarations dont la déclaration annuelle des données sociales (DADS). Elle est même devenue obligatoire pour la plupart des employeurs (hors EPA et employeurs FPT, FPH et FPE) depuis janvier 2017.



Il est intéressant d'utiliser ces déclarations comme base de données car les arrêts de travail sont indiqués au premier jours pour l'ensemble des salariés ayant souscrit à un contrat de prévoyance auprès de l'assureur. Ainsi, il est possible d'étudier l'ensemble des absences au sein du portefeuille même lorsque la franchise n'a pas été atteinte et donc que l'assuré n'a pas perçu d'indemnisation.

### 2.1.2 Fiabilité des données

La DSN ne peut pas être directement exploitée. Il faut avant cela définir un périmètre d'observation des contrats à étudier puis traiter les données qui semblent aberrantes. En effet, la base étant complétée grâce aux informations transmises par les entreprises, il n'est pas exclu que certaines entrées soient involontairement erronées. Cela est d'autant plus vrai que la déclaration n'est obligatoire pour la plupart des employeurs que depuis 2017. Au moins pour la première année, beaucoup d'entreprises manquent d'expérience sur la façon de compléter la déclaration et cela croît les potentielles erreurs.

Afin de pouvoir exploiter le plus justement les données, un travail de nettoyage et de fiabilisation est donc nécessaire. Il a fallu supprimer les déclarations incomplètes ainsi que celles présentant des anomalies.

Pour détecter les éléments à supprimer, une étude de la distribution du ratio d'évolution par taille d'entreprise a été faite. Il s'agit d'une étude sur les taux de déclaration des arrêts de travail de chaque entreprise. Les taux de déclaration de l'arrêt de travail sont calculés trimestre par trimestre comme suit :

$$\text{Taux de déclaration d'AT}_{\text{Trimestrei}} = \frac{\text{Nombre d'AT déclarés}_{\text{Trimestrei}}}{\text{Effectif de salariés exposés}_{\text{Trimestrei}}}$$

A partir de ce taux, le ratio d'évolution du taux de déclaration d'arrêt de travail entre deux années  $n$  et  $n + 1$  peut être construit.

$$\text{Ratio d'Evolution}_{n,n+1} = \frac{\text{Taux de déclaration d'AT}_{\text{Trimestrei},n+1}}{\text{Taux de déclaration d'AT}_{\text{Trimestrei},n}}$$

Le ratio d'évolution est calculé pour chaque entité en vue de définir la loi empirique suivie par ce ratio. Les quantiles empiriques construits à partir de la fonction de répartition des ratios d'évolutions permettent alors de définir les seuils  $i$  d'évolution du nombre d'arrêts d'une année à l'autre selon la taille de l'établissement.

L'hypothèse considérée est que le nombre d'arrêts de travail d'une entreprise en une année suit une loi binomiale de paramètre  $(n, p_n)$  avec  $n$  le nombre moyen de salariés dans l'année et  $p_n$  la probabilité d'avoir un arrêt de travail par salarié. La probabilité  $p_n$  est estimée en fonction de la taille  $n$  de l'entreprise à l'aide de l'ensemble des résultats observés les années passées.

Pour  $X$  le nombre d'arrêt de travail d'un établissement,  $X \sim \mathcal{B}(n, p)$

Dans un premier temps, pour chaque année déclarée par une entité, le nombre d'arrêt observés sur l'année ( $X_{\text{observe}}$ ) est comparé avec les résultats attendus ( $X_{\text{attendus}}$ ). En considérant un intervalle de confiance à 95% du nombre d'arrêts attendus, il faut que :

$$\mathbb{P}(X_{\text{observe}} \in [X_{\text{attendus}} \pm i\% \times X_{\text{attendus}}]) \geq 95\%$$

avec  $i$  l'évolution du nombre d'arrêts d'une année à l'autre selon la taille de l'établissement estimée précédemment.

Le taux d'évolution  $i$  est estimé en se basant sur les entités (groupées par taille) ayant des DSN complètes sur deux années consécutives. Pour l'estimation du nombre de sinistres attendus ( $X_{attendus}$ ), ce sont les déclarations complètes sur une année qui sont analysées.

Dans un second temps, au sein d'une même entité, lorsque les déclarations le permettent, l'évolution du nombre d'arrêts de travail par salarié est comparée d'une année à l'autre. Là aussi cette évolution doit s'inscrire dans l'intervalle de confiance de l'évolution par salarié modélisée à partir des déclarations complètes.

## 2.2 La base de données d'étude

### 2.2.1 Le risque étudié

Afin de pouvoir étudier le risque incapacité temporaire pour les garanties dont le type de franchise est discontinue, analyser la fréquence d'occurrence et la durée d'un arrêt de travail ne suffit pas. En effet, la franchise n'est pas renouvelée à chaque nouvel arrêt, mais lorsque sur une année, l'ensemble des jours d'absence est suffisant pour déclencher la garantie. Pour chaque assuré présent sur la base DSN, l'événement d'intérêt est donc le temps total d'absence en jours de cet individu sur un an glissant.

L'analyse de cet événement nécessite dans un premier temps de déterminer mensuellement le nombre de jours en arrêt de travail par salarié. Ensuite la durée sur 12 mois glissant est calculée suivant deux méthodes distinctes selon l'ancienneté de l'assuré dans la base DSN.

- les assurés observés sur une période inférieure à 12 mois consécutifs voient leur durée d'absence annuelle soit reconstituée à partir des informations disponibles, soit limitée et correspondante à leur temps de présence. Ces observations sont dites incomplètes et sont étiquetées comme tel via un indicateur.
- pour les assurés ayant 12 mois ou plus de présence consécutifs dans la base, la durée recherchée est calculée en moyennant l'ensemble des possibilités de recenser les jours d'absence sur 12 mois successifs.

Par exemple, pour une période de présence de Janvier 2018 à Février 2019, il s'agit de moyenner le nombre de jours d'arrêts sur les intervalles de Janvier 2018 à Décembre 2018, de Février 2018 à Janvier 2019 et de Mars 2018 à Février 2019.

La base finale comprendra les caractéristiques clés du salarié (sexe et collègue), son âge exact au moment du déclenchement de la garantie, l'entreprise et le contrat auquel il est associé, son exposition (temps de présence dans la base de données) hors arrêt de travail sur la période étudiée, son nombre de jours d'absence annuel estimé et un indicateur de censure lorsque l'observation est incomplète.

## 2.2.2 Le périmètre d'observation

### a) Les âges retenus

La garantie arrêt de travail concerne la part de la population en activité. En vue de couvrir au mieux l'ensemble des salariés, les salariés de moins de 20 ans et ceux de plus de 65 ans sont représentés en très faible proportion dans le portefeuille. Ces individus sont alors respectivement regroupés aux tranches d'âge 20 ans et 65 ans en prenant soins de nettoyer toutes les données semblant aberrantes ou extrêmes (comme les salariés de moins de 16 ans ou ceux de plus de 75 ans). Le périmètre d'étude est ainsi réduit aux salariés âgés entre 20 et 65 ans.

### b) Les catégories socio-professionnelles d'étude

Sept catégories socio-professionnelles sont recensées dans la déclaration sociale nominative. Dans un premier temps ces catégories sont réparties suivant quatre statuts - Cadres, Non cadres, Etam (Employés, Techniciens, Agent de Maîtrise) et Autres - comme suit :

CSP	Part	Statut
Agriculteurs	0,04%	Autres
Artisans, commerçants et chefs d'entreprise	0,56%	Autres
Cadres, professions intellectuelles supérieures	32,72%	Cadres
Employés	22,24%	ETAM
Ouvriers	17,73%	Non Cadres
Professions intermédiaires	23,94%	ETAM
Autres	2,77%	Autres

Bien que ne figurant pas dans la catégorie des cadres, le statut ETAM a la spécificité, en fonction des conventions collectives, soit de bénéficier des garanties cadres, soit de favoriser des comportements similaires à celui des cadres. Cela est dû notamment au salaire proposé aux individus de cette catégorie ou encore à une pénibilité du travail proche de celle des cadres. Dans les données, une rapide étude des salaires et des régimes de prévoyance montre que environ 29,86% des professions intermédiaires et employés peuvent être associés à des cadres. Les autres individus de ces catégories sont assimilés à des non cadres. Les statuts "Autres" représentant moins de 2,5% du portefeuille sont par défaut associés aux non cadres sans réel impact sur la suite de l'étude.

Ainsi la segmentation retenue considère deux catégories socio-professionnelles :

- les cadres - 39,37% du portefeuille,
- les non cadres - 60,62% du portefeuille,

### c) Les motifs d'arrêts de travail

Dans la Déclaration Sociale Nominative, les motifs des arrêts de travail sont regroupés suivant une dizaine de catégories. A partir de ces catégories il est possible de distinguer 3 principales causes d'absence :

- les accidents de trajet et maladies professionnelles - représentent 6,3% du portefeuille,
- la maladie - représente 87,2% du portefeuille,
- la parentalité (maternité, paternité, adoption...) - représente 5,6% du portefeuille.

Il existe des garanties spécifiques couvrant les différents motifs de la parentalité. L'étude de l'incapacité temporaire concerne donc uniquement les deux autres causes d'arrêt de travail.

## 2.3 Statistiques descriptives

### 2.3.1 Analyse de la population assurée

Dans un premier lieu, la figure 2.1 montre la pyramide des âges des assurés du portefeuille de l'année 2019. La proportion des hommes (en bleu) et des femmes (en orange) est comparée à la celle de la population active nationale en 2019 selon l'INSEE (représenté par les barres). La population active regroupe l'ensemble des personnes exerçant ou cherchant à exercer une activité professionnelle rémunérée. En 2019, 8,4% des personnes actives sont au chômage. La part des demandeurs d'emplois ne biaise que très légèrement la représentation de la population en activité en France.

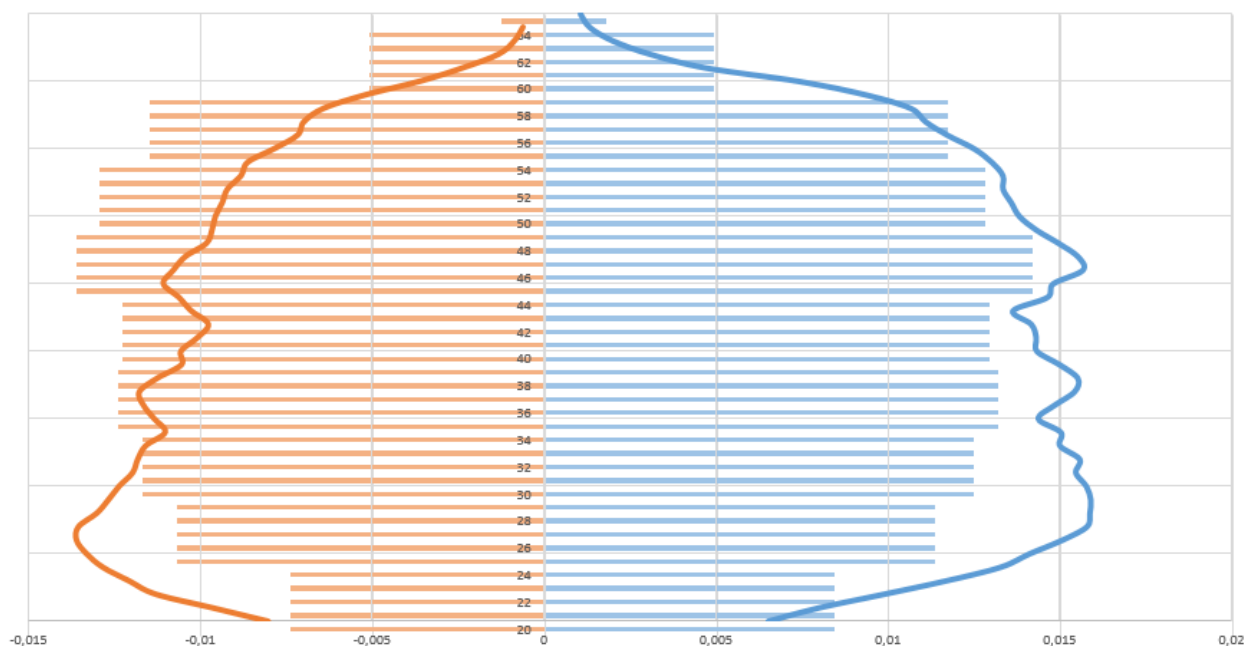


FIGURE 2.1 – Pyramide des âges de la population active nationale (barres) et de la population assurée (courbe) en 2019

Sans surprise les âges très jeunes et les grands âges sont moins représentés sur le marché du travail. Aussi, tout comme dans les statistiques nationales, les hommes actifs sont plus présents que les femmes. La part d'hommes assurés par AXA est en moyenne plus importante que la part nationale. Cela implique donc une proportion de femmes assurées plus faible que la moyenne française ; cette tendance se voit notamment chez les moins jeunes. Le portefeuille d'AXA présente également une population plus jeune que la démographie française.

Les garanties pour l'incapacité temporaire possèdent souvent des singularités dépendant du fait que le salarié soit un cadre ou non. C'est pourquoi il est important de distinguer ces deux catégories socio-professionnelles. Cependant, comme certains non cadres (ETAM) présentent un comportement de cadres voire bénéficient des garanties des cadres, ces derniers sont volontairement associés à des cadres en vue de mieux évaluer la part de contrats ou comportements cadres dont dispose AXA.

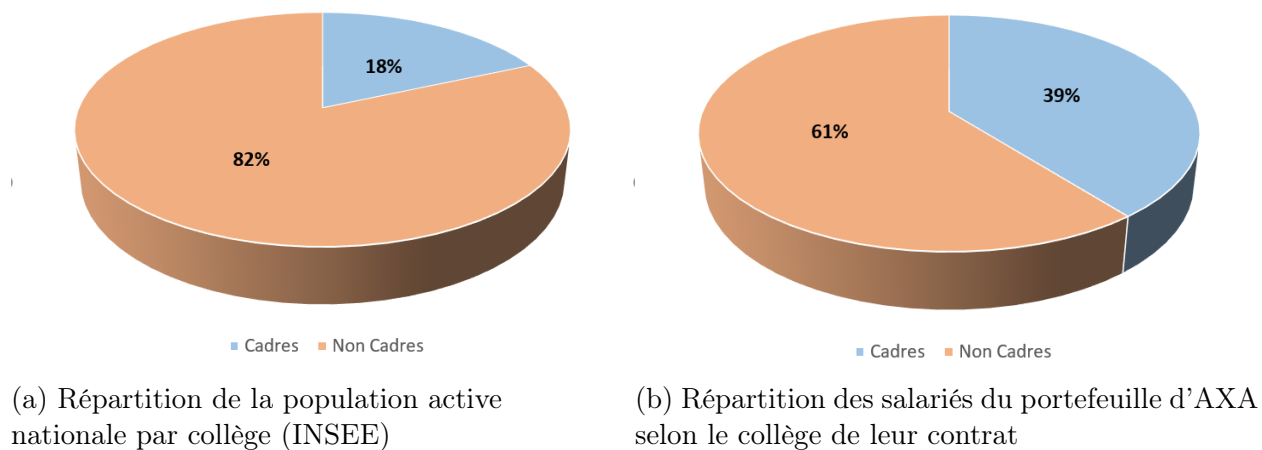


FIGURE 2.2 – Population nationale et étudiée selon leur CSP

Alors que la population nationale compte 18% de cadres parmi ses actifs, l'assureur compte 39% de cadres ou de non cadres considérés comme tel.

La figure 2.3 montre alors comment sont répartis les femmes et les hommes dans ces deux catégories. Alors que chez les non cadres la proportion hommes femmes est plutôt équilibrée (28% de femmes et 32 % d'hommes), chez les non cadres, le sexe masculin reste dominant avec une représentation légèrement inférieure à deux tiers des effectifs des cadres. La part totale de femme est estimée à 43% de l'effectif global. L'INSEE recense en 2019 environ 48% de femmes dans la population active. Cela s'accorde bien aux constatations faites sur la pyramide des âges Figure 2.1 précédente, qui met en évidence une population féminine moins importante que la moyenne nationale.



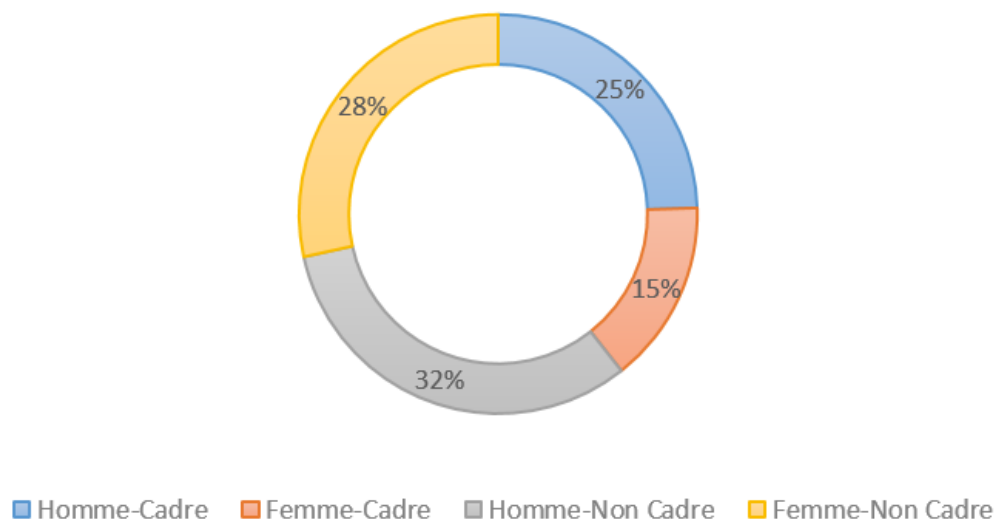


FIGURE 2.3 – Répartitions des salariés suivant leur sexe et collègue

Enfin, le graphique empilé ci-dessous permet d’apprécier la répartition des salariés cadres et non cadres selon leur âge.

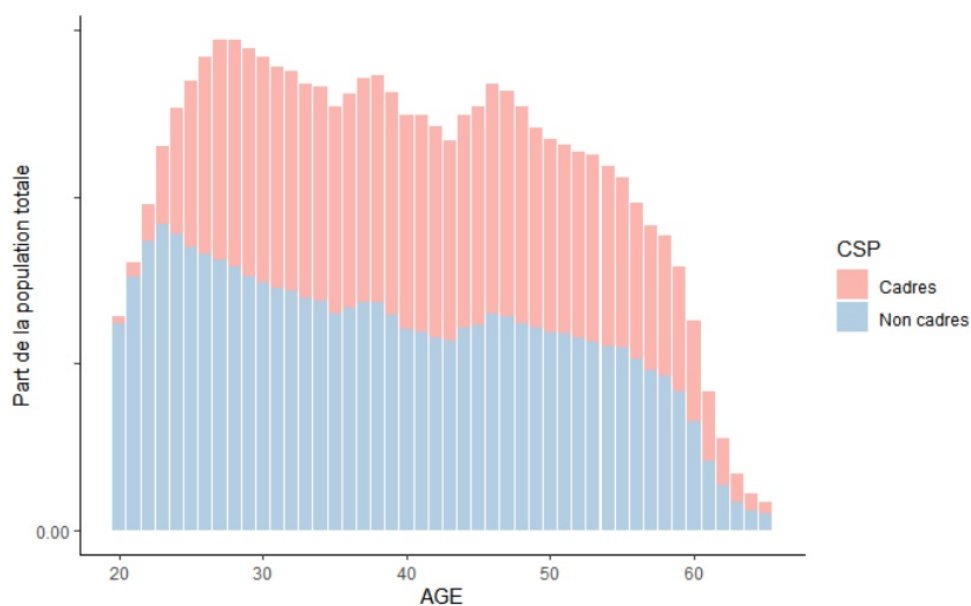


FIGURE 2.4 – Répartition des salariés cadres et non cadres par âge

Aux très jeunes âges, il n’est pas surprenant de constater une très faible représentation des cadres. En effet, ce statut s’obtient dans la plupart des cas soit à la suite d’études longues, soit plus tard dans la carrière comme une évolution professionnelle. Cette domination des non cadres s’équilibre après 25/30 ans où la part de cadre et de non cadre est plus ou moins la même.

Aussi quelque soit la CSP, les effectifs aux grands âges sont très faibles. A ces âges la probabilité de quitter le portefeuille est plus grande notamment suite à un départ à la retraite.

### 2.3.2 Analyse de la sinistralité

Pour tarifier l'incapacité temporaire, deux grands facteurs entrent en jeu. D'abord il faut analyser le nombre d'arrêts de travail survenus sur une année, et ensuite observer la durée de chacun d'entre eux. A partir de ces deux études, il est possible de modéliser le nombre de jours total d'absence des salariés sur une année.

Les premières statistiques descriptives sur la sinistralité décrivent donc le nombre d'arrêts de travail constatés sur l'année 2019 selon différents points de vue. Environ 30% des salariés du portefeuille d'étude sont concernés par l'arrêt de travail. Comme l'illustre la figure ci-dessous, la majorité d'entre eux (plus de 60%) ont un unique sinistre par année. Ensuite la distribution du nombre de sinistres par assurés est décroissante. 20% en ont deux, 9% en ont trois et moins de 8% ont en quatre ou plus.

