Mémoire présenté le :

pour l'obtention du Diplôme Universitaire d'actuariat de l'ISFA et l'admission à l'Institut des Actuaires

Par : LAMRI Yacine

Titre : Construction de lois de maintien en incapacité par pathologie et application à l'estimation du gain des contrôles médicaux

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

Membres présents du jury de l'ISFA

Entreprise : Malakoff Humanis

Nom :

Signature :

Directeur de mémoire en entreprise :

Nom : IORGOVAN Anca

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

Table des matières

AVANT-PROPOS

Pour des raisons de confidentialités, certaines données ont été modifiées. Par ailleurs, les résultats de cette étude, qu'ils soient explicites, sous-entendus ou masqués sont strictement confidentiels et ne doivent en aucune façon être exploités en dehors du cadre de ce mémoire.

RESUME

Depuis quelques années Malakoff Humanis a développé un premier suivi à travers un certificat médical d'incapacité de travail (CMIT) envoyé à l'assuré généralement au 31e jour d'arrêt. Cette démarche permet à l'assureur de récupérer des informations essentielles telles que la pathologie à l'origine de l'arrêt de travail ou les différents traitements prescrits par le médecin traitant. Ces données permettent d'avoir une meilleure compréhension du portefeuille.

Dans un contexte de dégradation du risque incapacité, il est important pour les compagnies d'assurance d'estimer la durée des arrêts de travail le plus finement possible dans le but de mieux piloter leurs provisions et faire face aux engagements vis-à-vis des assurés. Les assureurs construisent généralement des lois de maintien en incapacité temporaire de travail en utilisant des tables réglementaires du Bureau Commun d'Assurances Collectives (BCAC) ou des tables spécifiques certifiées.

L'objectif de ce mémoire est de présenter les différentes étapes dans la conception de lois de maintien en incapacité par pathologie. S'intéresser à cette segmentation est pertinent, car Malakoff Humanis développe son accompagnement d'aide au retour à l'emploi (ARE) ainsi qu'un programme de contrôle médical visant à détecter les arrêts frauduleux.

Après la présentation des différents risques et des différents enjeux, ces lois de maintien en incapacité sont construites puis segmentées par famille de pathologie à l'aide d'une base de données nettoyée préalablement et constituée d'arrêts de travail de 2016 à 2021. Les taux bruts de sortie étant erratiques, les tables seront lissées en utilisant des méthodes mathématiques telles que la moyenne mobile ou la méthode de Whittaker-Henderson.

Les différentes tables étant construites et lissées, il sera question de juger de leur pertinence et de leur prudence à travers un backtesting de manière à vérifier l'adéquation des prédictions avec les résultats réels.

Enfin, la dernière partie s'attachera à utiliser les résultats et les différentes durées modélisées par pathologie pour analyser l'efficacité du processus de contrôle médical implémenté depuis quelques années par Malakoff Humanis.

ABSTRACT

For several years Malakoff Humanis has developed a first follow-up through a medical certificate of incapacity for work sent to the insured generally on the 31st day of sick leave. This approach allows the insurer to obtain essential information such as the pathology or the different treatments prescribed by doctor. This data provides a better understanding of the portfolio.

In a deteriorating context of the number of work stoppages, it is important for insurance companies to estimate the duration of these stoppages as accurately as possible in order to better manage their provisions and meet commitments to insured. Insurers generally construct experience tables in incapacity using either the 'Bureau Commun d'Assurances Collectives' (BCAC) regulatory tables or specific certified tables.

The objective of this thesis is to present the different stages in the construction of experience table by pathology. Taking an interest in the pathology information is relevant for the group because Malakoff Humanis is developing its back-to-work assistance helping the injured worker to return to the workplace as well as a medical control program aimed at detecting fraudulent work stoppages.

After the presentation of the different risks and the different issues, experience tables in incapacity are constructed and then segmented by pathology from a previously cleaned database consisting of work stoppages from 2016 to 2021. Then, this tables will be smoothed using mathematical methods such as the moving average or the Whittaker-Henderson method.

The different tables being built and smoothed, it leads us to verify the relevance and the prudence through backtesting in order to verify the adequacy of the predictions with the actual results.

Finally, the last part will focus on using the results to analyze the profitability of the medical control process implemented several years ago by Malakoff Humanis.

REMERCIEMENT

Je remercie l'ensemble de l'équipe pédagogique de l'Institut de Science Financière et d'Assurances pour leur accompagnement et la qualité des cours dispensés.

Un remerciement particulier à Alexis Bienvenue, mon tuteur, pour son accompagnement et ses remarques pertinentes ainsi qu'à Frederic Planchet pour son cours de modèles de durée dont le contenu m'a énormément aidé lors de la réalisation de ce mémoire.

Je remercie également ma tutrice Anca Iorgovan pour l'accueil chaleureux au sein de l'équipe Prévoyance de Malakoff Humanis ainsi que pour m'avoir permis d'effectuer ce travail et Bienvenu Kenfack-Nanda, mon tuteur chez Malakoff Humanis pour le temps accordé durant mon année d'alternance. Je remercie également Marie Desruennes pour ses conseils avisés, sa disponibilité et pour m'avoir guidé dans mes travaux.

Enfin, je remercie l'ensemble des personnes rencontrées tout au long de mes études, qui m'ont beaucoup apporté tant sur le plan professionnel que sur le plan personnel. Pour finir, je tiens à remercier ma famille et mes proches pour m'avoir toujours soutenu durant ma scolarité.

INTRODUCTION

Dans un double contexte caractérisé par une augmentation du nombre d'arrêts de travail, notamment liés à des motifs psychiatriques, ainsi qu'un allongement de leur durée, Malakoff Humanis organise des contrôles médicaux afin de détecter et de réguler ces arrêts de travail anormalement longs. Dans le cadre de ce mémoire, nous nous proposerons de modéliser la durée des arrêts de travail pour incapacité par pathologie afin d'utiliser ces résultats dans le cadre d'une estimation de l'efficacité des contrôle médicaux en déterminant un gain sur le risque incapacité : gain en nombre de jours économisés ainsi qu'en euros économisés.

Depuis plusieurs années, un premier suivi du risque est effectué grâce à un certificat médical d'incapacité de travail (CMIT) envoyé à l'assuré en incapacité pour mieux comprendre l'arrêt de travail à travers des premières informations (pathologies, traitements prescrits…). Par ailleurs, ces certificats font ressortir une information intéressante à exploiter : la pathologie à l'origine de l'arrêt de travail. La pathologie à l'origine de l'arrêt de travail a un impact significatif sur la durée. Pour mieux comprendre ce lien, nous modéliserons les durées de maintien en incapacité pour chaque famille de pathologie.

Dans un premier temps, étant donné que nous travaillons sur des données pouvant être censurées, nous privilégierons les modèles non paramétriques de Kaplan Meier et de Nelson Aalen. En outre, nous disposons de variables explicatives telles que les catégories socioprofessionnelles et l'âge des assurés, que nous inclurons dans un modèle de Cox pour évaluer leur influence sur la durée de maintien en incapacité.
Pour améliorer la qualité de nos modèles, nous appliquerons différentes méthodes de lissage, notamment la méthode de Whittaker-Henderson.

Les différentes durées étant modélisées, nous les utiliserons pour mesurer l'efficacité des contrôles médicaux en comparant, pour un fichier contenant les arrêts de travail expertisés jugés non justifiés, la durée réelle et la durée modélisée déterminant ainsi un gain en nombre de jours économisés puis en euros économisés.

L'objectif final de cette étude est d'améliorer le ciblage des arrêts de travail nécessitant une expertise médicale.