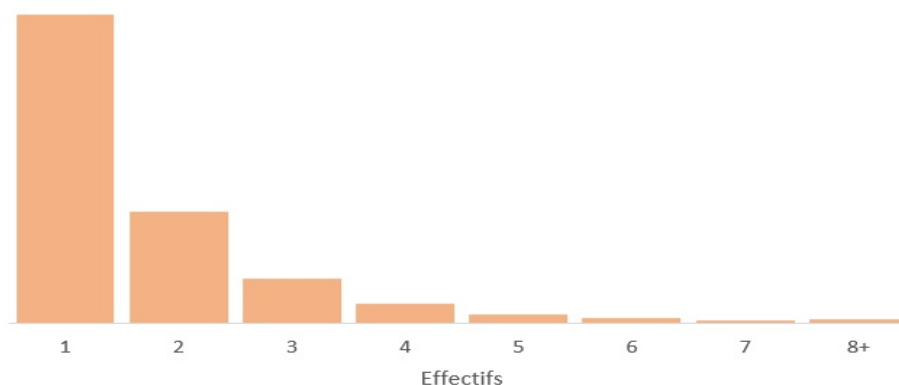


FIGURE 2.5 – Nombre de salariés suivant leur nombre d'arrêts de travail constatés sur une année (2019)

Bien que ressemblante, l'adéquation de cette distribution par une loi de Poisson semble cependant difficile car la modalité 0, représentant les salariés n'ayant eu aucune absence dans l'année, a un effectif important.

Cette même distribution du nombre d'incapacités temporaires par assuré est également présentée ci-dessous en segmentant les salariés selon leur sexe.

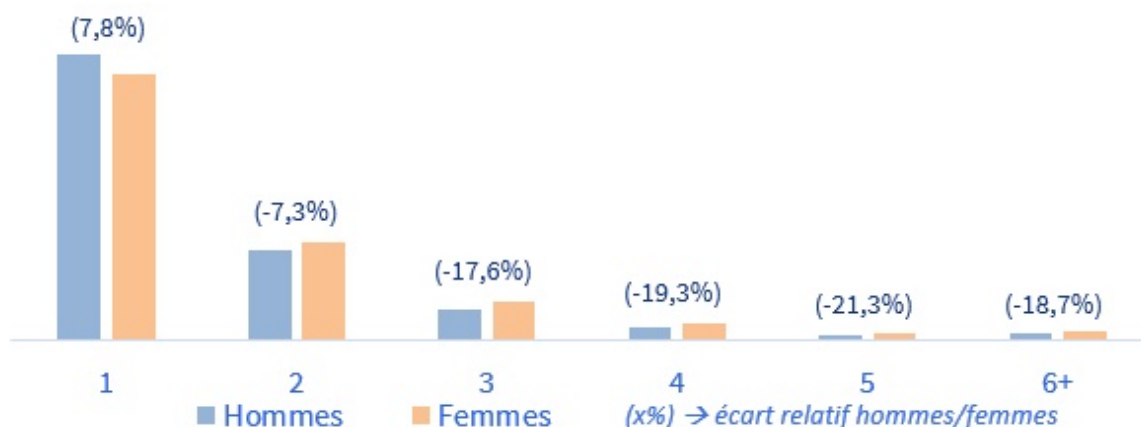
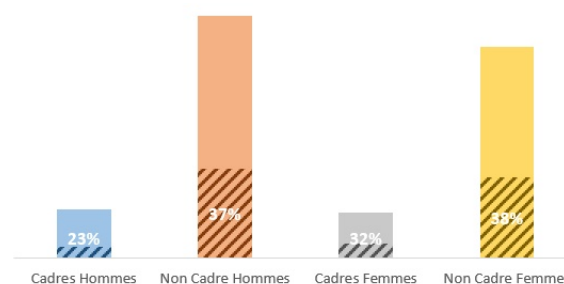
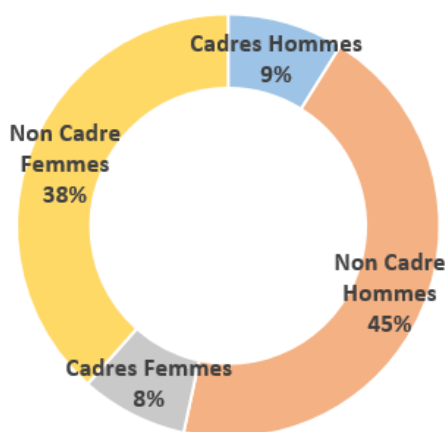


FIGURE 2.6 – Répartition de salariés suivant leur nombre d’arrêts de travail constatés sur une année

Cette représentation met en exergue le fait que les femmes ont plus tendance que les hommes à prendre des arrêts multiples. Bien que l’arrêt unique reste majoritaire pour les deux genres, il est plus fréquent d’observer plusieurs arrêts par an chez les femmes. Par exemple, en regardant la part de la population ayant exactement cinq absences, les femmes sont représentés à plus de 20% que les hommes. Ce phénomène est important à cibler car il permettra l’interprétation de résultats futurs.

Ensuite, les figures ci-dessous analysent quels sont les poids de chacune des sous-populations d’étude dans l’ensemble des arrêts de travail, mais également quelle est la part de salarié dans chacun de ces sous-groupes ayant au moins eu un arrêt de travail dans l’année. Par exemple 37% des hommes non cadres se répartissent à eux seuls 45% de l’ensemble des arrêts de travail recensés en 2019.



(b) Part des salariés concernés par l’arrêt de travail dans chaque catégorie

(a) Répartition des arrêts de travail selon le sexe et la CSP

FIGURE 2.7 – Poids de l’arrêt de travail selon le sexe et la CSP des salariés

La tendance selon laquelle les absences des femmes sont plus fréquents que ceux des hommes semble être contredite dans le diagramme circulaire 2.7a. En effet, comme le présente la figure 2.3, les femmes représentent 46% du portefeuille; cependant il est observé approximativement que seulement 47% des arrêts de travail les concernent. Ce chiffre semble faible lorsque la tendance montre que les femmes sont beaucoup plus concernées par les arrêts multiples que les hommes.

Cela s'explique par le fait que le diagramme ne tient pas compte de la répartition du portefeuille. Les femmes étant moins nombreuses que les hommes, représenter uniquement le poids de leur absences ne permet pas de rendre compte du phénomène décrit par la figure 2.6.

Le graphique en barre suivant explique alors qu'en prenant un poids égal pour chacune des catégories (femmes cadres, femmes non cadres, hommes cadres et hommes non cadres), la fréquence d'arrêts chez les femmes est plus importante que chez les hommes.

Nombre d'arrêts de travail pour 100 salariés		
	Cadres	Non Cadres
Femmes	178	212
Hommes	158	205

Les femmes ont en moyenne un arrêt de travail de plus que les hommes.

Pour la représentation graphique ci-dessous, l'ensemble du portefeuille d'étude est considéré comme un seul et même régime. L'observation porte alors sur la variation du nombre d'indemnisation lorsque la durée de la franchise varie.

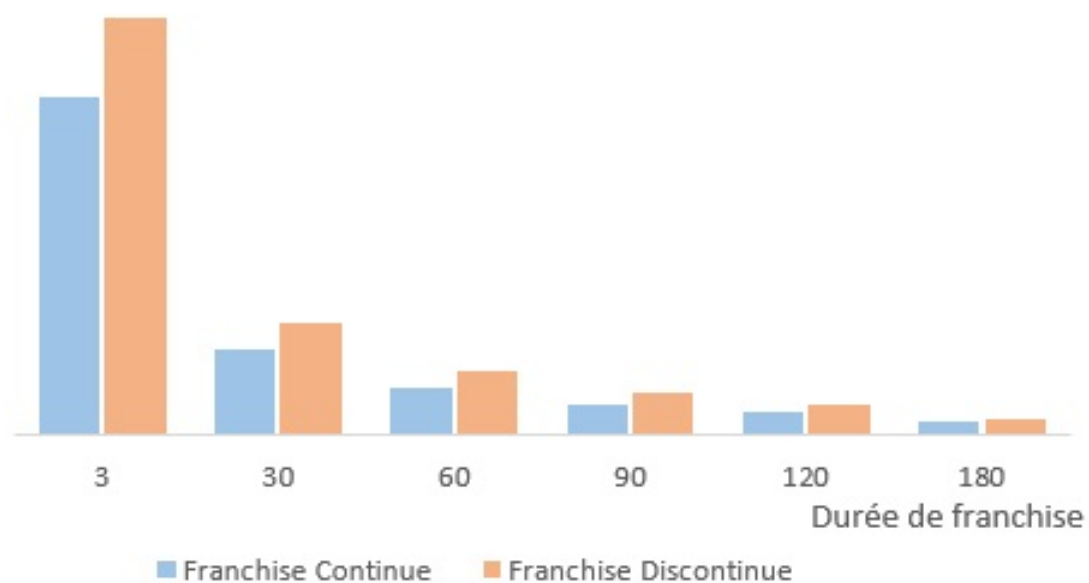


FIGURE 2.8 – Part d'arrêts de travail indemnisés selon la durée de franchise choisie

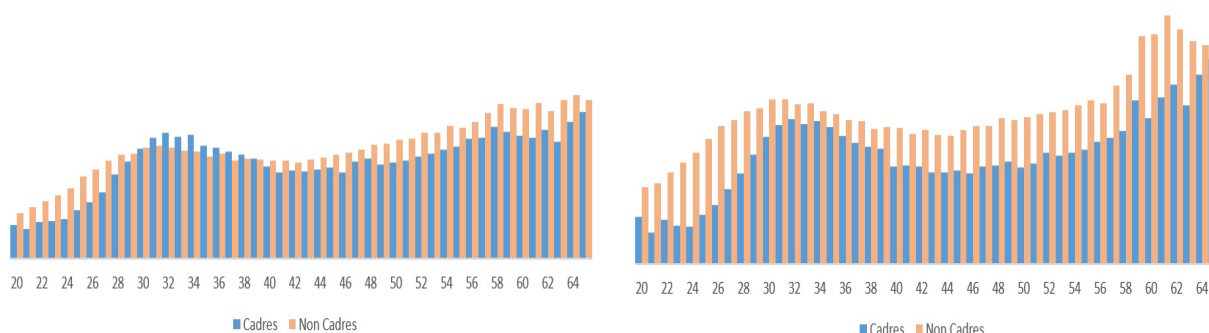
Évidemment, il est observé que sous la franchise discontinue la fréquence d'indemnisation est plus importante que sous la franchise continue. Pour rappel, cela s'explique par le fait

qu'en franchise continue, le délai de carence est appliqué à chaque nouvel arrêt de l'assuré alors qu'en franchise discontinue tous les jours d'arrêts des 365 précédant la date considérée entrent en compte dans la phase constitutive de ce délai.

Aussi, il n'est pas surprenant de voir que le nombre d'indemnisation décroît lorsque la franchise augmente. Plus le délai de carence est long, plus il est difficile de l'atteindre.

Il est intéressant de noter qu'il est très rare de voir souscrit un contrat rédigé avec une garantie incapacité temporaire de franchise discontinue de durée inférieure à 30 jours. Les franchises très courtes ne sont généralement écrites que pour le type continue.

Enfin, pour terminer cette analyse descriptive les durées des absences sont présentées. Dans le cas d'une franchise continue, est représentée la durée moyenne d'un unique arrêt de travail selon l'âge des assurés ; dans le cas discontinue, c'est le cumul des arrêts de travail sur une année glissante qui est traitée. Ces deux cas sont placés sous l'hypothèse d'une indemnisation au premier jour d'absence. Les salariés ayant en moyenne un à deux arrêts par an, il est normal de constater une durée moyenne en arrêt de travail une à deux fois plus grande en franchise discontinue.

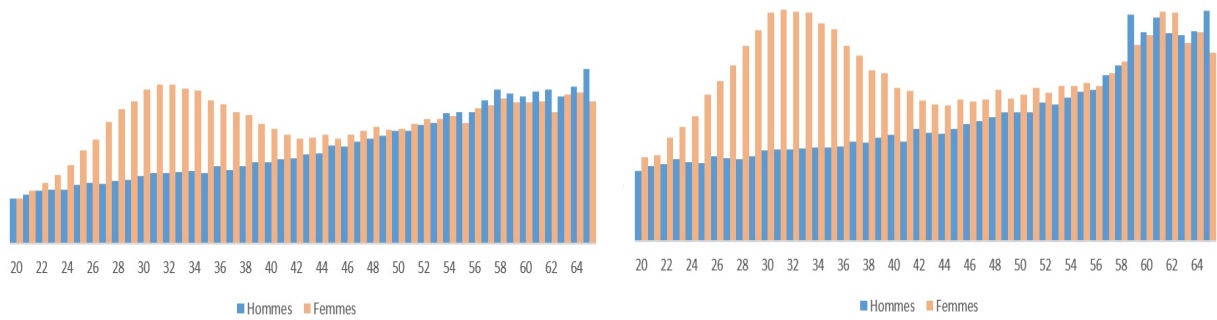


(a) Cas de la franchise Continue

(b) Cas de la franchise discontinue

FIGURE 2.9 – Durée moyenne de la période d'indemnisation de l'arrêt de travail en fonction de l'âge et de la CSP de l'assuré sous l'hypothèse d'une franchise 0 jour

Ici, la segmentation effectuée sépare les cadres et les non cadres tous sexes confondus. En franchise continue, le temps passé en arrêt est plus ou moins similaire quelque soit la catégorie socioprofessionnelle. Pour la franchise discontinue, les absences cumulées des non cadres sont légèrement plus longues que celles des cadres. Cela est dû à la fréquence des arrêts qui est plus importante pour les non cadres que pour les cadres.



(a) Cas de la franchise Continue

(b) Cas de la franchise discontinue

FIGURE 2.10 – Durée moyenne de la période d’indemnisation de l’arrêt de travail en fonction de l’âge et du sexe de l’assuré sous l’hypothèse d’une franchise 0 jour

Dans cette seconde représentation la segmentation est faite au niveau du sexe des salariés. Notamment en discontinue, les femmes présentent un pic de longévité en arrêt de travail aux alentours de la trentaine. Ce phénomène est impactant sur les résultats recherchés et sera donc présenté par la suite.

## 3 | La table d'entrée en incapacité

### 3.1 Notions clés de modèle de durée

Il existe de nombreux problèmes liés à des durées dans les applications statistiques comme la durée de chômage pour les contrats de prévoyance, la durée de scolarisation pour les rentes éducatives ou encore la durée de vie humaine pour les contrats vie/décès et les rentes viagères. Afin de modéliser le temps précédant l'occurrence d'un certain événement, ici la durée avant l'indemnisation des arrêts de travail pour les contrats incapacité/invalidité, il est nécessaire de s'intéresser aux modèles de durée.

L'étude de ces modèles nécessite une théorie spécifique. En effet, les durées ne sont pas des variables aléatoires comme les autres, elles contiennent dans leurs structures des particularités, décrites ci-dessous, dont il va falloir tenir compte.

- Une durée appartient à la classe particulière de variables satisfaisant une contrainte de positivité. Cette remarque a le mérite d'attirer l'attention sur un point : bien qu'il soit possible d'étudier une quantité équivalente sans la contrainte de positivité en passant par le logarithme de la variable étudiée, les distributions usuelles restent cependant différentes de ce que l'on peut trouver dans les domaines d'application "standards". Par exemple la gaussienne, qui peut prendre des valeurs négatives, ne sera plus la référence absolue, elle n'est pas adaptée pour décrire le phénomène souhaité.
- S'intéresser à une durée revient à s'intéresser au risque d'occurrence d'un événement au cours du temps (décès, accident, guérison...). Répondre à cette application introduit un changement de langage et un changement des quantités d'intérêt à estimer.
- Une spécificité purement statistique liée à la structure des données de durée est le phénomène des observations incomplètes. Une durée n'est pas une variable qui s'obtient instantanément. Par exemple estimer la taille moyenne des individus d'une population nécessite juste de mesurer un nombre suffisant individus pris au hasard dans cette population, alors que obtenir la durée de vie résiduelle des individus de cette même population requiert d'attendre le décès de chaque individu de l'échantillon. Or, sur la période d'étude, l'événement d'intérêt peut ne pas être observé pour l'ensemble de l'échantillon ou encore parasité par des événements annexe qu'il est impossible de contrôler (rupture de contrat, défaillance du système d'information, transfert de dossiers ...). Des observations incomplètes sont alors constatées sur la variable d'intérêt. Moins il est possible d'observer un individu pendant longtemps, moins l'événement étudié a de chance de se produire. Ne pas tenir compte de ce phénomène d'observation incomplète peut introduire un biais significatif de l'étude.

### 3.1.1 Les fonctions de base

#### a) La fonction de survie

Soit  $T_x$  une variable aléatoire positive représentant dans notre cas la durée de pré-couverture des arrêts de travail. La distribution de  $T$  est donnée par sa fonction de répartition qui décrit la probabilité que l'individu observé bénéficie de la couverture incapacité temporaire entre l'âge  $x$  et  $x + t$  :

$$F_x(t) = \mathbb{P}(T_x \leq t)$$

La fonction de répartition  $F$  définit de manière unique la loi de  $T_x$ . Un objet appelé la fonction de répartition empirique  $\hat{F}$  permet d'estimer cette fonction de répartition car elle converge presque sûrement vers  $F$ .

En effet pour un ensemble d'observations  $(T_1, \dots, T_n)$  i.i.d (indépendantes et identiquement distribuées) et de même loi que  $T_x$ , par la loi forte des grands nombres :

$$\hat{F}(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{T_i \leq t} \xrightarrow[n \rightarrow +\infty]{p.s.} \mathbb{E}[\mathbb{1}_{T_x \leq t}] = \mathbb{P}[T_x \leq t] = F(t)$$

La fonction de survie d'une variable aléatoire  $T$  est définie par :

$$S_x(t) = \mathbb{P}(T_x \geq t) = 1 - F(t^-)$$

Cette fonction est strictement équivalente à la fonction de répartition car il est possible de passer de l'une à l'autre de façon immédiate. Cependant, la fonction de survie est historiquement privilégiée dans la littérature de par l'application que vise l'étude des modèles de durées. Tout comme la fonction de répartition, la fonction de survie définit de manière unique la distribution de  $T$  et peut être estimée à partir de  $\hat{F}$ .

#### b) Le taux de risque instantané

Pour décrire une loi de probabilité, en général, une densité est un objet plus parlant que la fonction de répartition. Tout comme la densité, le taux de risque instantané associé à une fonction de répartition  $F$  notée  $\mu_F$  permet de visualiser un certain nombre de propriétés liées à des durées plus facilement. Il est le plus utilisé par les praticiens pour remplacer la fonction de densité dans les modèles de survie. Ce risque peut être défini dans le cas de variable discrètes ou continues.

##### Cas d'une variable aléatoire $T$ discrète :

Soit  $T \in \{t_1, \dots, t_n, \dots\} \subset \mathbb{R}$

$$\mu_F(t) = \mathbb{P}(T = t | T \geq t) = \frac{\mathbb{P}(T = t)}{\mathbb{P}(T \geq t)}$$

Il s'agit de la probabilité que l'événement se produise à la date  $t$  sachant qu'il ne s'est pas produit avant. C'est un indicateur parlant pour quantifier le risque d'occurrence.



Par exemple dans le cas où  $T_1$  représente l'occurrence de l'arrêt de travail. Lorsque pour un individu à un âge  $x_i$  (en jours),  $\mu_F(x_i)$  est grand, cet individu a de fortes chances d'être absent à cette date. Inversement,  $\mu_F(x_i)$  faible indique que cette probabilité est faible.

Une relation, assez simple à obtenir, relie la fonction de survie  $S$  et le taux de risque instantané  $\mu_F$  :

$$S(t) = \prod_{i|t-i < t} (1 - \mu_F(t_i))$$

Cette relation indique qu'à partir de la donnée du taux de risque instantané, il est possible de retrouver la fonction de survie et donc de caractériser de manière unique la loi de  $T$ .

### Cas d'une variable aléatoire $T$ continue :

Dans le cas continue, il n'est pas possible de définir  $\mu_F$  comme précédemment. En effet, dans ce cas, la probabilité  $\mathbb{P}(T = t)$  est toujours nulle.

Afin de prolonger la définition du taux de risque instantané, il faut étudier la probabilité que la variable aléatoire  $T$  appartienne à un intervalle infinitésimal  $[t, t + \delta t]$  rapportée à la taille de cet intervalle. La fonction s'écrit alors :

$$\mu_F(t) = \lim_{\delta t \rightarrow 0^+} \frac{\mathbb{P}(T \in [t, t + \delta t] | T \geq t)}{\delta t}$$

Les liens entre la fonction de survie et  $\mu_F$  deviennent alors :

$$\mu_F(t) = \frac{f(t)}{S(t)} \quad \text{où alors} \quad S(t) = \exp\left(-\int_{-\infty}^t \mu_F(u) du\right)$$

Les méthodes d'estimations qui suivront se basent sur ces fonctions décrites ci-dessus.

## 3.1.2 Les phénomènes d'observations incomplètes

L'existence d'observations incomplètes (censure et troncature) est une caractéristique non négligeable des données de survie. Il s'agit de perte d'informations due à la structure temporelle sous-jacente. Durant la récolte de données liées à des durées, des phénomènes parasites peuvent perturber l'observation. Ces phénomènes parasitent les méthodes d'estimation et donc entraînent une écriture particulière de la vraisemblance.

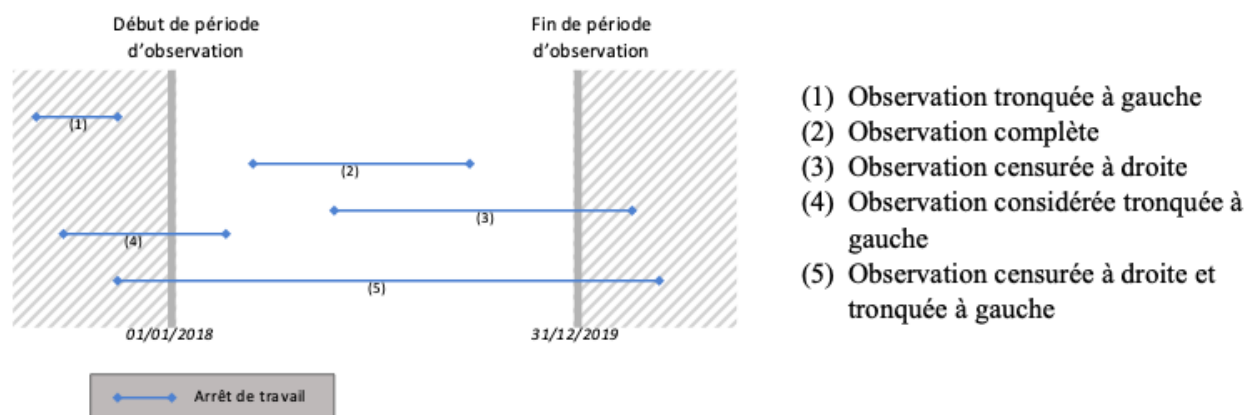


FIGURE 3.1 – Les différents types d'observations

### a) La censure à droite

La censure est le phénomène le plus couramment rencontré lorsque l'on recueille des données de survie.