1 Généralités sur la prévoyance et le risque arrêt de travail

1.1 La protection sociale et les garanties de base

1.1.1 La protection sociale

La protection sociale représente l'ensemble des mécanismes permettant aux individus de faire face financièrement à certains risques sociaux, c'est-à-dire à l'ensemble des situations pouvant provoquer une baisse des ressources ou une hausse des dépenses.

Les différents risques sociaux pouvant notamment être couverts sont :

- **Les dépenses en santé** : remboursement des frais médicaux (actes courants, hospitalisation, optique et dentaire)
- **L'arrêt de travail** : remboursement d'indemnités journalières (IJ) : incapacité, maternité et remboursements pour rentes d'invalidité
- **Le décès** : versement d'un capital ou d'une rente aux ayants droit au moment du décès de l'assuré
- **L'épargne-retraite :** constitution d'une épargne lors de la vie active permettant de bénéficier d'une rente ou d'un capital à la retraite
- **La dépendance** : versement d'un capital ou d'une rente lorsque l'assuré perd en autonomie
- **Le Chômage :** versement des indemnités de chômage (par Pôle Emploi)**.**

1.1.2 Le régime général de la Sécurité sociale

La protection sociale française repose sur le régime général de la Sécurité sociale.

Le régime général (ou régime de base) de la Sécurité sociale est le premier intervenant du marché en France. Il prend en charge l'ensemble des professions excepté les professions agricoles (prises en charge par la MSA mutualité Sociale Agricole).

La Sécurité sociale, obligatoire et universelle, créée en 1945 après la seconde guerre mondiale garantit que chaque individu bénéficiera en toutes circonstances d'une protection face aux risques sociaux. En fonction du type d'activité professionnelle, les assurés sont répartis selon 3 principaux régimes :

- Le régime général (ou le régime de base) s'adressant aux salariés ainsi qu'aux travailleurs indépendants depuis 2018.
- Le régime agricole s'adressant aux exploitants et aux salariés agricoles
- Les régimes dits « spéciaux » tels que celui de l'Assemblée nationale, de la RATP ou de la SNCF

Notre étude portera principalement sur le régime général. Celui-ci couvre près de 90% de la population.

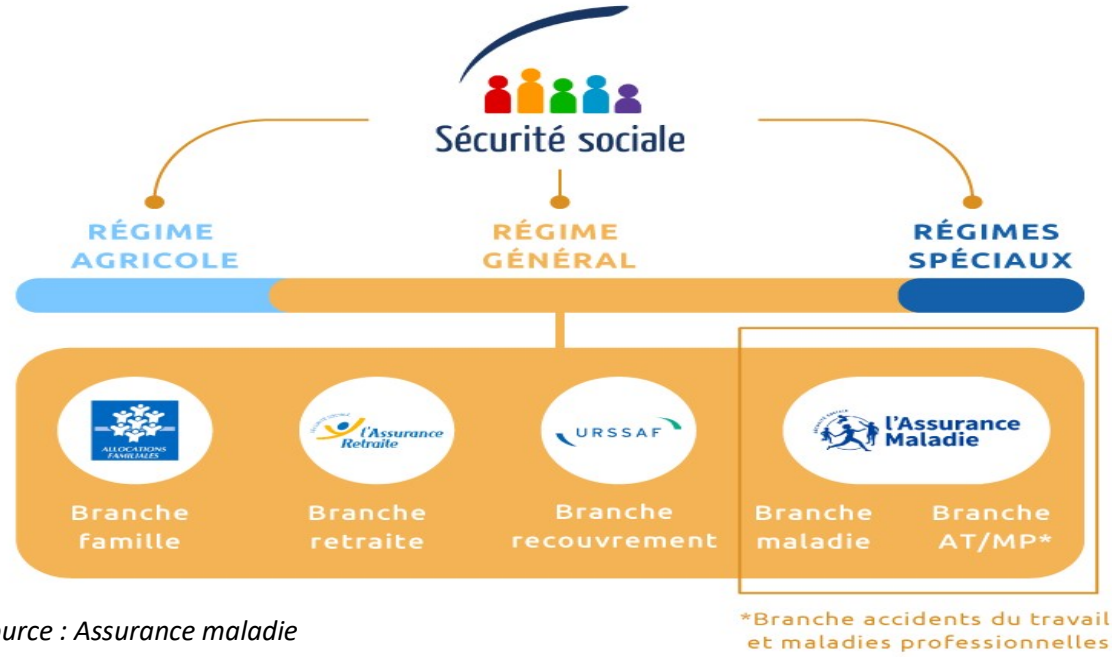
S'adressant initialement aux salariés du secteur privé, le régime général a été, au fur et à mesure de son extension, amené à intégrer dans sa couverture des populations comme par exemples les étudiants ou les demandeurs d'emploi.

Depuis le 1er janvier 2018, le Régime général couvre également les indépendants (artisans, commerçants, professions libérales non réglementées…)

Le Régime général est divisé en 5 branches d'activité ayant à leur tête une caisse nationale autonome. Une branche est une entité qui a à sa charge la gestion d'un ou plusieurs "risques". Ces risques sont définis comme des événements qui peuvent au cours d'une vie, porter atteinte à la sécurité économique d'une personne. Ils font donc l'objet d'une prise en compte, d'une réparation ou d'une rétribution. Les 5 branches sont :

- **La branche famille**, gérée par les Allocations familiales et s'occupant principalement des prestations familiales, c'est-à-dire l'ensemble des aides liées à la naissance, à la garde, aux différents handicaps et au logement.
- **La branche retraite**, gérée par l'Assurance Retraite, versant les pensions aux retraités issus de l'industrie, des services et du commerce.
- **La branche recouvrement**, gérée par l'Urssaf, collectant les contributions sociales auprès des entreprises, des particuliers et des travailleurs indépendants dans le but de les redistribuer aux autres branches pour financer la totalité des prestations.
- **La branche Maladie**, gérée par l'Assurance Maladie et recouvrant les risques maladie, maternité, invalidité et décès.
- **La branche accidents du travail**, gérée par l'Assurance Maladie et prenant en charge les risques professionnels : accidents du travail, accidents de trajet et maladies professionnelles.

Schéma du fonctionnement de la Sécurité sociale



Source : Assurance maladie

1.1.3 Assurances complémentaires

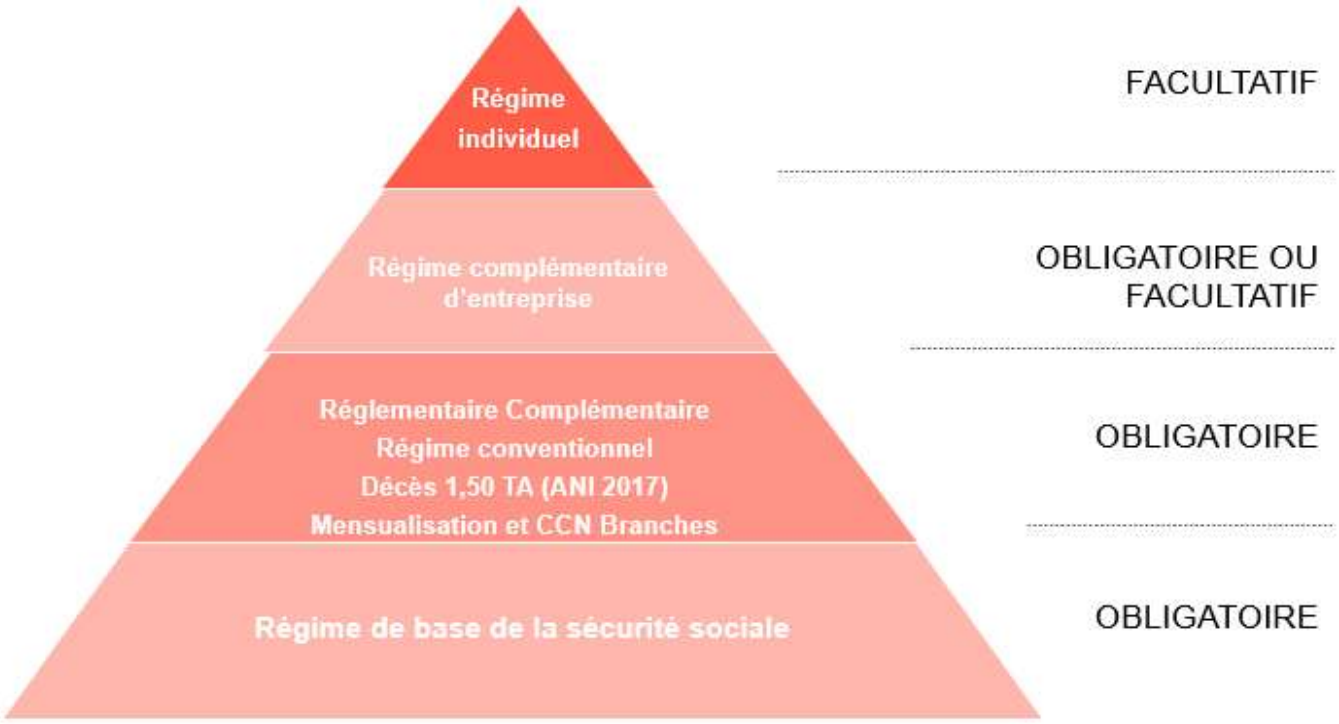
La dérive et les déficits constants enregistrés par le régime de base entraînent un réajustement des prestations servies par la Sécurité sociale qui sont fréquemment revues à la baisse.

C'est pour cette raison que les compagnies d'assurances (relevant du code des assurances), les institutions de prévoyance (relevant du code de la Sécurité sociale) et les mutuelles (relevant du code de la mutualité) proposent des assurances complémentaires aussi bien en prévoyance qu'en santé afin d'aider les assurés à pallier la baisse des prestations versées par la Sécurité sociale.

Certains de ces régimes sont obligatoires (régimes complémentaires de retraite des salariés) d'autres sont facultatifs (mutuelles de santé).

Les différents niveaux de la protection sociale sont synthétisés à travers le schéma ci-dessous :

Pyramide de la protection sociale du salarié



Source : Malakoff Humanis

1.2 Les opérations de prévoyance

1.2.1 Présentation de la prévoyance

Les opérations de prévoyance servent à faire face aux aléas de la vie.

Selon la loi n° 89-1009 du 31 décembre 1989, dite loi EVIN, la prévoyance regroupe « 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, des risques d'incapacité de travail ou d'invalidité ou du risque chômage ».

Ainsi, la prévoyance peut être collective, proposée par l'employeur ou souscrite de façon individuelle.

Au sens de la Sécurité sociale, un arrêt de travail peut être engendré aussi bien par une maladie que par un accident, dans le cadre du travail ou de la vie privée.

Nous distinguons 2 états :

- L'Incapacité Temporaire de travail, communément appelée Incapacité
- L'Incapacité Permanente de travail, communément appelée Invalidité

1.2.2 Le risque incapacité

D'après l'article L 321-1 du code de la sécurité sociale, un individu est en incapacité de travail s'il est dans « l'incapacité physique constatée par le médecin traitant, selon les règles définies par l'article L 162-4-1, de continuer ou de reprendre le travail ». L'arrêt de travail dure au maximum 3 ans. Au-delà de ce délai, si l'assuré est toujours dans l'impossibilité de reprendre son activité professionnelle, l'incapacité devient une invalidité, selon la décision de la sécurité sociale. Les 3 causes de sortie possibles de l'état d'incapacité sont : la reprise du travail (ou le licenciement), le décès et le passage en invalidité.

1.2.3 Le risque invalidité

D'après l'article L 341-1 de la Sécurité sociale, un individu est en état d'invalidité « lorsqu'il présente une invalidité réduisant dans des proportions déterminées (2/3 ou plus), sa capacité de travail ou de gain, c'est-à-dire le mettant hors d'état de se procurer, dans une profession quelconque, un salaire supérieur à une fraction de la rémunération normale perçue dans la même région par des travailleurs de la même catégorie, dans la profession qu'il exerçait avant la date de l'interruption de travail suivie d'invalidité ou la date de la constatation médicale de l'invalidité si celle-ci résulte de l'usure prématurée de l'organisme ».
3 catégories d'invalidité sont à distinguer :

- **Invalidité 1ère catégorie**

Si l'assuré peut exercer une activité professionnelle rémunérée malgré son invalidité, alors il est placé en invalidité dite de 1 ère catégorie.

- **Invalidité 2ème catégorie**

Si l'assuré ne peut plus exercer une activité professionnelle rémunérée à la suite d'un accident ou d'une maladie, alors il est placé en invalidité dites de 2ème catégorie.

- **Invalidité 3ème catégorie**

Si à la suite d'un accident ou d'une maladie, l'assuré ne peut plus exercer une activité professionnelle rémunérée et à besoin d'une tierce personne pour l'assister dans les gestes essentiels de la vie courante, alors l'assuré est placé en invalidité dites de 3ème catégorie.

Il est dit que l'assuré ne peut effectuer les gestes essentiels de la vie courante, lorsqu'il ne peut accomplir seul, totalement, habituellement et correctement au moins quatre des actes de la grille nationale (grille nationale annexée au décret nº 97-427 du 28 avril 1997 portant application de certaines dispositions de la loi nº 97-60 du 24 janvier 1997). Le schéma suivant résume les différentes catégories d'invalidité

Les 3 catégories d'invalidité

Source : Malakoff Humanis

1.3 Le cadre législatif du risque arrêt de travail

La garantie arrêt de travail est régie par un certain nombre de textes. Les organismes assureurs sont tenus de respecter une réglementation stricte afin de limiter leurs risques ainsi que le risque de leurs assurés.

1.3.1 Convention Collective Nationale des Cadres (1947)

La convention Collective Nationale des Cadres (1947) est le premier texte officiel de prévoyance collective. Elle oblige les entreprises employant des cadres ou des assimilés cadres, à cotiser obligatoirement à un régime de prévoyance collective et de verser une cotisation minimale obligatoire égale à la hauteur de 1,50% de la tranche A de rémunération, dont 0,76% affecté à la couverture du risque décès.

1.3.2 Loi de mensualisation de 1978

La loi du 19 janvier 1978 dite de mensualisation fait suite à l'Accord National Interprofessionnel (ANI) du 10 décembre 1977 et impose à l'employeur de maintenir le niveau de salaire des salariés en arrêt de travail pour cause de maladie ou d'accident. Cette loi est applicable à tous les salariés à l'exception des personnes travaillant à domicile, des saisonniers, des intermittents et des intérimaires.

Durant l'arrêt de travail et afin de prendre en considération la répartition inégale des jours entre les 12 mois de l'année, l'assuré en arrêt de travail peut bénéficier d'indemnités en complément des indemnités journalières.

Pour les percevoir, le salarié doit remplir les conditions suivantes :

- Il doit justifier, à la suite de l'ANI du 11 janvier 2008, d'une année d'ancienneté dans l'entreprise. (3 années étaient jusqu'alors requises)
- Son incapacité doit être attestée par un certificat médical suivi éventuellement d'une contre visite.
- Il doit être pris en charge par la Sécurité sociale.
- L'assuré doit être soigné sur le territoire national ou dans un état membre de la Communauté économique européenne ou de l'espace économique européen.

Cette indemnité complémentaire permet au salarié de percevoir, pendant les 30 premiers jours, 90 % de la rémunération brute qu'il aurait perçue s'il avait continué à travailler, puis les 2/3 pendant les 30 jours suivants.

1.3.3 Loi Evin du 31 décembre 1989

La loi Evin du 31 décembre 1989 a permis d'harmoniser les règles de prévoyance entre les différentes assurances. Elle a également permis de renforcer la protection des assurés.