Une durée de survie est dite censurée à droite si l'individu n'a pas subi l'évènement à sa dernière observation. La variable aléatoire de censure est sans précision supplémentaire sur la source de l'aléa. Ce phénomène entraîne des difficultés d'estimation dans la partie droite de la distribution qui est moins bien observée. En effet, les individus associés à une grande valeur de la durée étudiée ont plus de chance d'être censurés, ce qui entraîne une sous représentation des grandes durées parmi les observations complètes. Négliger la censure aboutit alors à une sous-estimation des valeurs prises par la variable de durée.

Dans le cadre de l'incidence en arrêt de travail, les causes possibles de censure à droite peuvent être :

- l'assuré n'a pas cumulé suffisamment de jours en arrêt de travail sur un an pour atteindre la franchise et donc n'a pas bénéficié de la couverture incapacité à la fin de l'étude (l'évènement d'intérêt n'a pas eu lieu) ;
- l'assuré a quitté l'étude en cours (évènement parasite) à une date à laquelle il n'a pas encore subi l'évènement étudié. Cela peut être du à une rupture de contrat, à un décès, ou tout simplement à la fin de la période d'observation.

### b) La troncature à gauche

La troncature à gauche est également un phénomène courant en analyse de survie. Ce phénomène impacte négativement la partie gauche des observations. Une observation est dite tronquée si elle est conditionnelle à un autre évènement. Dans le cadre de cette étude, une observation est considérée tronquée lorsque la date de début du contrat de travail d'un salarié est antérieure au début de la période d'observation (01/01/2018). Les absences durant ce laps de temps ne sont alors pas observées.

### c) Prise en compte des observations incomplètes

On peut alors déterminer les sinistres censurés, c'est-à-dire ceux dont la date de fin dépasse la période d'exposition de l'assuré (l'indicatrice de censure vaut alors 1) ainsi que les sinistres hors périmètre, c'est-à-dire ceux dont la date de fin ne dépasse pas le début de l'exposition, on supprimera alors ces sinistres.

On construit aussi l'indicatrice de troncature qui indique les sinistres tronqués, c'est-à-dire ceux dont la survenance est antérieure à la période d'exposition.

La variable "censure" pour chaque individu de la base de données indique alors le nombre de censures ou troncatures qui ont été recensées lors de l'observation des différents arrêts de travail de l'individu durant les deux années étudiées.

## 3.2 Méthodes d'estimation des taux d'incidence

De manière générale, l'incidence correspond au taux de nouveaux entrants dans un état pré-défini. Dans le cadre de l'arrêt de travail, l'incidence voulue est celle des salariés, selon leur âge, sexe et collègue, entrants dans l'état "*l'individu a bénéficié de la couverture pour la garantie incidence temporaire en franchise discontinue à l'âge  $x$* ". Cet événement d'intérêt sera noté dans la suite  $I_x$ .

Afin de calculer ces taux d'entrée en incapacité, la durée étudiée est, durant la période d'observation, le temps passé en activité à laquelle sont ajoutés les jours d'absences sous franchise (qui sont par définition non indemnisés). L'événement "entrée en incapacité" est donc défini comme "entrée en incapacité + maintien au-delà de la franchise". Ces observations sont notées pour la suite  $(T_1, \dots, T_n)$ . L'analyse est faite sur quatre sous groupes de salariés suivant des critères de sexe (femmes ou hommes) et de catégorie socio-professionnelle (cadres ou non cadres). Les populations d'intérêts sont donc les femmes cadres, les hommes cadres, les femmes non cadres et les hommes non-cadres.

Pour cela, il existe différents types d'estimateurs :

**Les estimateurs paramétriques :** Le choix d'un modèle paramétrique suppose que la loi de probabilité de la durée de vie  $T$  appartient à une classe de distributions de type connu, fonction de paramètres dont l'objectif sera de les estimer à partir d'un ensemble d'observations.

**Les estimateurs non paramétriques :** Les modèles non paramétriques permettent d'estimer l'une des différentes fonctions caractérisant la distribution de la variable  $T$  sans faire aucune hypothèse a priori sur celle-ci. Les modèles non paramétriques font alors appel à d'autres estimateurs empiriques en présence de censure et/ou de troncature.

L'avantage d'une estimation non paramétrique par rapport à une estimation paramétrique repose sur la convergence vers la vraie distribution sans émettre d'hypothèses limitant ainsi les erreurs du modèle. Afin de pouvoir utiliser de tels estimateurs, il est essentiel d'émettre l'hypothèse d'indépendance entre les salariés.

### 3.2.1 L'estimateur de Kaplan Meier

Pour rappel, les durées étudiées  $(T_1, \dots, T_n)$  sont les temps passés sur une année précédant une éventuelle application de la garantie incapacité temporaire.

#### a) Estimation en présence de censure

En présence du phénomène de censure, l'échantillon  $(T_1, \dots, T_n)$  est pollué par des variables de censure  $(C_1, \dots, C_n)$  de même loi  $C$ . Les données observées sont alors :

$$((Y_1, \delta_1), \dots, (Y_n, \delta_n)) \quad \text{avec} \quad \begin{aligned} Y_i &= \inf(T_i, C_i) \\ \delta_i &= \mathbb{1}_{T_i \leq C_i} \end{aligned}$$

Les  $\delta_i$  indiquent si une observation est complète ou non.

Afin de pouvoir estimer la distribution dans ce type de modèle, il est fondamental de formuler une hypothèse dite d'identification sur la façon dont les  $(C_i)_{i=1, \dots, n}$  et les  $(T_i)_{i=1, \dots, n}$  dépendent l'une de l'autre. Sans cette hypothèse, il est possible de montrer que la distribution de  $T_x$  ne peut être estimée en toute généralité. Usuellement il est supposé que  $T_i$  est indépendant de  $C_i$ . De façon générale, rien ne permet de tester cette hypothèse mais dans le cas de l'arrêt de travail, il paraît raisonnable de la supposer vraie.

Dans le problème de données censurées, la fonction de répartition empirique  $\hat{F}$  ne peut être calculée car les  $(T_i)_{i=1, \dots, n}$  ne sont pas observées directement. Il n'est donc pas possible d'estimer la fonction de répartition via  $\hat{F}$  ni par les quantités liées à cette fonction (moyenne et variance empirique). L'ensemble des informations disponibles sur la loi observée se résume en trois fonctions qu'il est possible d'estimer de façon consistante (par la loi des grands nombres) :

$$\begin{aligned} \hat{H}(t) &= \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{Y_i \leq t} & \xrightarrow[n \rightarrow +\infty]{p.s.} & \mathbb{E}[\mathbb{1}_{Y \leq t}] = \mathbb{P}[Y \leq t] = H(t) \\ \hat{H}_0(t) &= \frac{1}{n} \sum_{i=1}^n (1 - \delta_i) \mathbb{1}_{Y_i \leq t} & \xrightarrow[n \rightarrow +\infty]{p.s.} & \mathbb{P}[Y \leq t, \delta = 0] = H_0(t) \\ \hat{H}_1(t) &= \frac{1}{n} \sum_{i=1}^n \delta_i \mathbb{1}_{Y_i \leq t} & \xrightarrow[n \rightarrow +\infty]{p.s.} & \mathbb{P}[Y \leq t, \delta = 1] = H_1(t) \end{aligned}$$

A partir de ces trois fonctions, il est possible de retrouver la loi de  $\delta$  et la loi de  $Y$ , c'est à dire toute la loi des observations. L'estimateur de la fonction de répartition  $F$  est alors construit à partir de ces fonctions de base.

#### b) Construction de l'estimateur

Par souci de simplicité, le but est l'estimation de  $F$  à partir d'un estimateur discret. Même en cas de distribution réelle continue, la distribution discrète qui sera définie convergera de toutes les façons vers  $F$ . Chercher cet estimateur revient à déterminer un taux de risque instantané  $\mu_F$ . En effet, comme indiqué plus haut, il existe une relation entre le taux de risque instantané et la fonction de répartition discrète qui est :

$$\hat{S}(t) = 1 - \hat{F}(t^-) = \prod_{Y_i \leq t} (1 - \mu_{\hat{F}}(Y_i))$$

Estimer une telle fonction de répartition revient donc à déterminer l'estimateur  $\mu_{\hat{F}}$  de la fonction de risque instantané  $\mu_F$  en utilisant les fonctions  $H$ ,  $H_1$  et  $H_0$  qu'il est possible d'estimer. Ces fonctions peuvent se réécrire après quelques manipulations mathématiques :

$$\begin{cases} 1 - H(t) &= (1 - \mathbb{P}(C \leq t)) \times (1 - F(t)) \\ dH_1 &= (1 - \mathbb{P}(C < t)) \times dF(t) \\ dH_0 &= (1 - F(t^-)) \times d\mathbb{P}(C < t) \end{cases}$$

*Remarque : ces résultats sont obtenus en se basant essentiellement sur l'hypothèse d'indépendance entre les variables  $T$  et  $C$ . Avec une autre structure de dépendance entre ces deux variables, les formules obtenues auraient été différentes.*

Alors, à partir de l'expression de  $\mu_F$  également décrite dans la section précédente, la fonction devient :

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

Cette quantité est estimable à partir des observations par :

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

Enfin, en utilisant la relation entre la fonction de répartition et le taux de risque instantané, l'estimateur de la fonction de survie associée à la fonction de répartition cherchée s'écrit :

$$\hat{S}(t) = 1 - \hat{F}(t) = \prod_{Y_i \leq t} \left( 1 - \frac{d\hat{H}_1(Y_i)}{1 - \hat{H}(Y_i^-)} \right) = \prod_{Y_i \leq t} \left( 1 - \frac{\delta_i}{1 - \sum_{j=1}^n \mathbb{1}_{Y_j \geq Y_i}} \right)$$

Dans cette formule,  $\delta_i$  s'interprète comme étant le nombre d'indemnisation d'incapacité temporaire à l'instant  $Y_i$  et  $1 - \sum_{j=1}^n \mathbb{1}_{Y_j \geq Y_i}$  est le nombre de personnes en activité ou en arrêt de travail sous la période de franchise juste avant l'instant  $Y_i$ .

*Remarque : ce résultat est obtenu sous l'hypothèse d'absence d'ex-aequo dans l'échantillon. Autrement dit tous les  $Y_i$  prennent des valeurs distinctes. En cas d'ex-aequo, la définition est légèrement plus lourde mais assez similaire. Il faut introduire les différentes valeurs prises par  $Y_i$  dans le cas où plusieurs événements se produisent.*

### c) Estimateur en présence de censure et de troncature

En présence de variable de censure  $(C_1, \dots, C_n)$  de même loi que  $C$  et des variables de troncature  $(\tau_1, \dots, \tau_n)$  de même loi qu'une variable de troncature  $\tau$ . Les données observées sont alors :

$$((Y_1, \delta_1, \tau_1), \dots, (Y_n, \delta_n, \tau_n)) \quad \text{avec} \quad \begin{cases} Y_i \geq \tau_i \\ Y_i = \inf(T_i, C_i) \\ \delta_i = \mathbb{1}_{T_i \leq C_i} \end{cases}$$

Une observation  $(Y_1, \delta_1)$  a pour loi la loi de  $(Y, \delta)$  sachant  $Y \geq \tau$ .

La logique de construction de l'estimateur de Kaplan Meier peut alors être adaptée à ce cas de figure et aboutir à la définition suivante.

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

Cette formule est conditionnée aux mêmes hypothèses que précédemment ainsi qu'une hypothèse supplémentaire d'indépendance entre  $\tau$  et  $(T, C)$ .

#### d) Résumé qualitatif

L'estimateur de Kaplan meier ne met de la masse qu'aux observations non censurées car seules ces observations apportent une information complète sur l'événement étudié. Cette masse diffère d'une observation à l'autre selon le rang de l'observation dans l'échantillon. Le poids attribué est plus important pour les grandes observations non censurées afin de compenser le déficit de ces dernières. En effet, une observation ayant une grande durée augmente le risque de censure de cette durée contrairement aux petites valeurs de  $T_i$ . Cependant, avec cette méthode, la qualité d'estimation de ces queues de distribution est plus instable car les poids attribués peuvent vite devenir très grands. Dans notre cas cela correspond aux assurés ayant une fréquence d'arrêt de travail nulle ou très faible par rapport à la moyenne. Ce manque d'homogénéité dans les queues de distribution font que le modèle de Kaplan Meier perd de son efficacité pour estimés certaines valeurs.

### 3.2.2 L'estimateur de Hoem

L'approche la plus simple pour le calcul des taux d'incidence est l'estimateur binomial. Cependant, il ne considère pas les données sujettes aux phénomènes de censure et de troncature. Or exclure ces données entraînerait une perte d'information et une surestimation du risque d'incapacité temporaire.

L'estimation de Hoem est une généralisation de l'estimateur binomial permettant la prise en compte des données censurées et tronquées.

En notant,  $\alpha_i$  la date de début d'observation d'un individu d'âge  $x$ , et  $\beta_i$  la date de fin d'observation pour l'âge  $x$ , chaque individu d'âge  $x$ , a une exposition au risque sur un intervalle  $[\alpha_i, \beta_i]$  inclus dans  $[x, x + 1]$ . La durée  $(\beta_i - \alpha_i)$  correspond à la durée de présence effective d'un individu à l'âge  $x$ .

Par souci de simplification, l'hypothèse de linéarité des taux d'entrée en incapacité à un âge  $x$  est considéré. C'est à dire, pour  $t \in \mathbb{N}+$ ,

$$\mathbb{P}[T_x < t | T_x > x] = {}_t q_x = t \times q_x$$

Donc, sur l'intervalle  $[\alpha_i, \beta_i]$  l'occurrence de l'événement  $T_x$  à alors une probabilité de

$$\mathbb{P}[T_x < \beta_i - \alpha_i | T_x > x + \alpha_i] = {}_{\beta_i - \alpha_i} q_{x + \alpha_i} = (\beta_i - \alpha_i) q_{x + \alpha_i}$$

En considérant les notations :

- $N_x$  le nombre d'individus en activité ou sous période de franchise à l'âge  $x$ .
  - $q_x^i = (\beta_i - \alpha_i)q_{x+\alpha_i}$  la probabilité de bénéficier de la garantie incapacité temporaire dans l'année pour un individu d'âge  $x$
  - $(\delta_1, \dots, \delta_{N_x})$  variables aléatoires i.i.d de loi  $\delta_i \sim \mathcal{B}(q_x^i)$ .
- C'est à dire  $\delta_i = \begin{cases} 1 & \text{si l'événement } T_x \text{ se réalise} \\ 0 & \text{sinon.} \end{cases}$

Alors le nombre de réalisation de  $T_x$  entre l'âge  $x$  et  $x+1$  est décrite par une variable aléatoire de loi binomiale

$$D_x = \sum_{i=1}^{N_x} \delta_i \sim \mathcal{B}(N_x, q_x^i) \iff \mathbb{P}_i(D_x = d_x) = \binom{N_x}{d_x} \times q_x^{d_x} \times (1 - q_x)^{N_x - d_x}$$

Pour estimer le paramètre  $q_x$ , la méthode du maximum de vraisemblance associée à la réalisation d'un nombre  $d_x$  de réalisation de l'événement d'intérêt  $T_x$  pour l'âge  $x$  est utilisée.

Le principe de cette méthode des moments est de maximiser la vraisemblance de la loi de  $D_x$  selon  $q_x$ , ce qui revient à résoudre :

$$\operatorname{argmax}_{q_x} L(q_x) = \prod_{k=1}^{N_x} \mathbb{P}_i(D_x = d_x)$$

Puis l'estimateur  $\hat{q}_x$  est obtenu en déterminant le  $q_x$  annulant la dérivé  $\frac{d \ln[L(q_x)]}{dq_x}$ .  
Après calcul la formule de l'estimateur cherché s'écrit :

$$\hat{q}_x = \frac{d_x}{\sum_{k=1}^{N_x} (\beta_k - \alpha)}$$

L'estimateur de Hoem est ensuite appliqué aux données suivants les critères de segmentations. Cet estimateur à l'avantage d'être sans biais et est simple à implémenter informatiquement.

### 3.2.3 Résultats et choix

#### a) Présentations des résultats

Les deux méthodes d'estimation des taux d'entrée en incapacité décrites précédemment sont alors implémentées informatiquement. Les taux d'incidences sont calculés selon l'âge, le sexe et la CSP de chacun des assurés. La fréquence de mise en place d'indemnisation de l'incapacité dépendant également de la franchise, les taux bruts d'incidence sont évalués aux franchises clés 30, 45, 60, 75, 90, 100, 120, 150, 180 et 360 jours. Pour rappel, ces franchises sont celles le plus fréquemment employées dans les contrats. Les tables d'incidence pour les franchises intermédiaires sont par la suite déduites par interpolation linéaire. Les résultats obtenus pour la franchise 30 jours sont représentés ci-dessous :

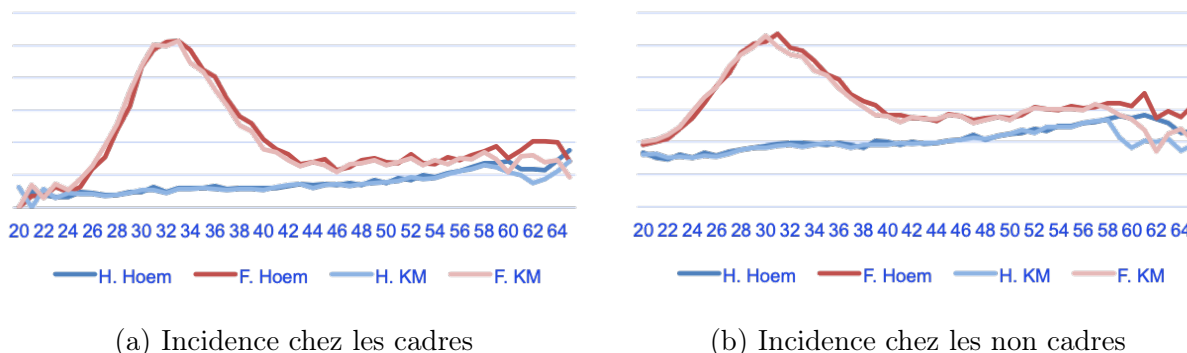


FIGURE 3.2 – Courbe des taux bruts d'incidence suivant les méthodes d'estimation de Kaplan Meier et Hoem

Pour chaque âge est alors représenté ci-dessous l'écart entre ces deux méthodes. Pour chaque âge est représenté le pourcentage  $x$  d'écart relatif entre les deux estimations défini tel que  $taux_{KaplanMeier} = taux_{Hoem} \times (1 + x)$ . Hormis pour les âges extrêmes dont la volatilité est liée au manque de données, l'écart entre les deux méthodes d'estimation étudiées précédemment est en moyenne inférieur à 5%. Cela permet de conforter l'idée de cohérence et d'exactitude de l'implantation des deux modèles.

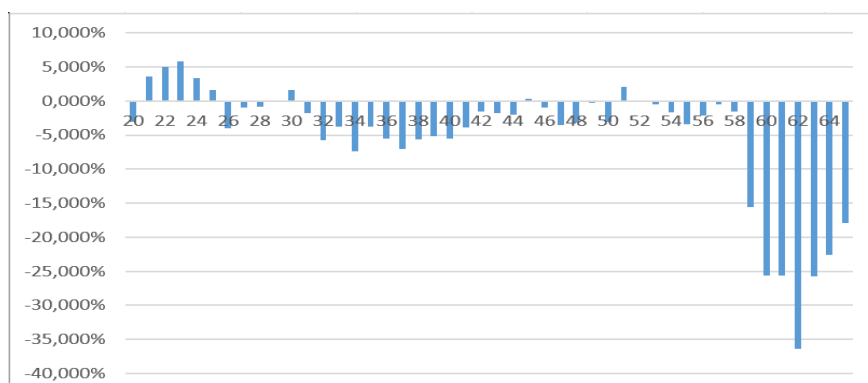


FIGURE 3.3 – Incidence chez les cadres

Au vu de ces graphiques le choix de méthode se porte sur l'estimateur de Hoem. Les deux principales raisons de ce choix sont :

- Pour les âges jeunes tout comme pour les âges élevés, la fluctuation des taux, notamment causée par le manque de données, semble moins conséquente et plus représentative de la réalité. En effet la méthode de Kaplan Meier étant basée sur une estimation asymptotique, elle est d'autant plus fiable que le nombre d'observations est important. Cela n'est pas forcément le cas pour les très grands et très petits âges, d'où le manque de stabilité des courbes ci-dessus à ces âges. En particulier, ne sont pas massivement représentés dans le portefeuille d'étude : les très jeunes cadres dû au niveau de qualification requis par ce statut, ainsi que les non cadres âgés car plus l'activité est dite pénible, plus l'âge de retraite est avancé. Ce soucis de masse de données a un impact moins important lorsque la méthode de Hoem est utilisée.



- Aux âges les plus représentés dans le portefeuille (entre 30 et 50 ans), le graphique en barre 3.3 montre bien qu'il est plus prudent de considérer la méthode d'estimation de Hoem. Les taux sont en moyenne 5% plus élevés en utilisant cette méthode plutôt que l'estimation de Kaplan Meier.

L'allure des courbes d'incidence est similaire quelque soit la franchise. La différence est que plus la durée de franchise est longue, moins l'événement d'intérêt se produit ; et donc les taux d'incidence sont logiquement plus faibles. Ce phénomène est observable dans les deux graphiques qui suivent.