Cette loi instaure notamment :

- Le caractère collectif de la souscription : la sélection médicale individuelle est interdite et tous les salariés sont couverts y compris ceux en arrêt de travail au moment de la signature de contrat d'assurance
- Le maintien de prestations au niveau atteint lors de la résiliation
- Le maintien de garanties décès au niveau atteint pour les personnes en arrêt de travail lors de la résiliation
- Une notice d'information sur les différentes garanties doit être communiquée aux assurés
- La communication par l'organisme assureur d'un rapport annuel sur les résultats du contrat à l'employeur

1.3.4 Loi du 8 août 1994

Les contrats de rente prévoient une revalorisation annuelle. Ainsi, par exemple, l'évolution d'un contrat de rente sera revue à la hausse en fonction de l'évolution de la valeur d'un point AGIRC.

Dans ce cadre, la loi du 8 août 1994 impose à l'employeur de revaloriser les rentes et les prestations versées par l'assureur en cours en cas de changement d'organisme d'assurance. Il s'agit d'un complément à la loi EVIN.

1.3.5 Accord National Interprofessionnel (ANI) de 2008

L'accord renforce certaines dispositions de la loi EVIN en instaurant le principe de portabilité des droits de la couverture santé et prévoyance. L'employeur a l'obligation de maintenir les garanties prévoyance et santé des salariés licenciés pendant une durée égale au minimum entre :

- La durée du dernier contrat de travail
- 12 mois

La couverture prévoyance et santé expirera au moment de la fin de la période de maintien des droits ou en cas de reprise d'un nouvel emploi.

1.4 Prise en charge de l'arrêt de travail

1.4.1 Indemnisation de l'arrêt de travail

Lorsqu'un individu est en arrêt de travail, afin de pallier la perte de ses revenus, la Sécurité sociale lui verse des indemnités journalières. Leurs montants ainsi que la date du début d'indemnisation dépendent de l'origine de l'arrêt. Une distinction est faite par la Sécurité sociale entre les arrêts de travail du type privé (maladie et accident de la vie courante) et les arrêts de travail de type professionnel.

Si **l'arrêt de travail est d'origine professionnelle**, l'assuré est indemnisé à partir du premier jour et pendant toute la durée de l'arrêt de travail jusqu'à la guérison ou la consolidation de l'accident du travail ou de la maladie professionnelle. Cette couverture concerne tous les salariés.

Pendant les 28 premiers jours, l'assuré en état d'incapacité reçoit par la Sécurité sociale, 60% de son salaire journalier de base dans la limite de 205,84€. Au-delà du 29eme jour, les indemnités journalières représentent 80% du salaire journalier et sont plafonnées à 274,46€ (en 2022)

Si en revanche, **l'arrêt de travail est de l'ordre du domaine privé**, les indemnités journalières ne commencent qu'à partir du 4^{e} jour et pour une durée maximale de 3 ans. En fonction de la durée de l'arrêt de travail :

- Si la **durée de l'arrêt de travail est inférieure à 6 mois :**

Dans ce cas, afin de percevoir des indemnités journalières, l'assuré doit satisfaire à l'une des deux conditions suivantes :

- Pouvoir justifier au moins 150 heures de travail au cours des 90 jours précédant l'arrêt.
- Avoir cotisé sur un salaire au moins égal à 1015 fois le montant du SMIC horaire au cours des 6 mois précédant l'arrêt.

- Si la **durée de l'arrêt de travail est supérieure à 6 mois :**

L'assuré doit justifier au moins 12 mois d'affiliation auprès de la Sécurité sociale et satisfaire à l'une des deux conditions suivantes :

- Avoir travaillé au moins 600 h au cours des 365 jours précédant la date de l'arrêt de travail.
- Avoir cotisé sur un salaire au moins égal à 2030 fois le montant du SMIC horaire au cours des 12 mois précédant l'arrêt.

Le montant des indemnités journalières versées par la Sécurité sociale est de 50% du salaire journalier de base dans la limite de 1,8 fois le SMIC. Si l'assuré a au moins 3 enfants à charge, à partir du 31ème jour, ses indemnités sont majorées et sont de l'ordre de 2/3 du salaire journalier.

A la suite de l'accord de mensualisation de 1978, les entreprises ont l'obligation de maintenir au minimum le salaire de leurs salariés (à partir d'un an d'ancienneté) en arrêt de travail à la hauteur de 90% du salaire durant les 30 premiers jours puis de 2/ 3 du salaire durant les 30 jours suivants.

Dans le cas d'un arrêt de travail d'origine privée, la franchise est de 7 jours et les indemnités journalières sont versées après ce délai.

En assurance, une franchise correspond au nombre de jours écoulés entre la date de survenance du sinistre et la date du début d'indemnisation par l'assureur.

A noter également que les durées d'indemnisation augmentent à hauteur de 10 jours par tranche de 5 ans d'ancienneté et sont limités à 90 jours. Par exemple, pour un assuré en incapacité ayant 12 ans d'ancienneté, le salaire sera maintenu à hauteur de 90% pendant 50 jours, puis un maintien à 66% pendant les 50 jours suivants.

1.4.2 Organismes d'assurance

L'employeur souscrit donc des contrats de prévoyance complémentaire auprès d'organismes d'assurance. Il existe 3 types d'organismes d'assurance porteurs de risques, habilités par l'article 1 de la loi Evin du 31 décembre 1989 et pouvant effectuer des opérations de prévoyance.

- Les sociétés d'assurance ;
- Les institutions de prévoyance ;
- Les mutuelles.

A. Les sociétés d'assurance

Les entreprises d'assurances sont régies juridiquement par le code des assurances tant pour leur fonctionnement que pour les contrats qu'elles émettent. Il existe 2 formes de sociétés d'assurance, les sociétés anonymes d'assurances reversant leurs bénéfices aux actionnaires et les sociétés d'assurances mutuelles (SAM) appelées aussi sociétés mutuelles d'assurances (SMA) qui ne bénéficient pas de capital social et n'ont pas d'actionnaires à rémunérer. Ces dernières sont gérées soit par des administrateurs bénévoles élus soit par des sociétaires.

Les activités de ces sociétés concernent tous les domaines de l'assurance. Elles ne peuvent cependant pas avoir en même temps des activités en assurance vie et en assurance non-vie.

Les sociétés d'assurance se révèlent être les leaders sur ce secteur de la prévoyance complémentaire.

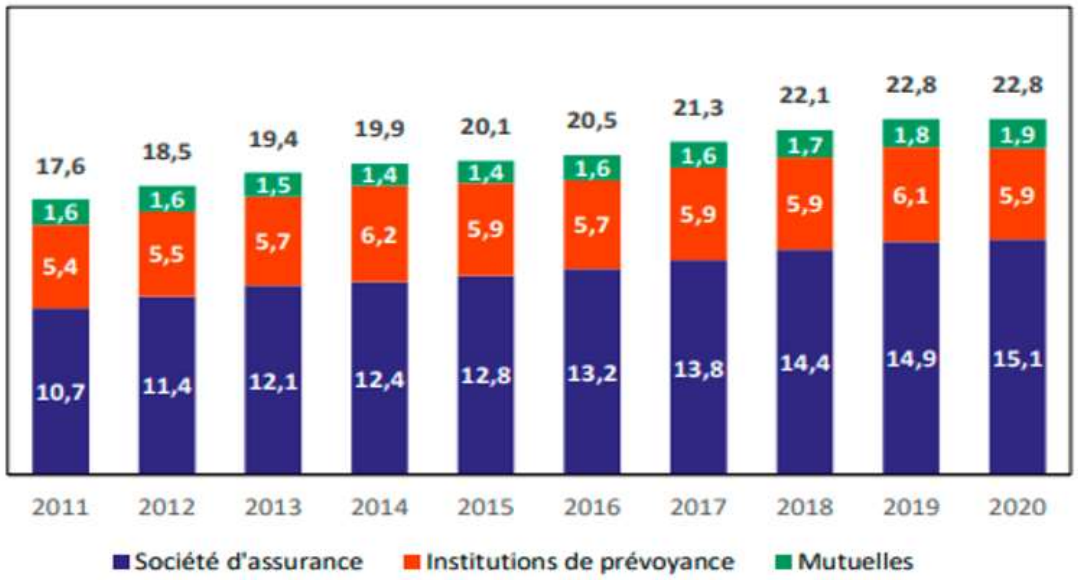
B. Les institutions de prévoyance

Les institutions de prévoyance sont des sociétés de droit privé, visant un but non lucratif, régies par le Livre IX du Code de la Sécurité sociale. Elles sont administrées paritairement par des partenaires sociaux. Leurs activités concernent principalement la prévoyance collective. En effet, les institutions de prévoyance n'ont le droit de faire de la prévoyance individuelle que sous certaines conditions.