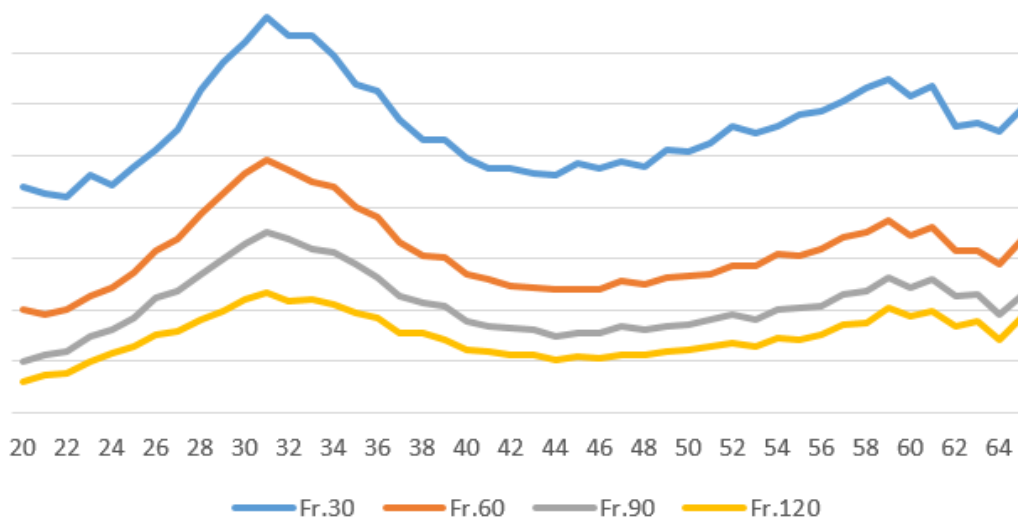


FIGURE 3.4 – Allure de la courbe d'incidence globale (tout sexes et CSP confondus) pour différentes franchises

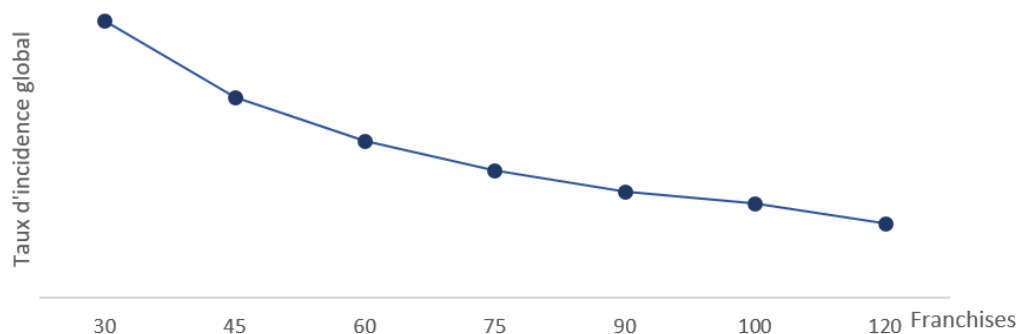


FIGURE 3.5 – Evolution des taux globaux d'incidence (tout âge, sexe et CSP) suivant le paramètre de franchise

## b) Interprétation des résultats

La figure 3.2 met en avant une particularité de la courbe des taux bruts d'incidence. Entre environ 25 et 45 ans, l'incidence chez la femme cadre ou non cadre peut doubler voir tripler pour certaines franchises. Le pic de cette bosse est observé chez les trentenaires. Plus haut dans cette étude, l'analyse descriptive des données mettait bien en évidence le fait que les femmes en général ont une fréquence d'arrêts de travail plus élevée que les hommes, ainsi qu'un maintien en arrêt plus long, notamment entre 25 et 45 ans. Ces deux phénomènes influent grandement sur l'incidence de la franchise discontinue car ils font croître la probabilité que la durée cumulée des jours d'absences sur l'année dépasse la durée de la franchise considérée.

### Les grossesses

Une des raisons principales de cette particularité entre 25 et 45 ans est la maternité. Bien que la base d'étude ne considère pas les absences de maternité (qui bénéficient d'une garantie propre) l'année de leur grossesse, le comportement des femmes diffère des autres années.

Pour étudier cela, le portefeuille est restreint aux femmes dont au moins un congé maternité à été recensé. Sous ces critères le congé maternité représente 7,5 % des arrêts de travail et concerne environ 16% des salariés de cette tranche. Ce chiffre est supérieur d'environ 17% aux observations nationales de l'INSEE. Cela s'explique notamment par la répartition par âge du portefeuille d'étude (cf. figure 2.1) où le nombre de femmes autour de la trentaine prédomine contrairement à ce qui est observé au niveau national. Bien que seulement 16% des effectifs féminin de 25 à 45 ans soient concernés par la maternité, ces futures mères représentent à elles seules un quart des arrêts de travail de l'ensemble des femmes de 25 à 45 ans.

Part de grossesse		
	Femmes	Femmes 25-45 ans
INSEE	6,64%	13,56%
Portefeuille	9,09%	15,93%

Durant l'année précédant ce congé, la fréquence des incapacités temporaires croît exponentiellement lorsque la date de congé maternité approche. Le graphique ci-dessous montre le mois précédant l'arrêt de travail pour maternité, la fréquence des arrêts de travail chez les femmes est 5 à 6 fois plus élevée que l'année d'avant. A noter que la date de début de grossesse est approximativement fixée car bien que la durée du congé prénatal est de 6 à 8 semaines, la mère a légalement le droit de renoncer à une partie de son congé sous certaines conditions.

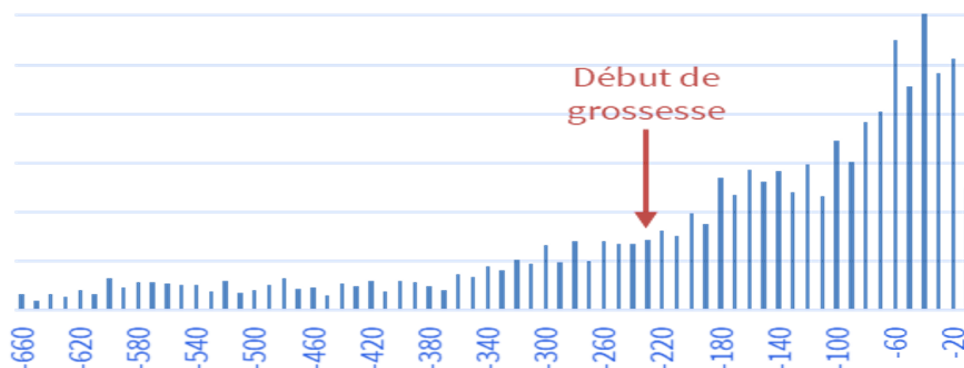
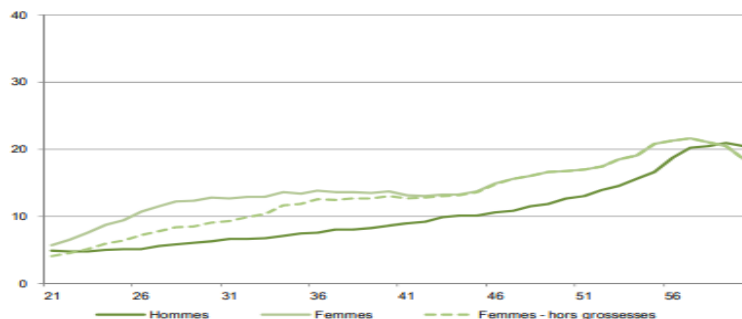


FIGURE 3.6 – Proportion des arrêts de travail en fonction du nombre de jours (par tranche de 10) précédant le congé maternité

En moyenne, l'année précédant le congé maternité les femmes âgées entre 25 et 45 ans cumulent 26 jours d'arrêts de travail répartis en 2 ou 3 absences distinctes contre 16 jours annuels étalés sur 1 ou 2 arrêts hors année de grossesse. La part des arrêts maladie liés à la grossesse (hors maternité) explique environ 40 % de l'écart entre femmes et hommes âgés de 21 à 45 ans.

A noter que ce phénomène est d'autant plus visible que la durée de la franchise est courte. Cela est notamment visible sur le graphique 3.5 où lorsque la durée de franchise augmente, le pic des 25-45ans s'atténue de plus en plus.

Cependant les grossesses n'expriment pas la totalité des raisons pour lesquelles les femmes sont plus souvent en arrêt de travail que les hommes. En effet, les rapports gouvernementaux du DREES (Direction de la Recherche, des Études, de l'Évaluation et des Statistiques) concernant les dépenses de santé (cités en bibliographie) montrent bien que ce phénomène de pic d'absentéisme des femmes entre 25 et 45 ans est également présent lorsque les femmes en période de gestation sont retirées du portefeuille d'étude.



Champ : Salariés âgés de 21 à 60 ans affiliés au régime général, hors sections locales mutualistes.  
Sources : Échantillon généraliste de bénéficiaires (EGB) 2011 2012, déclarations annuelles de données sociales (DADS) 2011.

FIGURE 3.7 – Nombre de jours d'arrêt maladie indemnisés (franchise continue) dans l'année par âge et par sexe

Bien que beaucoup plus atténuée sans les grossesses, la bosse reste présente. Ce phénomène a déjà fait l'objet d'études. Des éléments de réponses sont exposés dans le paragraphe qui suit.

### **L'absentéisme chez les femmes**

D'autres études annexes montrent également que les femmes sont, en général, plus touchées par l'absentéisme que les hommes. Ce constat est mis en avant dans le baromètre du groupe de conseil Ayming ainsi que dans les rapports gouvernementaux du DREES. Selon ces études, les trois principales causes sont :

- les troubles musculo-squelettiques : les femmes ont plus tendance que les hommes à occuper des postes à gestes répétitifs. Cela accroît leur risque à développer des troubles musculo-squelettiques générant souvent des arrêts de plusieurs mois.
- la "charge du foyer" : une grande partie du temps libre chez les femmes est utilisé à s'occuper du foyer. Lorsque l'on ajoute cette charge au temps de travail, l'organisme des femmes se fatigue plus rapidement donc engendre des absences qui durent plus longtemps.
- le nombre important de familles monoparentales, gérées à 83% par des femmes : un quart des familles est monoparental dont la plupart sont gérées par les mères. Ce troisième élément accentue la charge extra-professionnelle des femmes et donc leur tendance à être en arrêt de travail pour récupérer. A noter que cette situation concerne essentiellement les femmes entre 25 et 45 ans qui ont de jeunes enfants à leur charge. Ce phénomène participe donc à augmenter les pics d'absentéisme observé pour cette tranche d'âge.
- les femmes sont moins représentées dans les emplois les mieux rémunérés : cela explique à 5% l'absentéisme plus fréquents des femmes. Moins confiantes dans l'utilité de leur travail, elles sont aussi moins satisfaites de leur salaire compte tenu du travail réalisé. Tout cela fait qu'elles sont moins épanouies dans leur travail et donc plus sujettes aux risques psychosociaux.

Toutes ces raisons permettent de compléter l'étude faite sur les grossesses et donc de justifier la tendance des courbes de taux bruts observées.

### 3.3 Méthodes de lissage des taux bruts

Lors de la construction d'une loi d'incidence, après l'étape de l'estimation de taux bruts, il faut « lisser » les valeurs brutes obtenues pour représenter plus fidèlement la loi (inconnue) recherchée. En effet, ces valeurs présentent une certaine irrégularité, qui ne reflète pas le phénomène à mesurer mais plutôt l'imperfection des conditions de l'expérience dues à la fluctuation de l'échantillonnage.

Formellement, le taux d'incidence  $q_x$ , estimé par  $\hat{q}_x$  conduit à une erreur  $e_x = \hat{q}_x - q_x$ . L'objectif est de diminuer cette erreur, tout en construisant une courbe des taux en fonction de  $x$  plus « lisse » que la courbe des taux bruts.

**Régularité et précision :** La procédure de révision des données brutes fait intervenir deux types de contraintes à prendre en considération conjointement :

- la précision (ou fidélité) : Les taux ajustés doivent être proches des taux initiaux ;
- la régularité : les taux ajustés doivent être aussi réguliers que possible.

**Approche paramétrique ou non paramétrique :** Pour information, sémantiquement le terme «ajustement» est utilisé dans le cadre paramétrique et le terme «lissage» dans le cadre non paramétrique.

Le but de chacune de ces approche est de :

- Pour l'approche paramétrique : fixer une forme a priori pour la loi sous-jacente en vue d'ajuster une distribution définie par un ou plusieurs paramètres.  
(exemples : lissages par splines, modèle de Makeham)
- Pour l'approche non paramétrique : définir un certain nombre de traitement à appliquer aux données brutes initiales pour les rendre plus « lisses ».  
(exemples : moyennes mobiles<sup>2</sup>, lissages de type Whittaker-Henderson et leur extension dans un cadre bayésien plus général).

Dans le cadre de cette étude, le choix de la méthode de lissage s'est porté uniquement sur le lissage de Whittaker-Henderson. Il s'agit d'une méthode non paramétrique très utilisée pour le lissage des taux bruts. Au lieu de multiplier les méthodes de lissage en vue d'en faire une comparaison quantitative, il a été choisi de se concentrer sur une seule méthode régulièrement utilisée et de déterminer les paramètres entrant en jeu dans son application le plus justement que possible.

#### 3.3.1 Lissage de Whittaker-Henderson

Le lissage de Whittaker-Henderson est une méthode qui repose sur la minimisation d'une combinaison linéaire entre un critère de régularité et un critère de fidélité. L'idée de cette méthode est de trouver un compromis entre la fidélité aux taux brut et la régularité (le lissage) voulue.

La difficulté pratique de cette méthode réside dans le choix des paramètres de lissage optimaux. Ces choix, souvent fait subjectivement, seront basés dans cette étude sur des tests statistiques, améliorant ainsi la qualité de l'ajustement. Le test usuel du Khi-2, présenté dans un travail publié en 1981 par L. Giesecke et largement utilisé dans le cadre des ajustements paramétriques, sera complété par des analyses de sensibilités.

Au regard des données, le lissage est a dimension 1 car pour chaque âge n'est associé qu'un seul point.

### a) Le critère de fidélité

Le critère de fidélité consiste à prendre en compte le fait que les taux estimés par âge doivent être proches des taux bruts. Il mesure la distance euclidienne entre l'incidence lissée  $\tilde{q}_x$  et l'incidence estimée  $\hat{q}_x$  pondérée par le poids  $w_x$  associée aux estimations :

$$F = \sum_{x=20}^{65} w_x (\tilde{q}_x - \hat{q}_x)^2$$

Cette distance mesure l'écart aux valeurs estimées. Les poids sont, dans cette étude, définis *a posteriori* en fonction de l'effectif  $N_x$  d'âge moyen  $x$  observé et l'effectif moyen  $\bar{N}$  :

$$w_x = \frac{N_x}{\bar{N}}$$

Matriciellement, F peut être réécrit  $F = (\tilde{Q} - \hat{Q})^\top \times W \times (\tilde{Q} - \hat{Q})$  en notant :

$$\hat{Q} = \begin{bmatrix} \hat{q}_{20} \\ \hat{q}_{21} \\ \vdots \\ \hat{q}_{65} \end{bmatrix}; \tilde{Q} = \begin{bmatrix} \tilde{q}_{20} \\ \tilde{q}_{21} \\ \vdots \\ \tilde{q}_{65} \end{bmatrix}; W = \begin{bmatrix} w_{20} & 0 & \dots & 0 \\ 0 & w_{21} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & w_{65} \end{bmatrix}$$

Il n'est pas possible de définir de distance du type Khi-deux car la série de valeurs lissées  $\tilde{q}_x$  ne possède pas de forme explicite. La variance expromable des résidus du lissage n'a donc pas de variance.

### b) Le critère de régularité

Le critère de régularité repose sur l'hypothèse selon laquelle la courbe des taux estimés doit être aussi régulière que possible. Il mesure la distance entre les valeurs lissées par la méthode des différences avant :

$$R = \sum_{x=20}^{65-z} (\Delta^z \tilde{q}_x)^2$$

L'opérateur de différenciation avant est définie par récurrence comme étant :

$$\Delta^z q_x = \sum_{j=0}^z \binom{z}{j} (-1)^{z-j} q_{x+j} \quad \text{où } z \text{ fixe le degré du polynôme.}$$

Par exemple en prenant  $z = 1$ , le critère de régularité  $R$  est homologue à la distance euclidienne entre deux valeurs lissées consécutives :  $(\tilde{q}_{x+1} - \tilde{q}_x)^2$ . Pour tenir compte de suffisamment d'informations consécutives,  $z$  est généralement compris entre 2 et 5.

Matriciellement,  $R = (K_z \cdot \tilde{Q})^\top \cdot (K_z \cdot \tilde{Q})$ . La matrice spéciale  $K_z$  est de taille  $(65 - z, 65)$  et ses termes sont les coefficients binomiaux d'ordre  $z$  dont le signe alterne et commence positivement pour  $z$  pair.

$$\text{Par exemple pour } z = 3, K_3 = \begin{bmatrix} -1 & 2 & -1 & 0 & \dots & \dots & 0 \\ 0 & -1 & 2 & -1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & 0 & -1 & 2 & -1 \end{bmatrix} \in \mathcal{M}_{62,65}(\mathbb{R})$$

### c) Ajustement des taux bruts en taux lisses

La moyenne  $M$  de Whittaker-Henderson est une combinaison linéaire des critères de fidélité et de régularité. Un paramètre  $h$  permet d'accentuer ou non le critère de régularité :

$$\begin{aligned} M &= F + h \cdot R \\ &= (\hat{Q}^\top W \hat{Q} - \tilde{Q}^\top W \hat{Q} - \hat{Q}^\top W \tilde{Q} + \tilde{Q}^\top W \tilde{Q}) + h \times \tilde{Q}^\top K_z^\top K_z \tilde{Q} \end{aligned}$$

Les valeurs ajustées  $(\tilde{q}_x)_{20, \dots, 65}$  sont celles qui minimisent  $M$ .  $M$  étant convexe par combinaison linéaire de composantes convexe, minimiser  $M$  revient à résoudre,

$$\begin{aligned} \frac{\delta M}{\delta \tilde{Q}} = 0 &\Leftrightarrow 2W\tilde{Q} - 2W\hat{Q} + h \times 2K_z^\top K_z \tilde{Q} = 0 \\ &\Leftrightarrow \tilde{Q} = (W + hK_z^\top K_z)^{-1}W\hat{Q} \end{aligned}$$

### d) Choix des paramètres de lissage

#### Choix de $z$ :

Le paramètre  $z$  permet uniquement de déterminer le degré du polynôme de la courbe lissée. Son choix s'effectue donc simplement suivant l'allure de la courbe des taux brut. Plus cette dernière présente des pics de variations, plus la valeur de  $z$  sera grande. A noter que le choix de  $z = 1$  aboutira à une courbe lissée affine d'équation  $y = ax + b$ . Afin d'obtenir le meilleur choix possible pour  $z$ , plusieurs valeurs de ce paramètre seront testés. Au vu de l'allure de la courbe des taux bruts, les choix potentiels sont compris entre 2 et 5. La valeur permettant de minimiser l'écart entre la courbe des taux bruts et celle des taux lissés sera alors choisie comme paramètre.

**Choix de  $h$  :**

$h$  représente l'importance accordé au critère de régularité par rapport au critère de fidélité. Le choix de  $h$  présente une part d'intuition non négligeable. Ce choix ne s'appuie pas sur une méthode universelle mais sur des arguments quantitatifs raisonnables.

Prendre  $h = 1$  ne permet pas d'accorder la même importance aux deux critères analysés. En effet les deux phénomènes ne sont pas de la même échelle,  $F$  représente les valeurs d'une fonction mise au carré alors que  $R$  représente le carré d'une dérivée d'ordre  $z$ .

Règle 1 : Verifier l'adéquation post-lissage.

En considérant un nombre fini de candidats  $\mathcal{H} = \{h_1, \dots, h_k\}$  pour lesquels la courbe lissée  $(\tilde{q}'_x(h_j))_{j=1, \dots, k}$  est calculée, le test suivant est effectué en vue de vérifier que la courbe lissée avec un  $h_j$  donnée soit suffisamment en adéquation avec la courbe des données, :

$$\mathcal{H}_0 : q = \tilde{q}(h_j) \text{ contre } \mathcal{H}_1 : q \neq \tilde{q}(h_j)$$

Un candidat est recevable lorsque l'hypothèse  $\mathcal{H}_0$  n'est pas rejetée. Pour tester cette hypothèse, une possibilité est de faire un test du  $\chi^2$ . Il faut pour cela normaliser l'écart quadratique entre les taux bruts et la courbe lissée candidate.

$$D_n = \sum_{x=1}^m \frac{n_x}{\hat{q}_x(1 - \hat{q}_x)} (\hat{q}_x - \tilde{q}(h_j))^2$$

$n_x$  représente l'effectif à l'âge  $x$ .

Si  $\mathcal{H}_0$  est vrai,  $D_n$  est une somme de carré de loi  $\mathcal{N}(0, 1)$ . C'est à dire  $D_n$  est asymptotiquement équivalent à une loi  $\chi^2(k)$  avec  $k$  le nombre de degrés de liberté à déterminer. Il ne s'agit que d'une approximation asymptotique car d'une part les  $\hat{q}_x$  ne sont pas vraiment gaussien et d'autre part le terme de normalisation fait intervenir des estimateurs non paramétriques.

Si  $\mathcal{H}_1$  est vrai, alors la quantité  $D_n$  est "grande" parce qu'il existe au moins un  $x \in 1, \dots, m$  pour lequel l'écart  $\hat{q}_x - \tilde{q}(h_j)$  est significativement non nul. Et donc, cet écart multiplier par  $n_x$  fait croître fortement la valeur de  $D_x$ . En revanche sous  $\mathcal{H}_0$  cette quantité reste dans des limites "raisonnables".