C. Les mutuelles

Les mutuelles sont régies par le code de la mutualité. Elles sont définies comme des organismes à but non lucratif. Les mutuelles sont gérées par leurs assurés. En effet, les délègues à l'assemblé générale sont élus par les adhérents. Les mutuelles encadrent principalement l'assurance de personne et plus particulièrement le secteur de la santé. Elles ne peuvent assurer ni les biens, ni la responsabilité.

Montant des cotisations pour les 3 types d'organisations d'assurance. [2]

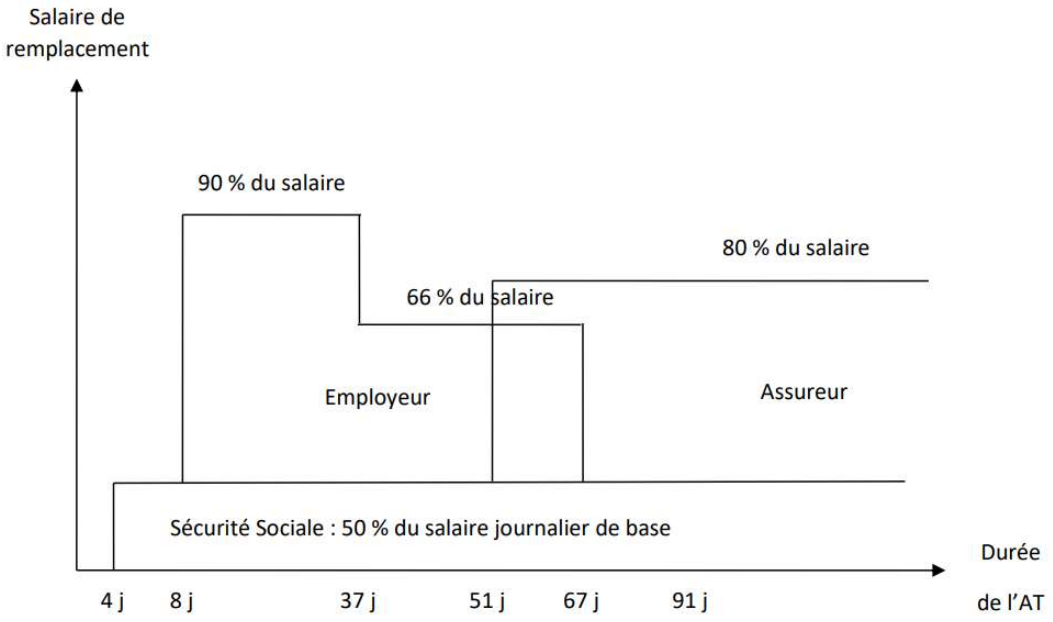


Source : FRANCE ASSUREURS, Le marché des assurances santé et prévoyance en 2020

Les sociétés d'assurance ont les plus grosses parts de marché en termes de cotisation.

En conclusion, le graphique suivant résume le processus de remboursement du salaire d'un salarié en incapacité de travail :

Exemple de remboursement pour un salarié en arrêt de travail. [3]



Source : Prévoyance collective. Support de cours ISFA

2 Etude des données

Le but de ce mémoire est de construire des tables de maintien en incapacité par pathologie.

Une table d'expérience pertinente implique des données fiables et robustes. C'est pourquoi un long travail de contrôle des données a été mené afin d'obtenir une base de données propre et exploitable. Après avoir présenté les différentes variables considérées et expliqué les différents retraitements effectués sur la base, des statistiques descriptives seront réalisées afin de mieux comprendre les différentes dynamiques de notre portefeuille constitué.

2.1 Démarches

À la suite du rapprochement de 2018 entre Malakoff-Médéric et Humanis, nos données proviennent de 2 bases totalement différentes correspondant à l'historique des données contenues dans le système d'information de Malakoff-Médéric et le système d'information d'Humanis.

De plus, pour cette étude, nous disposons d'une extraction provenant du service médical de Malakoff Humanis et contenant les incapacités de travail entre 2016 et 2021 pour lesquels un contrôle médical a été effectué et une pathologie est renseignée (c'est-à-dire pour lesquels le questionnaire médical a été renvoyé par l'assuré).

Le contrôle médical consiste à adresser au salarié en incapacité de travail d'au moins 30 jours ou à période de franchise atteinte, selon les contrats, un certificat médical d'incapacité de travail (CMIT) qu'il devra compléter ainsi que son médecin et retourner à l'attention du médecin conseil au service médical. Par exemple pour un contrat avec une franchise de 4 jours, le contrôle médical sera déclenché au 31ème jour ; pour un contrat avec une franchise de 60 jours, le contrôle médical sera déclenché au 61ème jour.

Sur la base du CMIT complété, la cellule médicale juge si des documents complémentaires et/ou une expertise médicale sont nécessaires.

Le Service Médical Conseil (SMC) effectue dans certains cas un contrôle des arrêts de travail et les arrêts de travail non justifiés sont classés en 3 catégories :

- Non médicalement justifié
- Exclusion médicale contractuelle
- Fausse déclaration

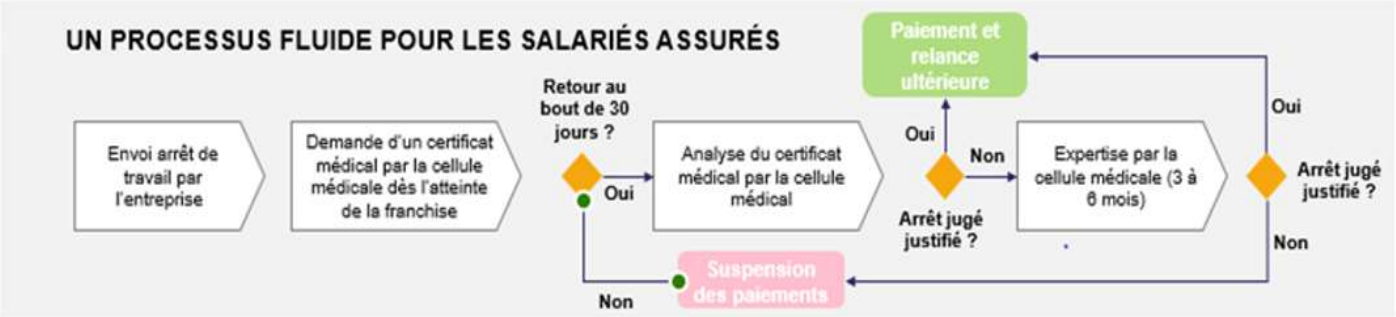
A noter que si le CMIT n'est pas renvoyé par l'assuré avant le 21^{e} jour, une relance est alors effectuée.