Tout cela aboutit à la forme de test suivant :

Pour un seuil  $S_\alpha$ , le test de niveau  $\alpha$  est défini par :

$$\mathbb{P}(D_n > S_\alpha | q = \tilde{q}(h_j)) \approx \alpha$$

Le seuil  $S_\alpha$  au delà duquel l'hypothèse est rejetée dépend du niveau du test  $\alpha$  qui pour rappel représente la probabilité de se tromper lorsque l'hypothèse nulle  $\mathcal{H}_0$  est vraie.



- Si  $D_n > S_\alpha$ , l'hypothèse  $\mathcal{H}_0$  est rejetée
- Sinon  $\mathcal{H}_0$  n'est pas rejetée et la courbe lissée de paramètre  $h_j$  est recevable.

En général, plusieurs éléments de  $(h_j)_{j=1,\dots,k}$  vont passer le test. A défaut d'avoir une méthode qui sélectionne un seul candidat, cette méthode permet d'éliminer les valeurs de  $h$  les moins optimales.

Le paramètre retenue doit permettre d'une part que la courbe soit visuellement acceptable et d'autre part que la courbe lissée reste stable lorsque le paramètre de lissage varie légèrement. En effet, en situation d'instabilité, une petite variation de  $h$  induit une grande variation du résultat obtenue. Dans ce cas, le résultat devient peut fiable car trop sensible à de faibles variations. C'est pour cela qu'il est nécessaire de faire des analyses de sensibilité avant de conclure.

### Règle 2 : Analyse de sensibilité.

Ici, le but est de savoir comment choisir le meilleur paramètre  $h$  parmi ceux pré-sélectionnés précédemment. Une façon de faire est d'utiliser une technique numérique. L'idée est de simuler un jeu de données artificiel puis vérifier si les résultats sont cohérents. Le jeu de données sera simulé à partir d'une courbe lissée raisonnable  $q(h_0)$ . Il faut donc que  $h_0$  un candidat sérieux et crédible qui a passé les tests précédents. Il est alors supposé que la vraie courbe est définie par cette courbe lissée. Ensuite à partir de cette courbe un grand nombre  $B$  de simulations sont lancées. Simuler une courbe signifie simuler artificiellement une courbe des taux bruts en utilisant leur loi asymptotique.

Pour une courbe lissée  $q(h_0) = (q_{20}, \dots, q_{65})$  auquel est associé le vecteur des effectifs  $(n_{20}, \dots, n_{65})$ , une simulation s'écrit :

$$\tilde{q}(h_0) = (\tilde{q}_{20}, \dots, \tilde{q}_{65}) \quad \text{avec} \quad \forall x \in \llbracket 20, 65 \rrbracket, \quad \tilde{q}_x \sim \mathcal{N} \left( q_x, \frac{q_x(1 - q_x)}{n_x} \right)$$

$B$  courbes des taux bruts simulés sont alors obtenus notés  $\tilde{q}_x^b$  avec  $b$  correspondant la  $b$ ème simulation. Il faut ensuite lancer les procédures de lissage pour chacune des courbes simulées avec chaque candidat de  $\mathcal{H}$ . Comme la courbe à partir de laquelle les simulations ont été faites est connue, il est possible de quantifier à quel point la méthode utilisant le paramètre  $h$  se trompe en calculant par exemple l'erreur quadratique :

$$e(b, h) = \sum_{x=1}^m n_x (\tilde{q}_x^b(h) - q(h_0))^2$$

. Enfin le choix se porte sur le  $h$  qui en moyenne donne le meilleur résultat. C'est à dire celui qui minimise la quantité :

$$\frac{1}{B} \sum_{b=1}^B e(b, h)$$

Bien que le choix des paramètres repose en partie sur l'intuition, la méthode proposée ci-dessus aiguille de manière non négligeable ce choix et a également le mérite de permettre de vérifier la cohérence des résultats.

**e) Résultats des tests de sélection :**

Les méthodes de sélection des paramètres de Whittaker-Henderson sont appliqués numériquement à chaque franchise pour chacun des segments de l'étude (les cadres femmes, les cadres hommes, les non cadres femmes et les non cadres hommes).

Afin de présenter le plus simplement que possible les résultats, ces méthodes de validation du lissage sont mises en pratique ici en considérant les taux d'incidences bruts sur le portefeuille global. Les taux bruts ont été calculés sans distinction de sexe ou de profession. En testant plusieurs couples de paramètres  $(h, z)$ , les courbes lissées obtenues sont présentés ci-dessous.

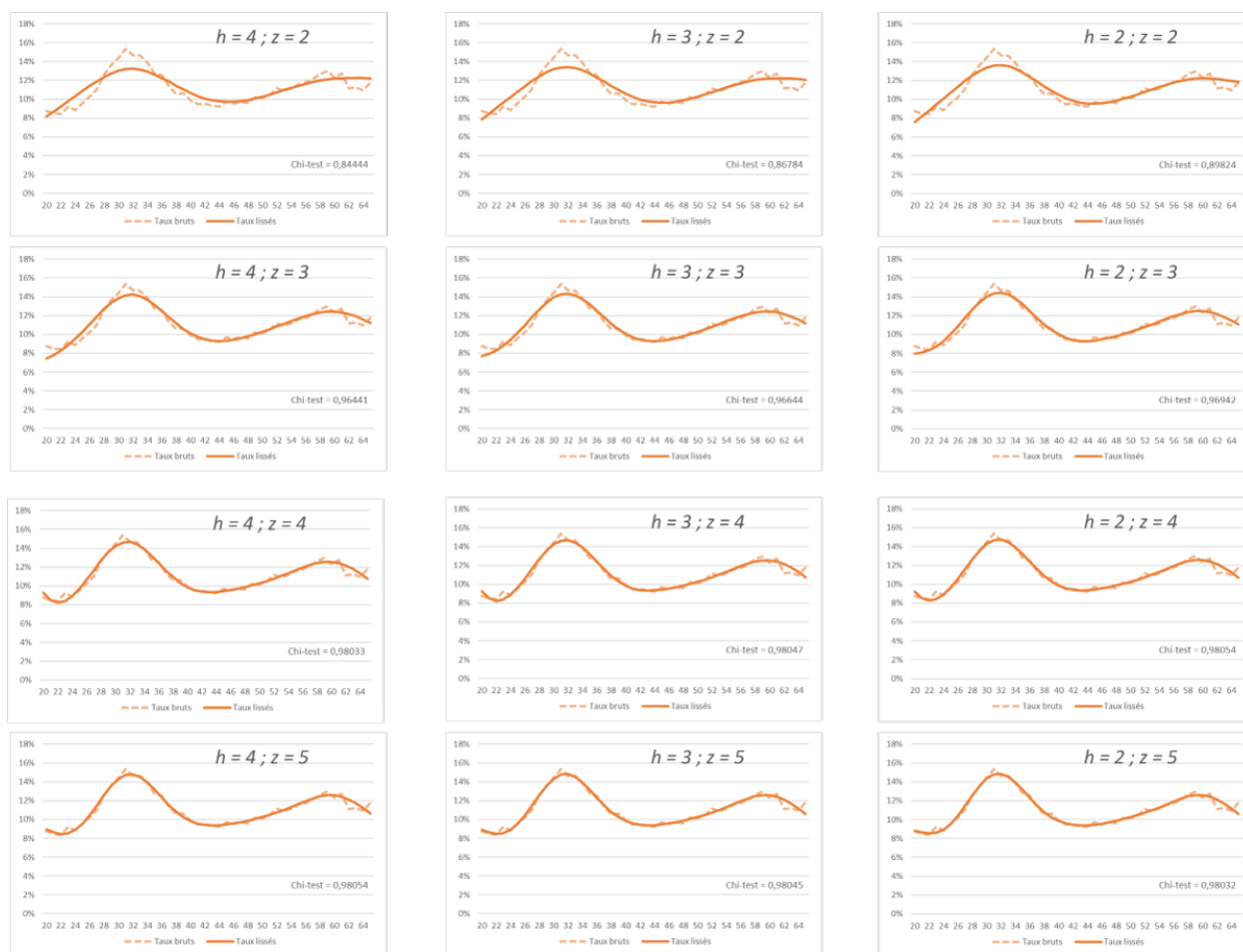


FIGURE 3.8 – Courbe des taux bruts et lissés d'incidence

Le constat est qu'il est difficile avec une simple appréciation visuel de décider quels seront les meilleurs paramètres à choisir. C'est ainsi qu'interviennent les tests présentés précédemment.

Les tests d'adéquations du  $\chi^2$  permettent alors de comparer les couples de paramètres :

		Paramètre h			
		2	3	4	5
Paramètre z	2	0,89824638	0,867846	0,84444799	0,826374610
	3	0,96942972	0,96644139	0,964416357	0,962952483
	4	0,98054192	0,98047267	0,980336	0,980139841
	5	0,98032964	0,98045262	0,9805441	0,980602973

A noter que ce tableau représente les statistiques de tests :

$$p = \mathbb{P}(D_n < S_\alpha | q = \tilde{q}(h_j))$$

Pour un seuil de confiance de 95%, l'hypothèse que la courbe des taux lissée et celle des taux bruts suivent la même loi n'est donc pas rejeté lorsque  $p \geq 0,95$ .

Ensuite, une fois que les couples qui ne sont pas rejetés sont sélectionnés, il faut effectuer les tests de sensibilités aux faibles variations.

La figure suivante présente les 5 premiers tests de sensibilité pour le couple de paramètre ( $h = 3 ; z = 5$ ).

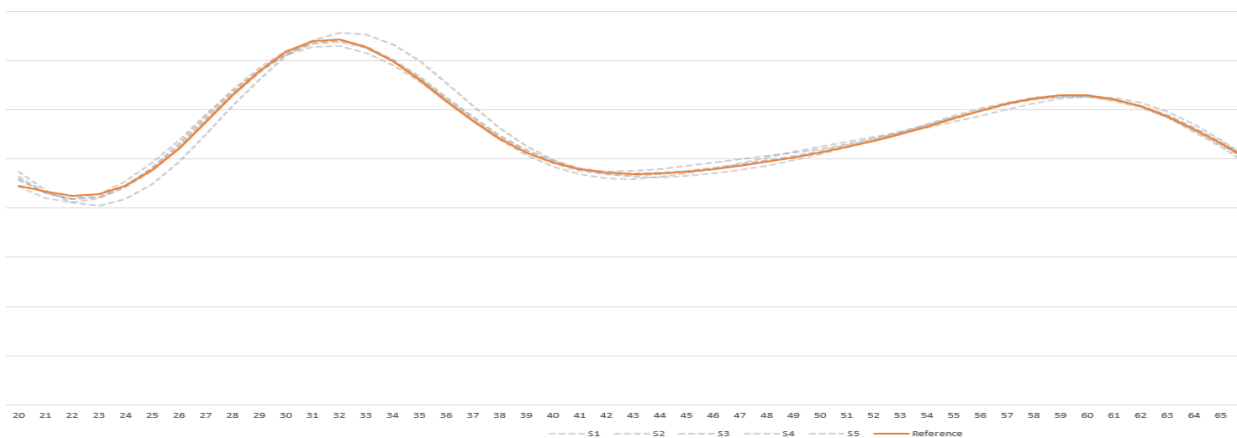


FIGURE 3.9 – Premières simulations du test de sensibilité pour le couple ( $h = 3 ; z = 5$ )

Ces simulations sont faites 1000 fois pour chaque candidat sélectionné. Puis le choix se fait sur le couple qui en moyenne minimise l'écart avec la courbe de référence.

### 3.3.2 Choix et tests de validation

Afin de déterminer cet écart minimal, un test type SMR a été effectué.

Le Standardized Mortality Ratio (SMR, également appelé en français Indice Standardisé de Mortalité, ISM) est le ratio entre le nombre de décès observés et le nombre de décès prévus. Dans le cas de l'étude de l'incapacité temporaire, la notion de décès est remplacée par l'entrée en incapacité. Ce test permet de vérifier que l'écart entre les observations et ceux prévus par les taux calculés précédemment est suffisamment faible. La statistique de test est donnée par :

$$SMR = \frac{\sum_x i_x}{\sum_x e_x \tilde{q}_x}$$

avec :

- $i_x$  le nombre d'arrêts observés à l'âge  $x$
- $e_x$  l'exposition à l'âge  $x$
- $\tilde{q}_x$  le taux d'incidence lissé à l'âge  $x$

Un lissage n'est pas rejeté lorsque la statistique est significativement proche de la valeur 1. L'hypothèse de test est donc :

$$\mathcal{H}_0 : SMR = 1 \text{ contre } \mathcal{H}_1 \neq 1$$

Ce test repose sur l'hypothèse que le nombre d'entrée en arrêt de travail suit une loi Poisson de paramètre  $\sum_x e_x \tilde{q}_x$ . Dans ce cas, en présence d'un nombre d'incapacité suffisamment grand, cette loi peut être approximée par une loi normale  $\mathcal{N}(\sum_x e_x \tilde{q}_x, \sum_x e_x \tilde{q}_x)$ . Ainsi, sous  $\mathcal{H}_0$ ,

$$(SMR - 1) \times \sqrt{\sum_x e_x \tilde{q}_x} \sim \mathcal{N}(0, 1)$$

L'hypothèse nulle est rejetée soit lorsque la p-value est inférieure au niveau de confiance, c'est à dire lorsque la valeur absolue de la statistique de test dépasse le seuil critique associé à  $\alpha$  et à une loi normale centrée réduite.

La méthode de lissage a une dimension de Whittaker Henderson ainsi que les critères de choix du paramètre implémentés informatiquement. Pour chacune des franchises étudiées, les taux bruts sont alors lissés plusieurs fois en utilisant divers paramètres afin de déterminer ceux qui répondent le mieux aux statistiques de choix.

Dans le cas d'une franchise discontinue 30 jours les résultats du lissage sont donnés ci-dessous.

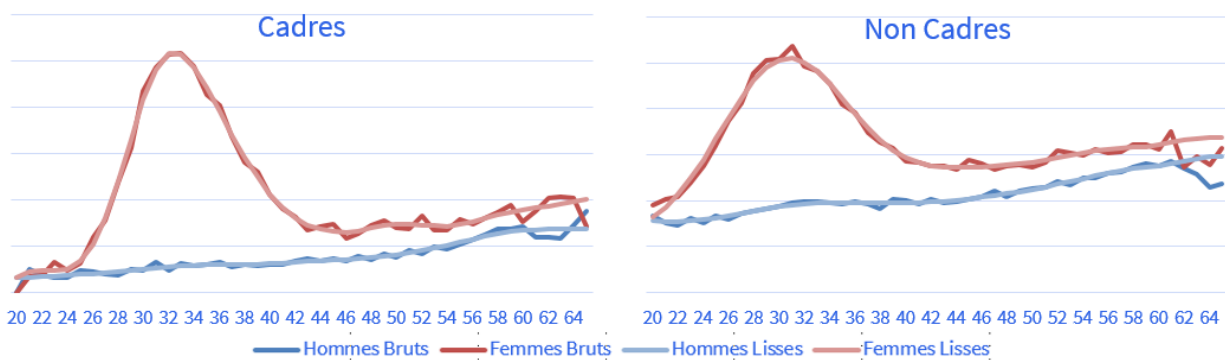


FIGURE 3.10 – Courbe des taux bruts et lissés d’incidence

Afin d’obtenir un tel lissage, les paramètres optimaux choisis sont présentés dans le tableau suivant. Le résultat du test type SMR pour les paramètres choisis est également indiqué à titre indicatif.

Cadres			
	$h$	$z$	Test SMR
Hommes	2	3	1,00015
Femmes	0,5	4	0,99997

Non Cadres			
	$h$	$z$	Test SMR
Hommes	2	3	1,0001
Femmes	1	3	1,0006

## 4 | La table de maintien en incapacité

### 4.1 Idée de conception

#### 4.1.1 Présentation générale

La loi Evin, outre le renforcement des procédures des assureurs, a introduit les tables de maintien en incapacité et en invalidité. Celles-ci ont été établies par le BCAC à partir d'une étude réalisée en 1996 sur un portefeuille rassemblant les plus grandes compagnies d'assurance françaises de l'époque. Ce portefeuille comprenait une cohorte d'invalides et/ou de personnes en incapacité issus des contrats assurés par les sociétés d'assurance. Essentiellement composée de salariés affiliés au régime général de la Sécurité Sociale, la table est représentative de l'ensemble des catégories professionnelles (ouvriers, cadres, ETAM) des entreprises françaises.

A noter que le BCAC a également construit une table de passage de l'état d'incapable à celui d'invalides. Ces tables ne seront pas traitées dans cette étude.

La table de maintien en incapacité présente deux entrées : l'âge à l'événement et la durée de présence en arrêt exprimée en mois. Elle est exprimée en termes de population restante : à l'ancienneté 0, on a 10 000 personnes. Au fur et à mesure que l'ancienneté croît, le nombre de personnes décroît.

Aujourd'hui, les provisions mathématiques pour les arrêts de travail sont encore calculées sur la base des anciennes tables de maintien du BCAC. Celles-ci, construites en 1993 et modifiées pour la dernière fois en 2010 à partir des données de portefeuilles de plusieurs assureurs (dont AXA), ne représentent pas la sinistralité actuelle et ne prennent pas en compte les spécificités du portefeuille d'assurés étudié.

De nouvelles tables ont été construites en 2013 par le BCAC mais ne sont pas utilisées en production pour le calcul des provisions mathématiques.

De plus, dans le cadre de la franchise discontinue, il est nécessaire d'établir une nouvelle table de maintien en incapacité car les durées étudiées diffèrent de celles de la table de maintien du BCAC. En effet, pour construire sa table, le BCAC étudie l'événement lié à la durée d'un arrêt de travail. Dans cette analyse, c'est l'ensemble des arrêts de travail sur une année qui est étudié.

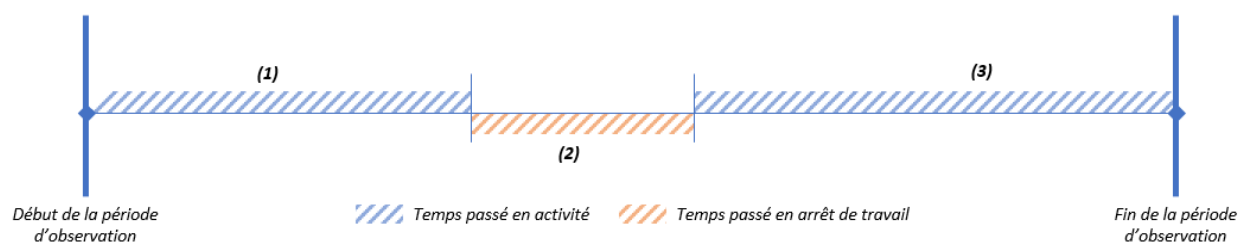
Les objectifs de conception d'une table de maintien en incapacité sont par conséquent les suivants :

- Construire une nouvelle table d'expérience basée sur les données plus récentes du portefeuille afin de représenter au mieux la sinistralité actuelle.
- Tenir compte de la sinistralité engendrée par la présence d'un franchise discontinuée dans les contrats.

### 4.1.2 Les événements observés

Dans le cas de la conception d'une table d'incidence, le risque étudié est l'entrée en incapacité. Autrement dit, l'observation est portée sur la longévité des salariés dans l'état d'activité. A l'opposé, l'étude sur le maintien en incapacité vise à obtenir les taux de sortie en incapacité. Il s'agit alors de considérer initialement un groupe d'individus en arrêt de travail en vue de déterminer leur longévité dans cet état.

Ce propos est illustré par l'observation (sur une période d'un an) du temps passé par un individu dans l'état d'activité et dans l'état d'arrêt de travail. Par souci de simplification, dans cet exemple, le temps passé en activité correspond à la période non indemnisée et celle passée en arrêt de travail à la période d'indemnisation.



Les durées (1) et (3) sont prises en compte dans l'élaboration de la table d'entrée en incapacité alors que la durée (2) est utilisée pour réaliser une table de maintien dans l'état d'incapacité temporaire. Il faut également rappeler de faire attention à la durée de franchise étudiée. En effet, sous une franchise de  $x$  jours, les  $x$  premiers jours d'arrêt de travail sont comptabilisés dans la durée (1) lorsque l'événement d'intérêt est l'entrée en incapacité. Ces jours de franchise sont également comptabilisés dans la durée (2) en vue d'étudier le maintien dans l'état d'incapacité. La franchise n'est pas un élément intervenant dans la conception de la table de maintien.

Tout comme dans la conception de la table d'incidence, les durées étudiées dans le maintien en incapacité sont sujettes aux phénomènes de censures et troncatures. Leurs définitions respectives restent inchangées, seules les causes à l'origine ces phénomènes sont différentes. En bref, sont censurés les individus dont la date de fin ou de début d'au moins un arrêt de travail sur le périmètre d'étude est inconnue. L'avantage d'étudier les données fournies par la DSN est que les sinistres sont donnés au premier jour sans prise en compte de la franchise. Cela permet alors de réduire la surestimation des taux de sorties due aux observations tronquées.

## 4.2 Elaboration de la table

Comme expliqué précédemment, une table de maintien est représentée sous forme d'un tableau à double entrée où pour chaque tranche d'âge est exprimée la survie des individus en arrêt de travail temporaire à différentes anciennetés dans l'état. Dans le cas du risque incapacité temporaire, pour une franchise continue, les anciennetés considérées sont de 0 à 36 mois (durée maximale de présence dans l'état). Dans le cadre de l'étude d'une franchise discontinue, conformément à la définition donnée, les durées maximales étudiées sont d'un an. Les anciennetés étudiées sont donc de 0 à 365 jours ici.

La base de données de la DSN dispose d'un volume de données assez important ; il est possible d'associer des individus à chaque ancienneté entre 0 et 365 jours. La table de maintien peut alors être déterminée jour par jour ce qui permet une plus grande justesse dans le calcul des coûts (exprimés au chapitre 5). Ce choix est notamment motivé par le fait que les anciennetés entre 1 et 30 jours (correspondant aux arrêts de travaux courts) sont particulièrement représentés dans le portefeuille. En regroupant mensuellement ces arrêts courts, il y a une perte d'information quant à l'allure de la courbe de maintien.

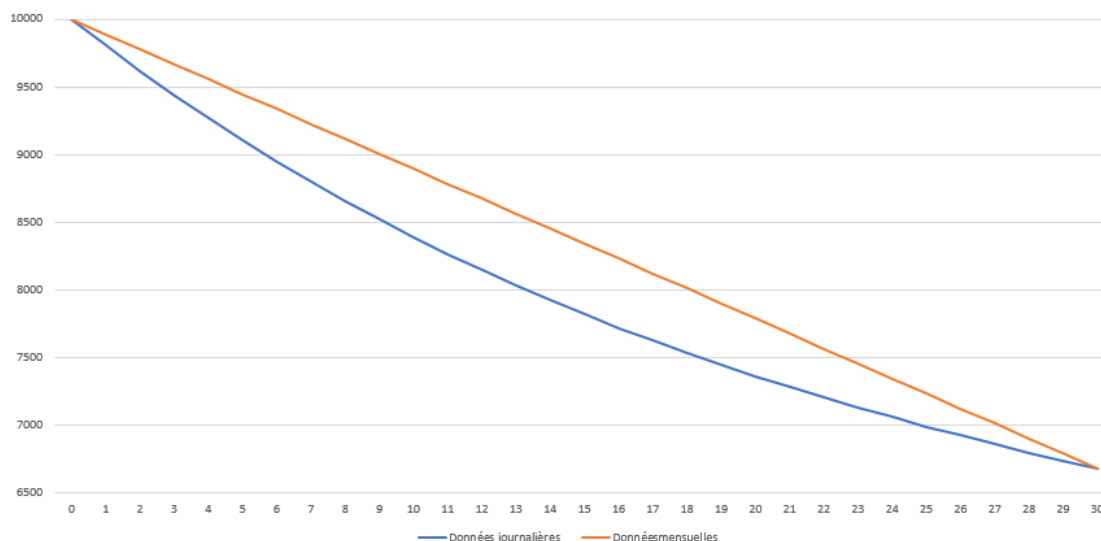


FIGURE 4.1 – Allure de la courbe de maintien globale (franchise continue) faites à partir de données mensuelles et journalières

La figure 4.1, présente dans le cadre de la franchise continue, sans distinction d'âge, l'allure de la courbe de maintien en arrêt de travail lorsqu'elle est construite à partir de données journalières (en bleu) ou à partir de données mensuelles (en orange). Dans le cas des données mensuelles, une régression linéaire a permis de reconstituer les informations intermédiaires. En effet, lorsqu'il n'est possible de disposer de données journalières, la méthode couramment utilisée pour obtenir les informations entre deux mois est une régression linéaire.



Cette figure montre bien que l'allure de la courbe selon le pas d'ancienneté utilisé est différente. Plus ce pas sera petit, plus la table de maintien sera représentative de la réalité du portefeuille. Il faut néanmoins bien faire attention à disposer de suffisamment de données pour accepter de réduire le pas. La plupart des méthodes d'estimation nécessitent un "grand" volume de données afin de réduire au mieux la volatilité des résultats. Bien que le terme "grand" reste subjectivement défini, il faut s'assurer avant d'échantillonner la base en segments plus petits que la volatilité alors engendrée reste acceptable en terme de risque.

La table de maintien est ensuite construite. Chaque cellule du tableau correspond à la fonction de survie  $S_x$  de la variable aléatoire  $T_x$  représentant pour un âge  $x$  la durée passé en arrêt de travail sur une année glissante.

Âge \ Période	0	1	...	365
20	10 000	$S_{20}(1)$	...	$S_{20}(365)$
...	10 000	...	...	...
65	10 000	$S_{65}(1)$	...	$S_{65}(365)$

Afin d'aboutir à une telle table, il faut d'abord étudier pour un âge  $x$  donné, les taux conditionnels de sortie de l'état d'incapacité temporaire notés  $q_t$  avec  $t$  l'ancienneté dans l'état. Il s'agit de taux de risque instantanés.

$$q_t = \mathbb{P}(T_x \leq t + 1 | T_x > t) = 1 - \frac{S_x(t + 1)}{S_x(t)}$$

### 4.2.1 Estimation des taux bruts

Tout comme dans le cas de la table d'incidence, il faut dans un premier temps estimer les taux bruts décrits ci-dessus. Seules les durées étudiées sont différentes. Elles ne portent plus sur l'entrée en incapacité mais sur la sortie d'incapacité.

Selon la littérature, la méthode la plus robuste de construction des taux bruts pour les lois de maintien est actuellement l'estimateur de Kaplan-Meier. Cette méthode consiste pour chaque ancienneté considérée à calculer les taux bruts de chaque âge, puis à rassembler toutes ces tables en un tableau à double entrée âge/ancienneté.

L'utilisation d'une méthode uni-dimensionnelle pour traiter d'informations bi-dimensionnelles (comme dans le cas de la table de maintien), interroge sur la perte d'information possible. En effet, ne pas prendre en compte de la loi conjointe selon les deux dimensions du problème est une source de biais. En pratique, la méthode de Kaplan Meier a l'avantage d'être peu pénalisante à ce sujet.

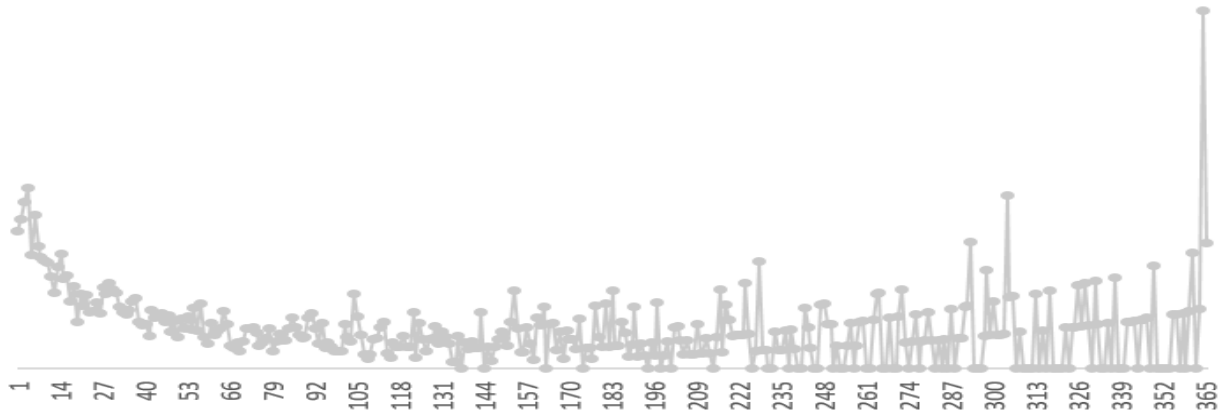
C'est pour ces raisons, qu'il a été choisi d'estimer les taux bruts de maintien dans l'état d'incapacité par la méthode de Kaplan-Meier comme mentionné ci-dessus.

A noter, qu'il existe des approches proposant des modèles naturellement bi-dimensionnels. C'est le cas du modèle de Lee-Carter [1992] ou des modèles de type géostatistique de DEBON et al. [2004]. Ces modèles initialement développés dans le contexte de la détermination de tables de mortalité prospectives peuvent être facilement adaptés au cadre de l'arrêt de travail. Cependant, une étude menée par LELIEUR [2005] montre une instabilité dans les estimations lorsque le modèle ne se base pas sur de très grands échantillons. Bien que LELIEUR propose un moyen de limiter cette difficulté dans le cas des tables prospectives de mortalité, la transposition au cadre de l'arrêt de travail proposée par WINTER [2005] est moins satisfaisante du fait de la forme particulière des taux de sortie pour ce risque.

A l'issue de la phase d'estimation, au lieu de calculer un unique taux d'entrée pour chaque âge comme dans le cas de la table d'incidence, est alors obtenu pour chaque âge un taux de sortie pour chaque période d'ancienneté dans l'état. En effet, pour rappel, le but de la table d'incidence est de déterminer la probabilité qu'un individu selon son âge bénéficie d'une indemnisation d'au moins un jour d'arrêt de travail ; alors que le but de la table de maintien, est pour chaque âge et chaque mois passé en arrêt de travail (durée cumulée) de pouvoir déterminer un taux de sortie en incapacité.

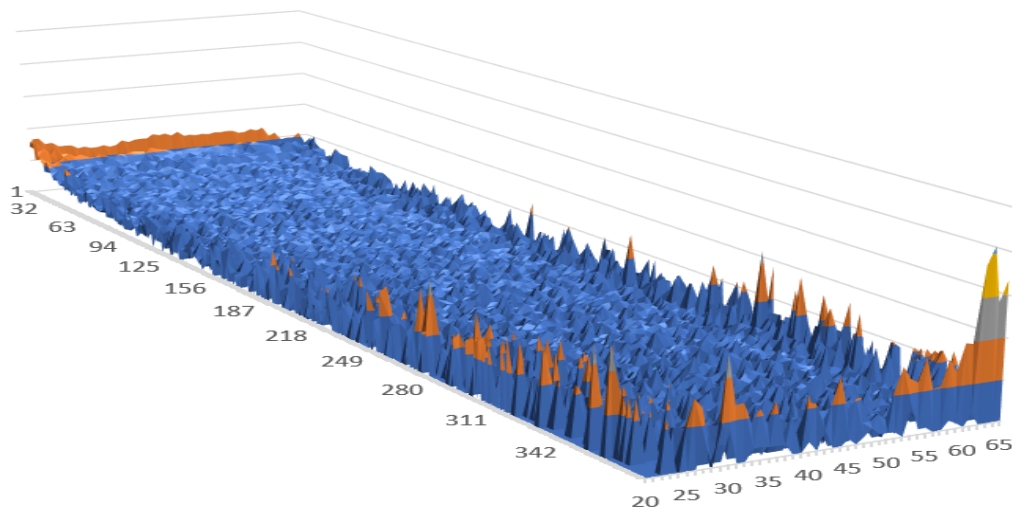
Ainsi pour chaque tranche d'âge  $x$ , il est calculé 365 taux de sortie correspondant à chaque ancienneté étudiée. Ces taux sont notés  $Q_x = \{q_1, \dots, q_{360}\}$ .

En fixant l'âge d'observation à 30 ans par exemple, la courbe des taux bruts de sortie à cette âge  $Q_{30}$  est représentée ci dessous. L'allure de cette courbe est plus ou moins similaire quelque soit l'âge  $x$  étudié. Cette courbe présente quelques points nuls causés par le manque d'informations associés au point d'abscisse considéré.



La tendance observée est d’abord un taux de sortie assez élevé en début de courbe. Cela montre une forte présence de durées annuelles en arrêt de travail plutôt courtes. Ensuite il faut faire attention avant d’analyser le pic de sortie associé grandes durées. Pour ces durées, le nombre de salariés observés est très faible. Cela se voit par la quantité d’informations manquantes qui s’amplifie en fin de courbe. De plus la méthode de Kaplan Meier est très sensible au volume de données utilisées. Il n’est donc pas immédiat de conclure à un fort pic de sorties après un an d’arrêt de travail.

Il est également possible de réaliser une représentation 3D de l’ensemble des résultats obtenus. Sur un tel graphe sont représentés les différents taux bruts de sortie selon l’âge et l’ancienneté. Il s’agit simplement d’une représentation graphique de la matrice  $[Q_{20}, \dots, Q_{65}]$ .



Comme dans le cas  $x = 30$  ans, les taux sont d'abord élevés pour les petites anciennetés. Quelque soit l'âge des individus, de nombreux individus ont une durée d'absence annuelle faible. Ensuite, le taux de sortie aux grandes anciennetés est particulièrement élevé pour les grands âges. Ce phénomène est causé par deux facteurs :

- Aux départs à la retraite. Les causes de sorties en arrêts de travail n'étant pas traités à part dans cette étude, il n'est pas possible d'identifier ici les salariés prenant leur retraite suite à un arrêt de travail.
- Aussi, les taux bruts calculés par la méthode de Kaplan Meier sont très erratiques aux âges limites. Cela est dû au volume de données qui est moins important à ces âges.

Aussi, la durée annuelle des arrêts de travail chez les jeunes excède moins fréquemment un an que chez le reste de la population. En effet, les taux de sorties inférieurs à un an sont généralement plus élevés pour cette sous-population.

## 4.2.2 Lissage de Whittaker-Henderson

Tout comme dans le cas de l'estimation des taux bruts, les méthodes statistiques couramment utilisées pour l'ajustement paramétrique des lois de maintien sont uni-dimensionnelles. Il est possible de citer par exemple les méthodes type Makeham, logit ou splines. Il existe néanmoins pour certaines de ces méthodes des extensions bi-dimensionnelles.

La méthode de lissage utilisée ici est une extension du lissage de Whittaker-Henderson en 2 dimensions. Cette méthode suppose que les taux considérés ne dépendent plus uniquement de l'âge  $x$  des individus mais également de leur ancienneté  $i$  en arrêt de travail. Ainsi sont étudiés les estimations  $\hat{q} = (\hat{q}_{x,i})_{20 \leq x \leq 65, 1 \leq i \leq 365}$  définis comme précédemment. Matriciellement cela donne :

$$\hat{Q} = \begin{bmatrix} \hat{q}_{20,1} & \hat{q}_{20,2} & \cdots & \hat{q}_{20,365} \\ \hat{q}_{21,1} & \hat{q}_{21,2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \hat{q}_{64,365} \\ \hat{q}_{65,1} & \cdots & \hat{q}_{65,364} & \hat{q}_{65,365} \end{bmatrix}$$

Le critère de fidélité s'écrit alors :

$$F = \sum_{x=20}^{65} \sum_{i=1}^{365} w_{x,i} (\tilde{q}_{x,i} - \hat{q}_{x,i})^2$$

Le critère de régularité dans le cas en deux dimensions suppose de distinguer la régularité verticale et la régularité horizontale qui elle agit sur les  $q_{x,i}$  à  $x$  fixé. Pour ce faire, sont introduits les opérateurs  $\Delta_v$  et  $\Delta_h$  qui permettent d'agir sur les  $q_{x,i}$  respectivement à  $i$  et  $x$  fixés. Ainsi les indices de régularité vertical et horizontal sont définis respectivement par :

$$R_v = \sum_{i=1}^{365} \sum_{x=1}^{65-z} (\Delta_v^z \tilde{q}_{x,i})^2 \text{ et } R_h = \sum_{i=1}^{365} \sum_{x=1}^{65-z} (\Delta_h^z \tilde{q}_{x,i})^2$$

La moyenne  $M$  de Whittaker-Henderson s'écrit alors :

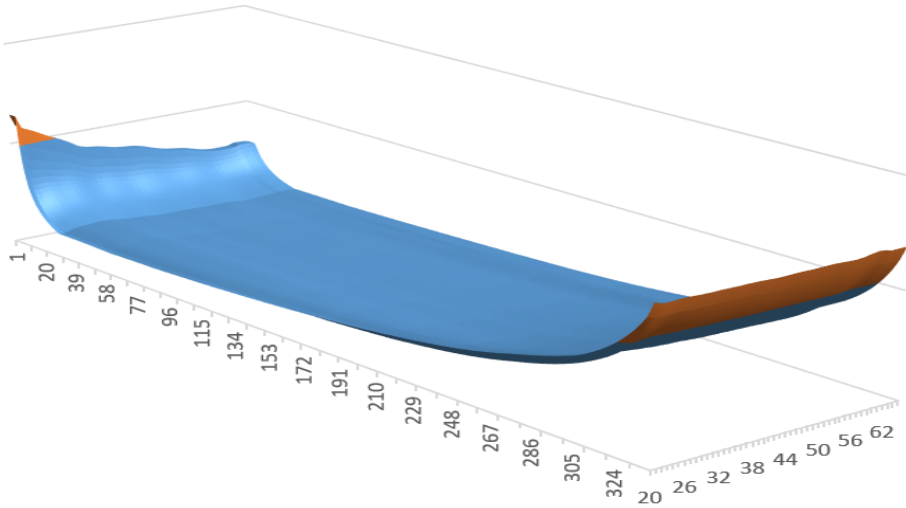
$$M = F + \alpha \times R_v + \beta \times R_h$$

Les paramètres à définir dans ce cas sont  $(z, h, \alpha, \beta)$ .

Finalement, en se basant sur les mêmes notations matricielles établies dans le lissage des taux d'incidence, la minimisation de  $M$  permet d'obtenir les taux lissés suivants :

$$\tilde{Q} = (W + \alpha K_z^{v\top} K_z^v + \beta K_z^{h\top} K_z^h)^{-1} W \hat{Q}$$

L'implémentation et les tests de lissages faits à partir de cette méthode ne sont pas étudiés dans le cadre de cette étude. Le résultat final obtenu est représenté graphiquement ci-dessous.



### 4.2.3 Validation et résultats

Une fois les taux de sortis ajustés sont obtenus, il est alors possible d'établir la table de maintien recherchée. Il s'agit d'appliquer la formule établie dans les notions clés de modèle de durée et permettant une relation entre la fonction de survie et les taux de risque instantanés. Pour chaque âge  $x$ ,

$$S_{x,i} = \prod_{i|t-i < t} (1 - \tilde{q}_{x,i})$$

La table de maintien lissée obtenue peut ensuite être comparée aux tables déjà existantes du BCAC établies dans le cas continue. Pour plus de visibilité, le graphique ci-dessous montre une moyenne des taux par tranche d'âge. L'observation est faite sur douze mois car au-delà la notion de franchise discontinue n'a plus de sens.

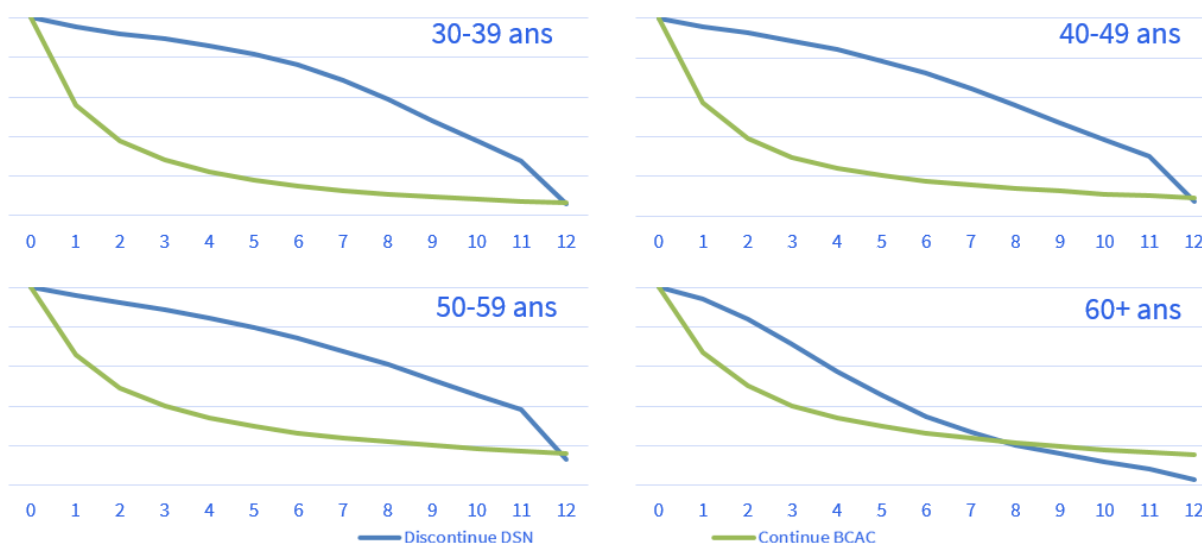


FIGURE 4.2 – Table de maintien sur 12 mois pour la franchise continue (BCAC) et discontinue

Globalement les taux de maintien en discontinue sont bien plus élevés. Cette observation est souhaitée dans la mesure où les durées d'absence observées ici sont plus longues que celle d'une étude faite sur le maintien d'arrêts indépendants. Aussi il est logique d'observer un rapprochement des courbes pour les grandes anciennetés. En effet, un arrêt de travail discontinu de douze mois est équivalent à un arrêt continu de douze mois. Enfin, plus l'âge étudié est élevé, plus l'écart entre les deux courbes diminue. Cela s'explique par la tendance des plus jeunes à avoir des arrêts de travaux plus courts et plus fréquents. La durée des arrêts en cumulés sur l'année est alors plus longue que lorsque les arrêts sont étudiés individuellement et indépendamment les uns des autres. A contrario, aux grands âges, les arrêts étant plutôt longs et peu fréquents sur une année, la durée du cumul des arrêts tend à se rapprocher de la durée d'un seul l'arrêt.

Il faut cependant noter que pour les grands âges et/ou pour les grandes anciennetés, le manque de données impacte les queues de distributions.

La table de maintien peut également être représentée entièrement via le graphique en plan suivant. La représentation est faite uniquement pour les 100 premiers jours en vue de plus de visibilité graphique.