Une attention particulière sera portée aux arrêts de travail

- Ne présentant pas de date de reprise prévue
- Avec une pathologie psychiatrique (maladie non objectivable)
- Avec des durées particulièrement longues, au regard de la pathologie

Ce dernier point nécessite ainsi une étude actuarielle et la construction de lois expérimentales par pathologie.

Processus de contrôle des arrêts de travail



Source : Malakoff Humanis

2.2 Présentation des données

2.2.1 Présentation des variables

Le point de départ de notre étude est la base de données extraite par le service médical contenant les différentes pathologies. La période d'observation retenue pour l'ensemble de l'étude correspond aux événements survenus entre le 01/01/2016 et le 31/12/2021.

Par ailleurs, nous étudierons uniquement les arrêts de travail pour incapacité temporaire. Les incapacités permanentes de travail (invalidités) ne seront pas traitées. Ainsi les arrêts de travail étudiés sont exclusivement des incapacités.

Les variables renseignées dans les bases de données et qui seront utilisées sont :

Variables	Précisions
Numéro_Externe_RP	Numéro de l'arrêt de travail permettant l'identification du sinistre
SIREN	Permet l'identification de l'entreprise
Date_naissance_assuré	Date de naissance de l'assuré
Pathologie	Pathologie à l'origine de l'arrêt de travail
SEXE	Sexe de l'assuré
CSP	Catégorie socioprofessionnelle de l'assuré au moment de l'arrêt
Motif	Motif de l'arrêt de travail
Salaire	Salaire de l'assuré au moment de l'arrêt de travail (annuel brut)
Date_de_survenance	Date de survenance de l'arrêt de travail
Date_fin_Période	Date de fin de période de paiement
Date_deb_Période	Date de début de période de paiement

2 variables vont être rajoutées à notre base de données

- L'âge calculé à partir de la date de naissance
- La famille de pathologie

L'âge calculé et utilisé dans la suite de ce mémoire correspond à l'âge de l'assuré au moment où survient son incapacité de travail.

Notre base de données contient 448 pathologies différentes que nous regroupons à l'aide de l'expertise fournie par le Service Médical Conseil en 21 familles de pathologie.

2.2.2 Classification des pathologies en familles de pathologie

Les modalités de la variable pathologie étant trop nombreuses, on se propose de regrouper ces modalités en différentes familles de pathologie. A l'aide d'une expertise médicale on effectue la classification avec les familles de pathologie suivantes. Cette classification a été validé par le médecin conseil

Les différentes familles de pathologie créées

Familles de pathologie
Pathologies psychiatriques
Pathologies rhumatologiques
Traumatologie
Cancers
Pathologies gynéco-obstétriciennes
Pathologies neurologiques
Pathologies cardiovasculaires
Pathologies appareil digestif
Coronavirus
Tumeurs bénignes
Pathologies endocrinologiques
Pathologies respiratoires
Pathologies ORL
Pathologies pneumologiques
Pathologies néphrologiques
Pathologies ophtalmologiques
Pathologies hématologiques
Pathologies infectieuses
Pathologies dermatologiques
Pathologies systémiques e auto immunes
Pathologies urologiques

A titre d'exemple, la famille « Pathologies psychiatriques » regroupe toutes les pathologies suivantes :

Pathologie	Classe Pathologie
Anorexie mentale	Pathologies psychiatriques
Dépression nerveuse	Pathologies psychiatriques
Schizophrénie et psychoses	Pathologies psychiatriques
Spasmophilie Tétanie	Pathologies psychiatriques
Troubles fonctionnels d'origin	Pathologies psychiatriques
Attaque de panique	Pathologies psychiatriques
Burn out	Pathologies psychiatriques
Boulimie	Pathologies psychiatriques
Dépendance - sevrage - ethylisme - toxicomanie	Pathologies psychiatriques
Névrose	Pathologies psychiatriques
Psychose (bouffées délirantes aigues, psychose hallucinatoire,etc.)	Pathologies psychiatriques
Schizophrénie	Pathologies psychiatriques
Troubles anxieux et phobiques	Pathologies psychiatriques
Troubles de l'humeur- Bipolarité	Pathologies psychiatriques
Anorexie	Pathologies psychiatriques
Troubles obsessionels compulsifs (TOC)	Pathologies psychiatriques

Certaines maladies sont dites « non-objectivables ». Il s'agit d'affections physiques ou psychiques considérées comme non qualifiables par des professionnels de santé.

Une personne qui souffre d'une telle maladie nécessite néanmoins un suivi thérapeutique constant, voire des arrêts de travail plus ou moins longs et plus ou moins à répétition.

Parmi les maladies non objectivables les plus courantes, on retrouve :

- La fatigue chronique
- Les diverses pathologies psychiques et états dépressifs
- Les affections psychosomatiques
- L'épuisement professionnel et le burn-out
- Les pathologies du dos comme la lombalgie, la sciatique, la hernie discale…

2.3 Contrôles de cohérence des données

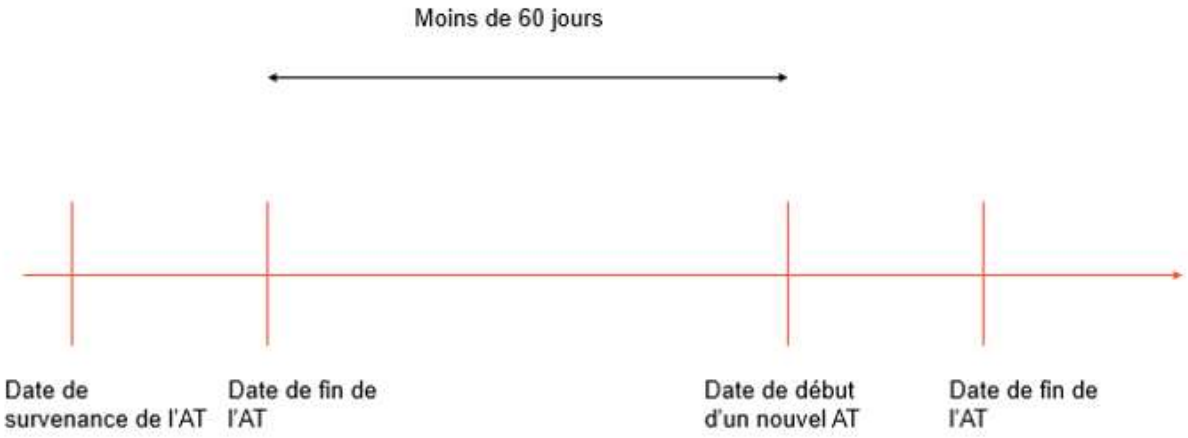
Différents types de contrôles ont été effectués pour rendre les données les plus fiables possibles. Les extractions ont été faits via SAS alors que les travaux de fiabilité de données ont été réalisés sous L'étude se porte sur la base des arrêts de travail (AT) pour lesquels une pathologie a été renseignée.

2.3.1 Traitement des rechutes

Traitement des rechutes :

Tout nouvel arrêt de travail qui survient après une reprise de travail de moins de 60 jours suite à un précèdent arrêt ayant la même pathologie est comptabilisé comme un seul même arrêt de travail. On parle d'une rechute.

Graphique expliquant la classification d'un sinistre en « rechute »



2.3.2 Critère d'unicité :

Il est également indispensable de vérifier le critère d'unicité pour ainsi éviter les doublons qui viendraient perturber l'étude.

Pour cela, on détermine notre clé primaire constituée de la date de survenance de l'arrêt de travail, de la date de naissance et du SIREN de l'entreprise.

Ces 3 données permettent d'identifier de manière unique chaque arrêt de travail. On procède ainsi à une suppression du doublon afin de ne pas introduire un biais dans l'analyse de données.