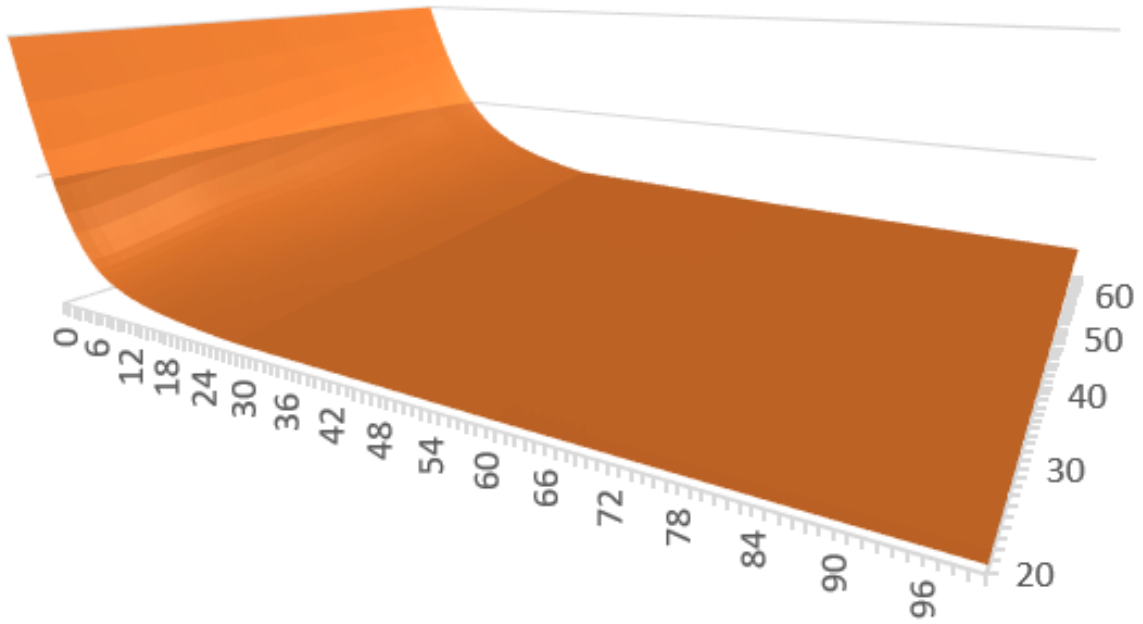


FIGURE 4.3 – Table de maintien sur 100 jours en franchise discontinue

## 5 | La tarification

### 5.1 Méthode de tarification "Best-Estimate"

#### 5.1.1 Tarification d'une tête

La vison "Best-Estimate" est obtenue en utilisant une méthode "Coût x Fréquence".

La fréquence est tout simplement obtenue en lisant la table d'incidence. Pour un âge donné  $x$  et une durée de franchise  $f$ , l'occurrence de l'événement "avoir un arrêt de travail indemnisé" est de probabilité donnée par la table d'incidence associée à la franchise  $f$  et pour la ligne correspondant à l'âge  $x$ .

Le coût n'est lui pas directement obtenu par simple lecture d'une table. L'évaluer revient à déterminer une annuité à partir de la table de maintien élaborée. En effet, l'annuité se définit comme la valeur actuelle d'une série de flux futurs. Dans le cas de l'étude de l'incapacité temporaire, les flux futurs concernés sont les indemnités journalières que recevront les assurés sur la durée estimée de leurs absences. L'annuité pure, c'est à dire sans tenir compte des frais de gestion ou autres frais, ce calcul pour un âge  $x$  et une franchise  $f$  de la façon suivante :

$$a_{(x,f)} = \sum_{j=1}^{1095} \frac{{}_j p_x}{(1+i)^{j+f}} \times \frac{1}{f p_x}$$

avec

- $x$  l'âge étudié
- $i$  le taux d'actualisation journalier. Pour un taux annuel de 3%, ce dernier est environ égal à 0,0082%.
- ${}_j p_x$  la probabilité de rester en incapacité entre l'âge  $x$  et  $x + j$
- 1095 correspond à une durée de 3 ans en jours. C'est la durée maximale d'indemnisation dans le cas d'une garantie incapacité temporaire.

A partir de cette formule, il est possible de calculer pour tous les âges, la valeur estimée de l'annuité. Le graphe ci-dessous représente l'évolution de l'annuité selon l'âge. Les résultats déterminés à l'aide de la table de maintien élaborée dans le cadre de la franchise discontinue sont également comparés aux annuités basées sur la table de maintien du BCAC qui est effectué dans un cadre continu.



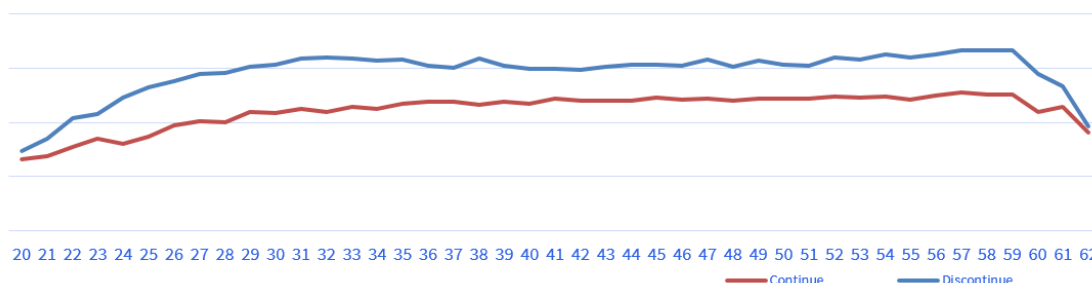


FIGURE 5.1 – Annuité de la franchise 30 jours

Il est évidemment observé qu'en discontinue les annuités sont plus importantes qu'en continue. La raison est encore une fois le fait que les durées en discontinues sont plus longues. Aux grands âges, comme déjà expliqué précédemment, les arrêts sont le plus souvent uniques. C'est pour cela que les annuités sont proches.

Il est intéressant de noter que le taux d'actualisation dans le cas d'une garantie incapacité temporaire a très peu d'impact sur les annuités comme le montre le graphique ci-dessous. Cela s'explique par le fait que l'indemnisation de la garantie incapacité temporaire a une durée de vie maximale de 3 ans. Passée cette période, l'assuré n'est plus considéré comme en incapacité mais comme un invalide. Lorsque le taux est inférieur ou égal à 3%, une actualisation sur 3 ans ne permet pas d'impacter significativement l'annuité.

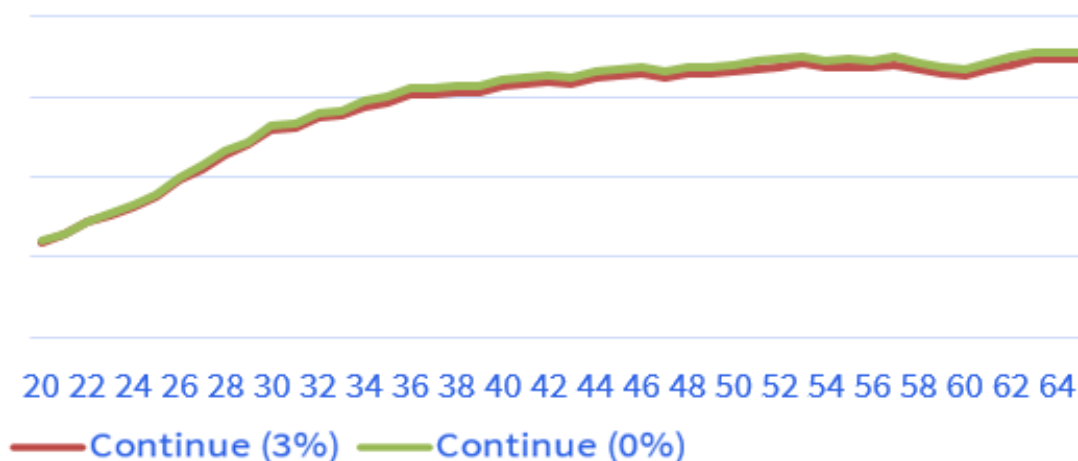


FIGURE 5.2 – Impact du taux d'actualisation sur l'annuité

Ainsi, à partir des deux tables élaborées précédemment, le tarif de l'arrêt de travail selon l'âge a l'allure suivante :

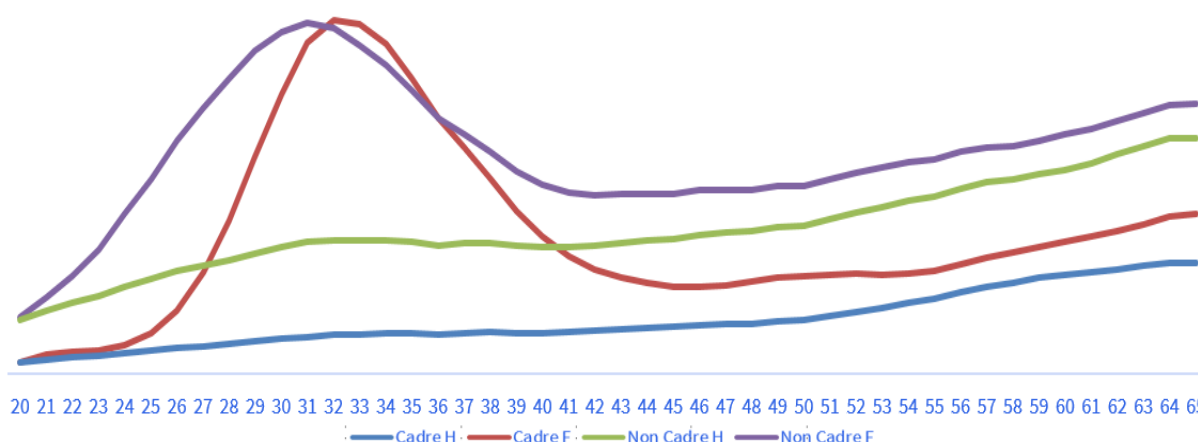


FIGURE 5.3 – Tarification de la garantie incapacité temporaire franchise 30 jours discontinue

Le tarif est basé sur des tables d'incidence et de maintien construites à partir des DSN. Cependant la déclaration sociale nominative recense quelque soit la durée de franchise appliquée au contrat, l'ensemble des arrêts de travail des bénéficiaires au premier jour. Il faut donc faire attention au biais issu du phénomène d'aléa moral. En effet, le comportement des assurés selon la durée de franchise de leur contrat peut être différent. Un assuré atteignant rapidement la fin de sa période sous franchise, sachant ses futures absences indemnisées par l'assureur pour le reste de l'année, pourrait avoir tendance à s'arrêter plus facilement et fréquemment.

## 5.1.2 Correctif âge actuariel

### a) Problématique et étude

Le tarif déterminé précédemment est soit celui d'un unique individu selon son âge, sexe et collègue soit celui d'un groupe d'individus tous du même âge. Cependant une des difficultés de la tarification en assurance collective provient du fait que l'assureur ne dispose généralement pas des informations individuelles de chacun des salariés mais plutôt à une démographie moyenne des effectifs. Ainsi la tarification de l'incapacité temporaire au sein d'un établissement s'effectue dans la majorité des cas à partir de l'âge moyen de ses salariés. Cela peut poser problème dans la mesure où l'âge moyen ne tient pas compte de la dispersion de l'effectif autour de cet âge moyen.

Par exemple, en considérant deux entreprises composées chacune de 2 salariés dont l'âge moyen est 30 ans telles que :

- L'entreprise n° 1 est constituée de deux femmes cadres ayant toutes deux 30 ans ;
- L'entreprise n° 2 est constituée de deux femmes cadres ayant 20 ans pour l'une et 40 ans pour l'autre.

Au vu des graphiques ci-dessus, le tarif calculé pour la société n° 1 sera supérieure à celle de la société n° 2. Ce propos est illustré ci-dessus en se basant sur la courbe de tarif des femmes non cadres pour une franchise discontinue de 30 jours. Les "Tarif 1" et "Tarif 2" représentent respectivement les prix calculés pour l'entreprise n° 1 et l'entreprise n° 2.

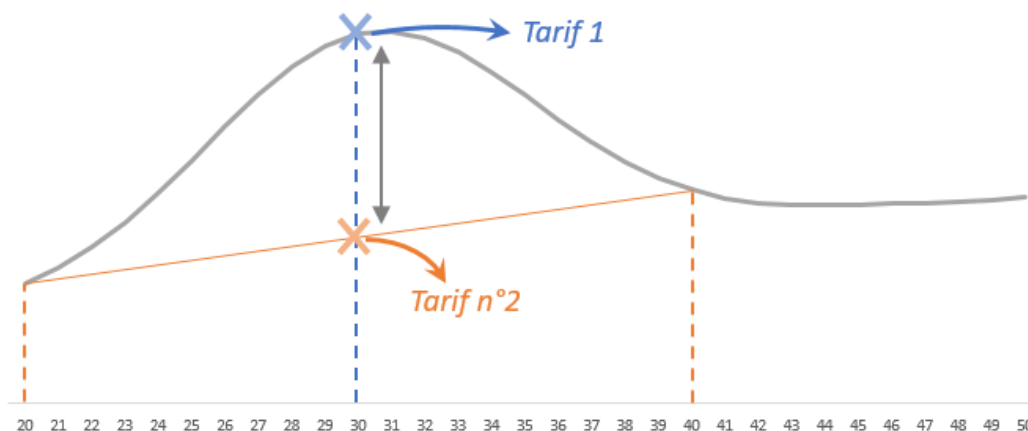


FIGURE 5.4 – Exemple de tarification des entreprises n° 1 et n° 2

Bien que les deux entités présentent un effectif d'âge moyen 30 ans, le calcul du tarif salarié par salarié montre qu'en moyenne le risque est moins élevé pour la deuxième entité. Or lorsque seul l'âge moyen des salariés est disponible, la tarification des deux sociétés aboutit à un même prix : "Tarif 1". Dans le cas de l'entreprise n° 2 le tarif est beaucoup plus élevé que le coût obtenu si l'information sur l'âge de chacun des salariés est disponible.

Cela prouve bien qu'il peut y avoir une différence entre "déterminer l'âge moyen d'une entreprise puis tarifier à partir de cet âge moyen" et "déterminer le coût de chacun des salariés et moyenner ces coûts pour obtenir un tarif".

Le but du correctif âge actuariel est alors de rectifier le coût établi à partir des caractéristiques moyennes de l'entreprise afin qu'il puisse tenir compte au mieux de la dispersion des effectifs. Pour ce faire, l'âge moyen des salariés est modifié de façon à ce que le tarif calculé en à partir du nouvel âge moyen se rapproche au mieux du tarif qui aurait été obtenu s'il avait été possible d'évaluer le coût de chacun des salariés et d'en faire une moyenne. La DSN a l'avantage de fournir pour chaque entité du portefeuille des informations d'âge de sexe et de collègue pour chaque salarié de ladite entité. Il est donc possible d'une part d'établir un tarif individuel pour chacun des individus d'une entreprise puis d'autre part de comparer la moyenne de ces coûts avec le tarif calculé uniquement avec la démographie moyenne de la compagnie.

L'étude menée est résumée ci dessous :

**Calcul du tarif tête par tête (*salarié par salarié*) :** Le coût de chaque individu de la base de donnée d'étude est d'abord calculé en fonction de son sexe, âge et collègue suivant la méthode "coût x fréquence". Ensuite les salariés appartenant au même Régime Professionnel de Prévoyance (RPP) d'une entreprise sont regroupés. Le tarif final par RPP est alors calculé en moyennant ces coûts individuels de ses adhérents. A noter que pour plus de précision, pour chaque RPP ce coût final est segmenté suivant le sexe et la CSP des salariés. Dans une vision "Best-Estimate", cette méthode est vivement recommandée. Cependant elle nécessite la connaissance de chacun des salariés associés au contrat étudié ; or ces données ne sont pas forcément accessibles à l'assureur.

**Calcul du tarif démographie moyenne :** A chaque RPP est associé l'âge moyen de ses adhérents en prenant soin de segmenter les individus par sexe et CSP. Puis le tarif "coût x fréquence" est calculé a partir de ces informations moyennes. Ce tarif alors obtenu est équivalent à considérer soit que tous les salariés d'une même entité ont le même âge qui correspond à l'âge moyen de la structure, soit l'entité est composée d'un seul salarié dont dans ce cas est également l'âge moyen. Autrement dit, ce tarif ne tient donc pas compte de la dispersion des salariés autour de l'âge moyen contrairement au tarif tête par tête.

**Estimation du correctif d'âge :** L'objectif est de modifier l'âge d'incidence permettant que le tarif démographie moyenne se rapproche au mieux du tarif tête par tête. Par soucis de simplicité, seul le paramètre "âge d'incidence" variera. L'annuité retenue est celle associée à l'âge moyen des salariés du RPP considéré. Il s'agit alors de déterminer l'âge d'incidence qui permet de minimiser l'écart entre le tarif global estimé tête par tête et le tarif calculé à partir de l'âge moyen.

## b) Résultats et interprétation

Dans cette première figure sont représentés d'une part les résultats de la tarification tête par tête (en pointillé) et d'autre part celle faite démographie moyenne selon l'âge moyen du RPP étudié. Considérer la dispersion des effectifs autour de l'âge moyen, comme c'est le cas dans le tarif tête par tête, permet de réduire considérablement le pic de la courbe du tarif fait pour un seul salarié.

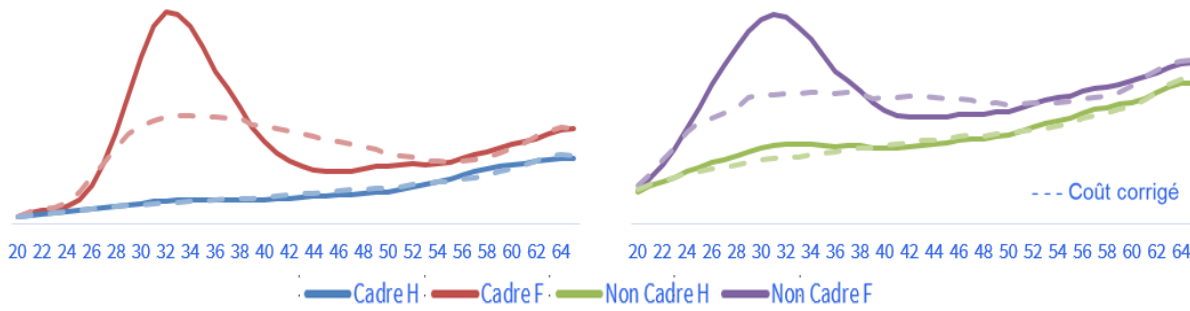


FIGURE 5.5 – Comparaison entre le tarif tête par tête et et le tarif corrigé démographie moyenne selon le sexe et la CSP de l’assuré

Lorsque l’âge moyen d’un contrat est autour de 40 ans, le tarif tête par tête est plus élevé que celui fait sur une groupe d’assurés tous âgés de 40 ans. La raison est qu’il faut prendre en compte la part non négligeable des salariés plus jeunes et plus âgés pour lesquels le tarif est plus élevé. De même pour un âge moyen de 30 ans, ce tarif sera plus faible car le contrat concerne des individus dont l’âge requiert un tarif moindre que celui des salariés ayant 30 ans.

Ce correctif ne concerne peu voire pas du tout les RPP d’âge moyen extrême, c’est à dire ceux proches de 20 ans et de 65 ans, car à ces âges moyens, les effectifs des contrats sont très faibles et souvent les individus ont tous le même âge. La dispersion des individus associés aux contrats ayant un très grand ou très faible âge moyen est très faible et donc le tarif fait sur une tête (ou sur un groupe d’individus tous du même âge) est très proche de celui fait en tenant compte de la dispersion des effectifs.

Le graphique ci-dessous illustre ce phénomène en indiquant sur les lignes les différents âges moyens des RPP du portefeuille et en colonne les âges de chacun des individus des RPP considérés. Ainsi, il est possible de voir la dispersion des individus autour de l’âge moyen.

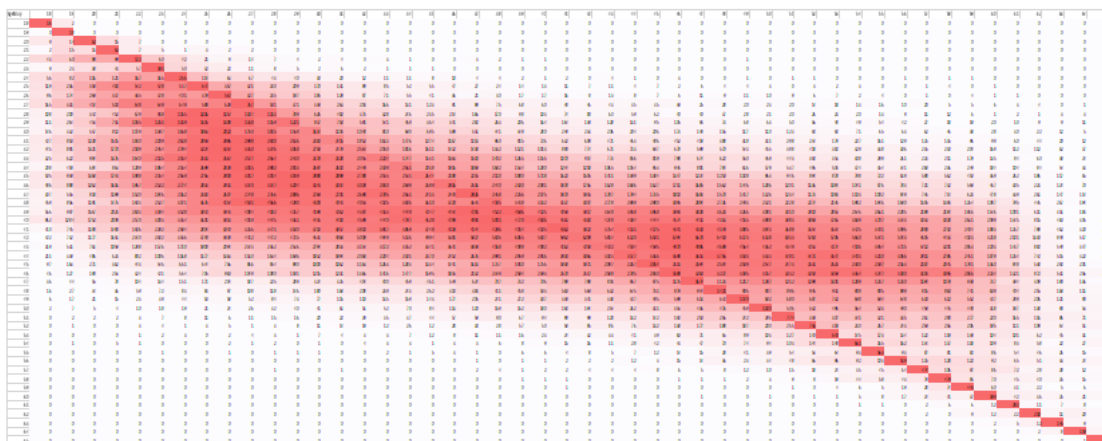


FIGURE 5.6 – Dispersion des salariés selon leur âge (colonnes) et en fonction de l’âge moyen du contrat auquel ils ont souscrit

Lorsque le RPP est d'âge moyen très jeune, il n'y a que très peu de dispersion. Pour un régime ayant une moyenne d'âge inférieure à 40 ans, l'âge des salariés excède très peu la moyenne observée. Un correctif d'âge actuariel dans ce cas donnera un poids plus important aux coûts des plus jeunes. Dans le cas des âges moyens très vieux, la dispersion est quasiment nulle et donc le tarif tête par tête et celui démographie moyenne est quasiment la même. Cependant pour une moyenne d'âge tournant autour de 40 ans, il est observé une dispersion plutôt homogène des effectifs. Dans ce cas, le poids de chaque âge sera plus ou moins le même dans le calcul du tarif à l'âge moyen.

Une fois le correctif de coût établi, le but est de trouver quel âge faut-il appliquer à la table d'incidence afin que le tarif " coût x fréquence " coïncide avec le coût corrigé. Le problème à résoudre peut s'écrire mathématiquement :

$$i_0^x = \operatorname{argmin}_{i=20,\dots,65} [\text{incidence}_i \times \text{annuite}_x]$$

avec

- $x$  l'âge moyen de du contrat étudié ;
- $i_0^x$  l'âge actuariel recherché pour une démographie d'âge moyen  $x$  ;
- $\text{incidence}_i$  le coefficient de la table d'incidence associé à l'âge  $i$  ;
- $\text{annuite}_x$  le coefficient de la table d'annuité associé à l'âge  $x$ .

Cet exercice permet de déterminer les résultats suivants dans le cadre de la franchise discontinuée 30 jours :

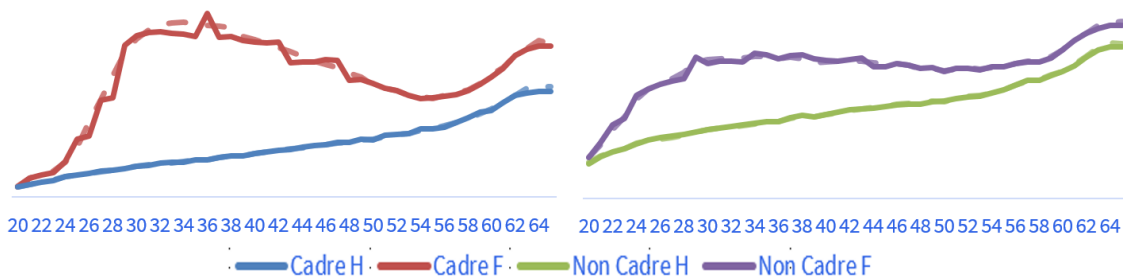


FIGURE 5.7 – Comparaison entre le tarif corrigé démographie moyenne et le tarif sous correctif âge actuariel selon le sexe et la CSP de l'assuré

Ainsi, la même comparaison faite à la figure 5.5 est possible mais cette fois-ci avec l'application du correctif âge actuariel.

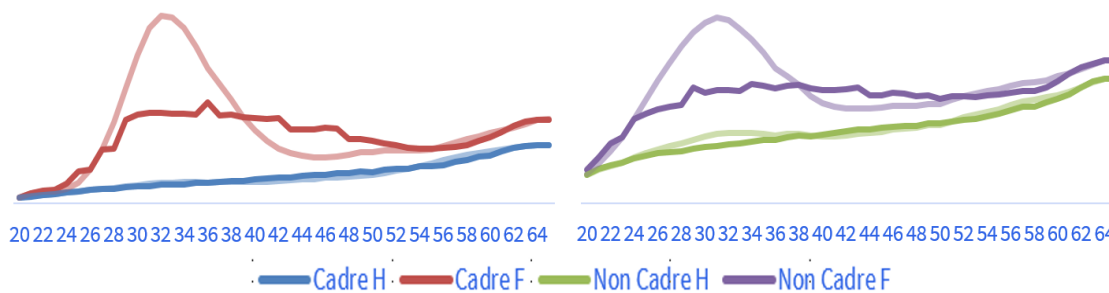


FIGURE 5.8 – Comparaison entre le tarif tête par tête et le tarif sous correctif âge actuariel selon le sexe et la CSP de l'assuré

Les nouvelles courbes de tarif obtenues ne sont alors plus aussi lisses que initialement. Cela s'explique par le fait que les différents taux sont calculés pour des âges entiers. Lorsque l'âge à l'incidence est corrigé, il est remplacé par un autre âge également entier. En utilisant des correctifs d'âge entier, la maille de l'étude est trop grossière pour obtenir une courbe plus fidèle aux résultats souhaités. Cependant il est possible d'affiner la maille en considérant des âges de tarification décimaux. L'incidence à ces âges est simplement obtenue par interpolation linéaire entre deux âges entiers. Les résultats suite à l'application de correctifs d'âge actuariels décimaux deviennent alors plus fidèles à la courbe de tarif corrigé.

## 5.2 Comparaison au tarif existant

### 5.2.1 La tarification du BCAC

#### a) Les principes de tarification de l'arrêt de travail selon le BCAC

Réglementairement, la tarification du risque Arrêt de travail n'est pas encadrée. En l'absence d'historique de résultats d'un régime et sur des bases de calcul purement théorique, les tarifs proposés par les organismes assureurs peuvent varier du simple au double. En assurance collective, les organismes assureurs ne disposant pas d'un fichier des assurés, pour effectuer la tarification, ils se servent généralement des barèmes mis à leur disposition par le BCAC (Bureau Commun des Assurances Collectives).

Le tarif est obtenu à partir de :

- la détermination de la période couverte (type et durée de franchise)
- l'estimation du coût (niveau des garanties)
- un taux de base (survenance)

A partir de ces éléments, le BCAC a établi un barème de correctifs permettant l'estimation du coût de l'arrêt de travail. Ce barème est établi pour un groupe d'homme de 40 ans d'âge moyen. Ensuite des correctifs concernent la catégorie socio-professionnelle, l'âge, le sexe de l'assuré ainsi que le type et la durée de la franchise sont appliqués selon les caractéristiques de la population assurée et les clauses de franchise du contrat.

**Le correctif d'âge :** Comme vu précédemment, le risque varie avec l'âge. Les barèmes fournis par le BCAC se base sur le coût d'un groupe d'hommes de 40 ans d'âge moyen, puis des correctifs sont appliqués pour les autres âges.

**Le correctif de sexe :** Généralement, le risque est plus élevé pour les groupes féminins que pour les groupes masculins. C'est la raison pour laquelle des correctifs sont appliqués pour prendre en compte la démographie du groupe étudié. Cependant, un même niveau de tarification que pour les hommes est appliqué aux femmes occupant un poste de cadre.

**Le correctif socioprofessionnel :** En règle générale, les salariés sont d'autant plus présents que leur travail est plus qualifié et leur responsabilité plus importante. Les emplois considérés comme pénibles, sont sujets à des absences plus fréquentes. Quel que soit le secteur d'activité, les cadres sont en moyenne moins souvent absents que les non cadres. Il est donc nécessaire d'appliquer un correctif.

**Le correctif franchise :** Le barème fourni par le BCAC intègre un correctif selon la durée de franchise. En effet, le tarifs pour les franchises courtes doit être plus élevé que celle des franchises longues car l'occurrence est plus importante pour ces franchises. Aussi, les tarifs fournis par le BCAC sont déterminés pour une franchise continue. Les contrats prévoyant des franchises discontinues ou en relais, sont tarifés via une table correspondance avec une franchise continue.

Cependant, le barème du BCAC et ses correctifs ont été établis à partir de statistiques des années 1960. Depuis cette date, les taux d'absentéisme ont évolué en raison du contexte économique (chômage, précarité de l'emploi, diminution de la durée du travail) mais aussi grâce à une meilleure prévention. Pour tenir compte de ces évolutions, de nouveaux correctifs tarifaires à appliquer sur les barèmes cités ci-dessus ont été érigés.

#### b) La table d'équivalence entre la franchise continue et discontinue

Comme brièvement expliqué précédemment, afin de tarifer la franchise discontinue, le BCAC a élaboré la table d'équivalence ci dessous :

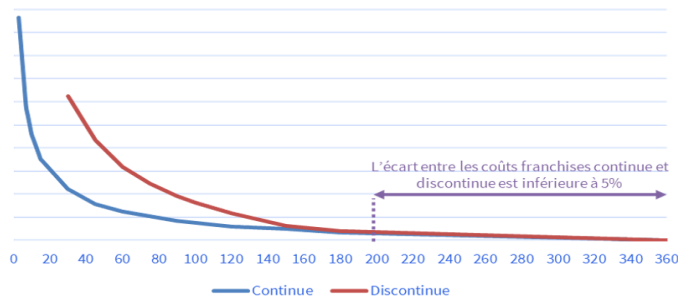
<b>Franchise discontinue</b>	30	45	60	75	90	100	120	150	180	360
<b>Equivalence continue</b>	7	20	45	60	75	90	120	150	180	360

FIGURE 5.9 – Table d'équivalence de la franchise discontinue en franchise continue

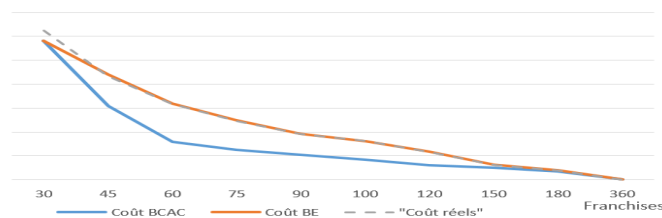
Ainsi, le coût d'une garantie sous la franchise discontinue 90 jours est équivalent au coût d'une garantie sous la franchise continue 75 jours. Tarifer la franchise 90 jours discontinue revient donc à tarifier la franchise 75 jours continue.



Cependant, cette table de correspondance ne reflète plus la réalité du portefeuille. En effet, les coûts obtenus suivant la méthode BCAC se révèlent moins conséquents que ceux calculés en vision "Best-estimate". Les trois figures qui suivent permettent de visualiser ces disparités. Par souci de confidentialité les valeurs présentées ci-dessous ont été modifiées.



(a) Step 1



(c) Step 3

Franchise Discontinue	Equivalence BCAC	Equivalence observée
30	7	7
45	20	11
60	45	19
75	60	27
90	75	37
100	90	44
120	120	66
150	150	117
180	180	169

(b) Step 2

Dans un premier temps, le graphique 5.10a permet de comparer à durée de franchises égales les différences moyennes de coûts entre les garanties sous franchises continues et discontinues. Alors que le BCAC estimait à 120 jours la durée à partir de laquelle la différence entre les deux types de franchises était négligeable, cette figure montre que les coûts réellement observés en franchise discontinue se confondent avec les coûts du continue à partir de 200 jours avec une marge d'erreur inférieure à 5%.

De plus, à partir du graphe 5.10a, il à été possible d'obtenir le tableau 5.10c. Celui-ci permet de constater quelles conversions de franchises discontinue/continue sont appropriées pour s'approcher au mieux des coûts discontinus calculés. Bien que cette méthode de conversion ne sera pas reprise en vision "Best-Estimate", cela ajoute un facteur expliquant en quoi le tarif BCAC n'est pas adapté au portefeuille d'étude. Une franchise discontinue 90 jours considérée équivalente à une franchise continue 75 jours par le BCAC peut être assimilée à une franchise continue 37 jours lorsque l'on s'appuie sur les nouvelles tables calculées.

Les chiffres de la table 5.10c se vérifient à l'aide du graphe 5.10b qui prouvent bien que les nouvelles équivalences sont adaptées à la vision réelle des coûts de l'arrêt de travail en franchise discontinue.

### 5.2.2 Les autres impacts clés de la tarification "Best Estimate"

Outre le problème décrit précédemment, le passage au "Best-Estimate" met en exergue d'autres impacts accentuant l'écart entre le tarif du BCAC et le tarif de l'incapacité temporaire se basant sur le portefeuille. En effet, le BCAC estimant le coût de la franchise discontinu à partir des coûts de la franchise continue, les défaillances de tarification dans le cas du continu impactent également le cas discontinu. Les deux principales faiblesses de la méthode de tarification actuelle sont présentées ci-dessous :

#### a) La segmentation des cadres

Dans sa méthode de calcul, la méthode du BCAC fait une distinction de tarif entre les hommes et les femmes non cadres mais pas chez les cadres. Cela vient probablement du fait que les données à la base de cette méthode datent des années 1960. Selon l'INSEE, à cette époque seulement 40 à 45 % des femmes de 30 à 50 ans étaient déclarées actives. Désormais, la proportion de femmes actives à ces âges est supérieure à 80 % et elles s'érigent beaucoup plus aisément vers des postes à responsabilité. Aussi, la part globale de cadres dans la population active a triplé entre les années 60 et aujourd'hui passant de 5% à plus de 16%. Comme le montrent les résultats précédents, le comportement des femmes est différent de celui des hommes quelques soit l'emploi occupé. C'est pourquoi il semble désormais pertinent de segmenter les femmes et les hommes également chez les cadres.

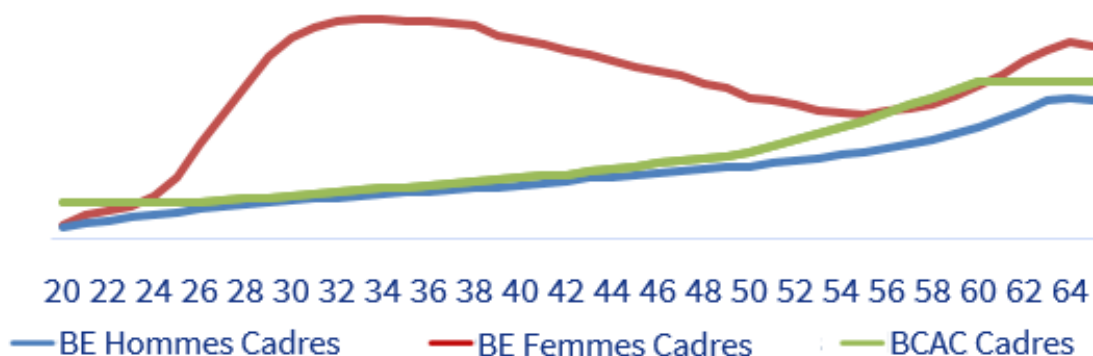


FIGURE 5.11 – Franchise discontinuée 30 jours - Comparaison Tarif "Best-estimate" et BCAC

Dans le cas de la franchise discontinuée 30 jours, la méthode du BCAC permet d'obtenir un tarif assez proche de celui calculé en ne tenant compte que du comportement des hommes cadres. Cependant ce graphe montre également que lorsque l'effectif assuré est composé d'une part non négligeable de femmes, le budget global calculé a de forte chance d'être sous-tarifé. Négliger ces comportements présente alors un risque pour l'assureur.

**b) La sous-tarification aux franchises courtes**

Ici, sont considérés comme courtes les durées de franchises inférieures à 10 jours. Au premier abord cela semble sans intérêt de les étudier dans le cas de la tarification de la franchise discontinue car il est très rare de trouver dans des durées de franchises inférieures à 30 jours. Seules les franchises de type continues sont concernées par les courts délais de carence. Néanmoins, dans la méthode de tarification du BCAC, sous les mêmes conditions démographiques, une garantie avec une franchise discontinue 30 jours est à iso-tarif avec une garantie ayant une franchise continue 7 jours. Savoir bien tarifier la franchise continue courte est donc un élément clé de la tarification des franchises discontinues.

La tarification de l'incapacité temporaire franchise continue en vision "Best Estimate" a été faite suivant la même méthode que celle présentée dans cette étude. Seule l'événement considéré a changé. Au lieu d'étudier le cumul des arrêts de travail d'un salarié sur une année, la durée de chacun des arrêts du portefeuille a été étudiée séparément. Le tarif global obtenu dans le cas des franchises inférieures ou égales à 10 jours est présenté dans le graphique ci-dessous. Ce tarif est comparé à celui calculé via la méthode BCAC.

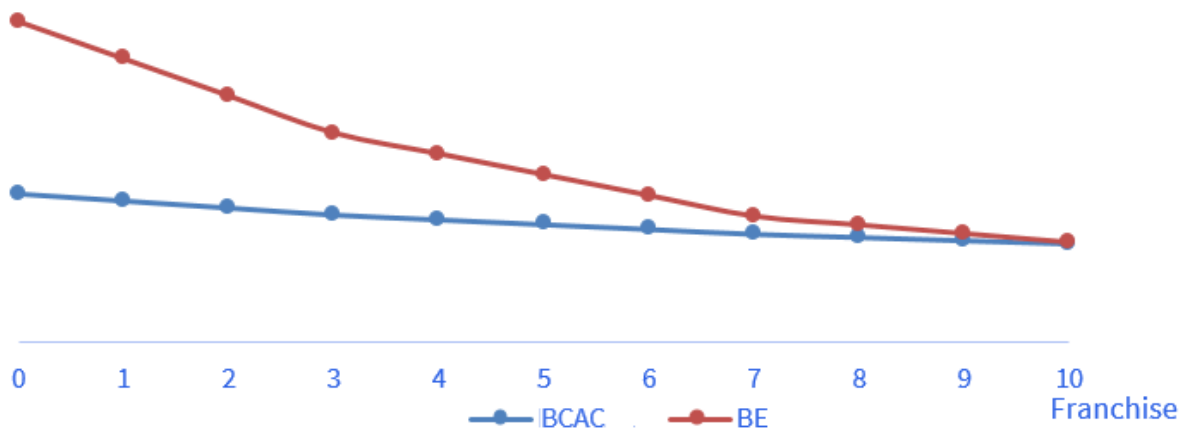


FIGURE 5.12 – Tarifs des franchises continues courtes

Dans le cas des franchises très courtes (inférieures à 5 jours), le tarif en vision "Best Estimate" est approximativement une à deux fois plus cher que les résultats obtenus avec les tables du BCAC. Cet écart de tarif se réduit au fur et à mesure que la durée de franchise augmente. Ainsi, dans le cas particulier de la franchise 30 jours, la sous-tarification de la franchise continue 7 jours est un élément jouant dans la différence de tarif entre la méthode présentée dans cette étude et celle actuellement utilisée.

# Conclusion

Ce mémoire avait pour objectif d'établir une méthode de tarification propre à la garantie incapacité temporaire sous franchise discontinuée à partir des données expérimentales propre au portefeuille assuré. Le but était d'actualiser la méthode dérivée fournie par le BCAC, qui pour tarifier une telle garantie se basait sur des correctifs du coût établi dans le cas d'une franchise continue. Les éléments de cette étude montrent qu'il est possible d'obtenir une méthode de tarification qui simplifie et clarifie les calculs de coût tout en reflétant mieux la réalité des sinistres observés les années passées.

Pour cela, il a d'abord fallu comprendre le risque étudié afin de retraiter les données obtenues par la DSN et ainsi représenter au mieux ce risque dans la base d'étude. Il est apparu que l'événement étudié dans le cadre de la garantie incapacité temporaire en franchise discontinuée est différent que celui dans le cadre de la franchise continue.

Ensuite, afin d'aboutir à une méthode simple de tarification de forme "*Coût × Fréquence*", deux tables ont dû être élaborées.

D'une part, la *Fréquence* est déterminée à partir d'une table d'entrée en incapacité aussi dit table d'incidence. Le but était de définir selon les caractéristiques d'âge, de sexe et de collège des salariés, mais également selon la durée de franchise, un taux de nouveaux entrants dans un état d'incapacité temporaire. S'agissant d'un problème d'analyse de survie (ici dans l'état d'activité), les modèles très connus de Kaplan-Meier et de Hoem suivi par la méthode de lissage de Whittaker-Henderson ont permis d'aboutir au résultat souhaité. Ces taux d'incidence obtenus sont par la suite soumis à des correctifs d'âge actuariels en vue de répondre à la problématique de calcul à l'âge moyen propre à l'assurance collective.

D'autre part, afin de construire la partie *Coût* de la formule, il était nécessaire d'établir une table de maintien propre au risque étudié. Cette table a été élaborée parallèlement à cette étude suivant des modèles de durées semblables à ceux étudiés lors du travail sur l'incidence. A partir de cette dernière, il a été possible de calculer des annuités selon la franchise voulue représentant le coût cherché.

C'est ainsi qu'ont pu être conçus des tarifs pour un groupe d'assurés suivant leurs caractéristiques d'âge, de sexe et de collège ainsi que suivant la durée de carence associée à la garantie d'incapacité temporaire (sous franchise discontinuée) souscrite.

Bien évidemment, cette étude présente de nombreux axes d'amélioration qui n'ont pu être pris en compte ici. De manière non exhaustive, figure parmi ces points :

- Le caractère récent de la DSN fait que la profondeur des données analysées n'est limitée qu'à deux ans (2018 et 2019). Avec une meilleure ancienneté sur la sinistralité, les résultats des tables seront de moins en moins volatiles et plus représentatifs de la réalité.
- L'ajustement et la validation des tables ne reposent que sur la méthode de Whittaker-Henderson. Ce modèle de lissage est couramment utilisé en analyse de survie et majoritairement considéré fiable une fois minutieusement paramétré comme c'est le cas ici. Néanmoins, il est possible de challenger cette méthode avec d'autres méthodes de lissage connues telles que le lissage par splines ou par la méthode des noyaux.
- La méthode d'élaboration des correctifs d'âge actuariel peut également être challengée en confrontant les résultats obtenus à de nombreuses simulations démographiques. Le but est que le tarif obtenu tête par tête concorde significativement au tarif démographie moyenne.

Finalement, comme expliqué, l'approche présentée dans ce mémoire a pour but in fine de remplacer l'actuelle méthode de tarification de l'arrêt de travail du BCAC. Ici, une solution est proposée pour les garanties sous franchise discontinuée. D'autres études répondent aux garanties souscrites sous d'autres types de franchises ou aux garanties liées à l'invalidité. Le prochain objectif est donc une phase de test de toutes ces refontes afin de redéfinir et réajuster les différents travaux. Il est absolument nécessaire que les nouvelles primes calculées soient représentatives de la sinistralité à venir.

# Bibliographie

## Cours et ouvrages

O.LOPEZ (ISUP 2019) *Survival analysis and other topics*

C.CORNIALLE (ISUP 2019) *La prévoyance collective*

G.BIESSY (ISUP 2020) *Modélisation en Assurance Dépendance*

F.PLANCHET (ISFA 2020) *Modèles de durée : méthodes de lissage et d'ajustement*

GRUPE AYMING (2018) *Le 10ème Baromètre de l'Absentéisme® et de l'Engagement d'Ayming*

C.MARBOT, C.POLLAK (2014) *Les dépenses de santé en 2014*, DREES

C.INAN (2013), *Les absences au travail des salariés pour raisons de santé : un rôle important des conditions de travail*, Dares Analyses

L.GIESECKE (1981), *Use of the chi-square statistic to set Whittaker-Henderson smoothing coefficients*

## Mémoires d'actuariat

P. MBOW (2019) *Construction de lois d'incidence en arrêt de travail à partir des DSN*

J-B. ROUSSEL & R. ZAOUI (2014) *Loi expérimentale de maintien en incapacité temporaire avec un pas quotidien*

A.GAUMET (2001) *Construction de tables d'expérience pour l'entrée et le maintien en incapacité*

A.DE LA MORINERIE (2016) *Conceptualisation d'un modèle multi-états en arrêt de travail et application à une loi d'incidence en incapacité*