2.3.3 Traitement des valeurs manquantes et des valeurs aberrantes

Afin d'avoir la base de données la plus propre, nous supprimons de notre base les données incohérentes.

La date de début et la date de fin d'indemnisation pour l'assuré i sont posées comme étant :

- $dt\_deb\_indem_i = \min(Date\_deb\_Période_i)$
- $dt\_fin\_indem_i = \max(Date\_Fin\_Période_i)$

Nous vérifions successivement que la date de survenance ainsi que la date du début d'indemnisation sont bien inférieures à la date de fin de l'arrêt. Dans le cas où la date de début et la date de fin de l'arrêt de travail seraient identiques, la durée de l'arrêt de travail est nulle. Ces AT ne sont pas comptabilisés et sont surement dus à des erreurs internes de saisies. De manière analogue, si la date de survenance du sinistre est supérieure à la date de début d'indemnisation de l'arrêt, ces données sont retirées de notre base.

Nous supprimons également de notre base de données le cas de figure où la date de survenance de l'arrêt de travail est inférieure à la date de naissance de l'assuré.

Schéma représentant la cohérence temporelle des différentes dates de notre étude

De plus, nous sélectionnons dans notre base uniquement les arrêts dont la durée d'indemnisation est inférieure à 1096 jours. Seul cas contraire, les scénarios de rechutes comptabilisés comme des arrêts de travail et dont la durée d'indemnisation peut être supérieure à 1096 jours.

Les données ne vérifiant pas les points suivants seront également supprimées :

- Assuré présentant dans nos bases 2 dates de naissance différentes ou 2 sexes différents
- Date de naissance de l'assuré, date de survenance du sinistre et date de sortie de l'état incapacité non renseignées
- Pas de montant d'IJ (ce qui indique qu'il n'y a pas eu de prestations)

Notre base finale retraitée et nettoyée est constituée de 56 489 observations sur la période 2016-2021.

2.4 Statistiques descriptives

Réalisons des statistiques descriptives sur quelques variables de notre base de données.

2.4.1 Statistiques par année de survenance

Nombre d'arrêts de travail par année de survenance

Année de survenance	Nombre d'arrêts de travail	Pourcentage
2016	4 372	7,8%
2017	6 031	10,7%
2018	7 811	13,8%
2019	8 507	15,1%
2020	13 317	23,5%
2021	16 451	29,1%
Total	56 499	100%

La volumétrie est beaucoup plus importante sur les années 2020 et 2021. Ceci est dû à la différence de politique de relance des CMIT à partir de 2018. De plus, les délais de traitement des CMIT sont de plus en plus rapides.

2.4.2 Statistiques par famille de pathologie

Afin d'avoir des groupes représentatifs, un regroupement de familles de pathologie est effectué. Dans notre étude, nous nous intéresserons uniquement aux pathologies les plus représentées. Les autres seront regroupées au sein de la modalité : « Autres pathologies »

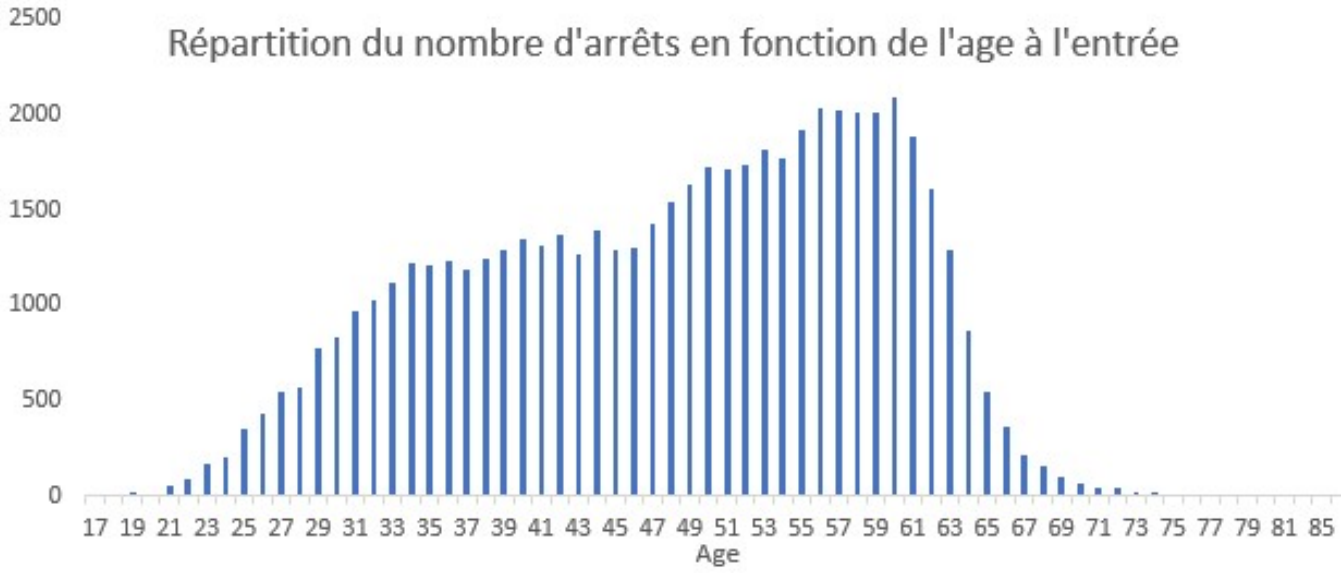
En excluant les données censurées au 31 décembre 2021, nous regarderons la durée moyenne des arrêts en fonction des différentes pathologies.

Répartition des différentes familles de pathologie au sein de notre portefeuille.

Famille de pathologie	Nombre d'arrêts de travail	Pourcentage	Durée moyenne (en jours)
Pathologies psychiatriques	16 907	29.9%	317
Pathologies rhumatologiques	10 858	19.2%	303
Traumatologie	9557	16.9%	285
Autres pathologies	5527	9.8%	289
Cancers	5215	9.2%	455
Pathologies gynéco-obstétriciennes	3663	6.5%	134
Pathologies neurologiques	2070	3.7%	367
Pathologies cardiovasculaires	1963	3.5%	331
Coronavirus	729	1.3%	97
TOTAL	56489	100%	306

La durée des arrêts de travail associés à la pathologie « Cancers » est la plus élevée. Remarquons également, les pathologies psychiatriques d'une durée assez élevée : 317 jours en moyenne.

2.4.3 Statistiques par tranche d'âge



La figure ci-dessus nous illustre la répartition du nombre d'arrêts de travail en fonction de l'âge d'entrée. Il est à remarquer que les arrêts de travail sont principalement répartis sur les âges élevés. L'âge moyen d'entrée en incapacité de notre portefeuille est de 56 ans.

4 tranches d'âge sont construites : Moins de 40 ans/ 40-49 ans/ 49-55 ans/ plus de 55 ans. Nous allons étudier la répartition des différentes familles de pathologie au sein de ces tranches d'âge.

Les résultats sont synthétisés dans le tableau suivant :

Tranche d'âge	Nombres d'arrêts	Pourcentage	Durée moyenne (en jours)
Moins de 40 ans	14 532	25,7%	304
40-49 ans	12 234	21,7%	297
49-55 ans	12 289	21,7%	306
Plus de 55 ans.	17 434	30,9%	349
Total	56 489	100%	306

Les durées des arrêts de travail sont en moyenne plus élevées pour les personnes de plus de 55 ans.

2.4.4 Statistiques par sexe

Sexe	Nombres d'arrêts	Pourcentage	Durée moyenne (en jours)
Hommes	21 782	38,5%	324
Femmes	34 707	61,5%	293
Total	56 489	100%	306

Notre portefeuille est disproportionné avec beaucoup plus de femmes que d'hommes. Les durées moyennes d'arrêts de travail des femmes sont en moyenne inférieures à celles des hommes.

2.4.5 Statistiques par catégorie socio-professionnelle

CSP	Nombres d'arrêts	Pourcentage	Âge Moyen	Part d'hommes	Part de femmes	Durée moyenne (en jours)
Cadres	10 744	19%	51 ans	50,09%	49,91%	307
NCA	45 745	81%	47,24 ans	35,85%	64,15%	301
Total	56 489	100%	47,95 ans	38,56%	61,44%	306

Notre portefeuille est inégalement réparti entre les non-cadres (NCA) extrêmement majoritaires et les cadres. Notons également le fait que la population cadre est en moyenne légèrement plus âgée.

2.4.6 Statistiques par origine du sinistre

Origine du sinistre	Nombres d'arrêts	Pourcentage	Durée moyenne (en jours)
Privée	51 338	90,9%	303
Professionnelle	5 151	9,1%	346
Total	56 489	100%	306

Les incapacités de travail présentes dans notre portefeuille sont engendrées principalement par une maladie ou un accident dans le cadre de la vie privée.

Nos données ne sont pas uniformes et notre population semble présenter une certaine hétérogénéité. De plus certaines variables comme l'âge ou la famille de pathologie semblent avoir un impact non négligeable sur la durée de l'arrêt de travail

2.5 Segmentation par pathologie

Afin de vérifier statistiquement que certaines de nos variables sont liées à la durée de maintien en incapacité, nous effectuons un test du Khi-deux d'indépendance.

2.5.1 Test du Khi-Deux

Le test d'indépendance du Khi-deux est un test statistique permettant d'analyser la corrélation entre 2 variables. Il n'est pas forcément obligatoire pour ces variables d'avoir le même nombre de modalités.

Les 2 hypothèses testées sont

H_0 : les 2 variables choisies sont indépendantes

H_1 : les 2 variables choisies ne sont pas indépendantes

Pour les 2 variables que l'on veut tester, le test se base sur un tableau de contingence à double-entrée recensant les effectifs observés.

Exemple de tableau de contingence pour 2 variables A et B

A|B	B_1	B_2	B_3	Total
A_1	n_{11}	n_{12}	n_{13}	$n_{1\_}$
A_2	n_{21}	n_{22}	n_{23}	$n_{2\_}$
A_3	n_{31}	n_{32}	n_{33}	$n_{3\_}$
A_4	n_{41}	n_{42}	n_{43}	$n_{4\_}$
Total	$n_{\_1}$	$n_{\_2}$	$n_{\_3}$	N

L'intersection de la i^{eme} ligne et de la j^{eme} colonne, n_{ij} correspondant aux observations dont les variables $A\ et\ B$ prennent respectivement les modalités A_i et B_j

On a également :

- $n_{\_j} = \sum_{j=1}^{p} n_{ij}$
- $n_{i\_} = \sum_{i=1}^{q} n_{ij}$

On pose $x_{ij} = \frac{n_{i\_} \times n_{\_j}}{N}$ l'effectif théorique si A et B sont indépendants

La statistique correspondant au test du Khi-deux observée peut, ensuite, être calculée via la formule :

$$\chi^2{}_{obs} = \sum_{i=1}^{p}\sum_{j=1}^{q} \frac{(n_{ij} - x_{ij})^2}{x_{ij}} = \frac{(Effectifs\ observés - Effectifs\ théoriques)^2}{Effectifs\ théoriques}$$

$\chi^2{}_{obs}$ suit sous l'hypothèse H_0 une loi du Khi deux à (p-1) (q-1) degrés de liberté :

$$\chi^2{}_{obs} \sim \chi^2{}_{(p-1)(q-1)}$$

En comparant la statistique calculée avec $\chi^2{}_{(p-1)(q-1)}$ que l'on lit dans la table du Khi-deux suivant les degrés de liberté, l'hypothèse d'indépendance des variables est acceptée ou rejetée.

Si :

- $\chi^2{}_{obs} > \chi^2{}_{(p-1)(q-1)}$ l'hypothèse d'indépendance H_0 est rejetée au risque d'erreur α
- $\chi^2{}_{obs} > \chi^2{}_{(p-1)(q-1)}$ l'hypothèse d'indépendance est acceptée au risque d'erreur α

La variable de référence sera la durée de maintien en incapacité (répartie par tranche) pour laquelle un test d'indépendance de Khi-deux sera réalisé avec successivement les variables suivantes :

- Famille_de_pathologie
- Age, pour laquelle des tranches d'âge ont été utilisées
- MOTIF_SIN
- SEXE
- CSP

Les différents tests seront modélisés via le langage R à travers la notion de p-value correspondant à la probabilité que le Khi-Deux soit supérieur à l'observé $\chi^2{}_{obs}$.

Pour un risque d'erreur α = 5%, l'hypothèse d'indépendance sera rejetée si la p-value est inférieur à 5%

Les résultats obtenus sont les suivants :

Test d'indépendance du Khi-Deux pour la durée de maintien en incapacité

Variable	p-value	Décision
Famille de pathologie	<2,2e-16	l'hypothèse H_0 est rejetée
Age	<2,2e-16	l'hypothèse H_0 est rejetée
MOTIF_SIN	<2,2e-16	l'hypothèse H_0 est rejetée
SEXE	0,1184	l'hypothèse H_0 est validée
CSP	0,1324	l'hypothèse H_0 est validée

Les variables « Famille_de_pathologie », « Age » et « MOTIF_SIN » sont significatives. Ainsi, il y a bien une corrélation entre ces variables et la durée de maintien en incapacité. Cependant, quelle est la variable la plus discriminante ?

Le mécanisme d'inversion du cycle de production place les assureurs dans un état qui les oblige à prédire le risque afin de tenir leurs engagements vis-à-vis des assurés. C'est pour cette raison qu'à travers un arbre de régression CART ((Classification And Regression Trees) nous allons déterminer et classifier les variables les plus discriminantes permettant d'expliquer le risque.

De plus, cette méthode met en lumière les différents sous-groupes permettant ainsi d'identifier l'hétérogénéité du portefeuille. Nous illustrerons également le caractère discriminant de la variable liée la pathologie.

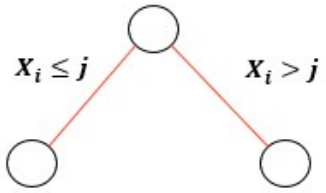
2.5.2 Modélisation CART

La modélisation CART introduite par Breiman en 1984 désigne une méthode statistique permettant de construire des arbres de décision. Le but de l'arbre est de partitionner la population initiale en sous-populations. On parle d'arbres de classification dans le cas d'une variable réponse qualitative et d'arbres de régression si la variable réponse est quantitative.

Un arbre est constitué d'un ensemble de nœuds. La totalité des observations est regroupée au niveau du nœud initial appelé la racine. Un arbre binaire partitionne chaque nœud en 2 sous-populations de manière à obtenir des échantillons les plus homogènes.

Le but de chaque division est de réduire au maximum l'hétérogénéité du groupe. Lorsque le partitionnement est interrompu, les nœuds finaux sont appelés des feuilles auxquelles sont attribuées des valeurs si la variable à expliquer est quantitative, une classe si la variable à expliquer est qualitative.

Schéma de subdivision d'un nœud



Le critère le plus utilisé pour arrêter la division est le nombre minimal des échantillons dans un nœud. Si l'effectif du nœud est inférieur à un seuil fixé, on ne divise plus et on considère le nœud comme une feuille.

Dans notre étude, la variable à expliquer est la durée de l'arrêt de travail et les différentes variables explicatives sont l'âge de l'assuré à la date de survenance, le motif du sinistre, le sexe, la catégorie socio-professionnelle et la famille de pathologie.

La méthode CART nous aide à sélectionner les facteurs qui expliquent le plus fortement la durée de maintien en incapacité.

Construction de l'arbre

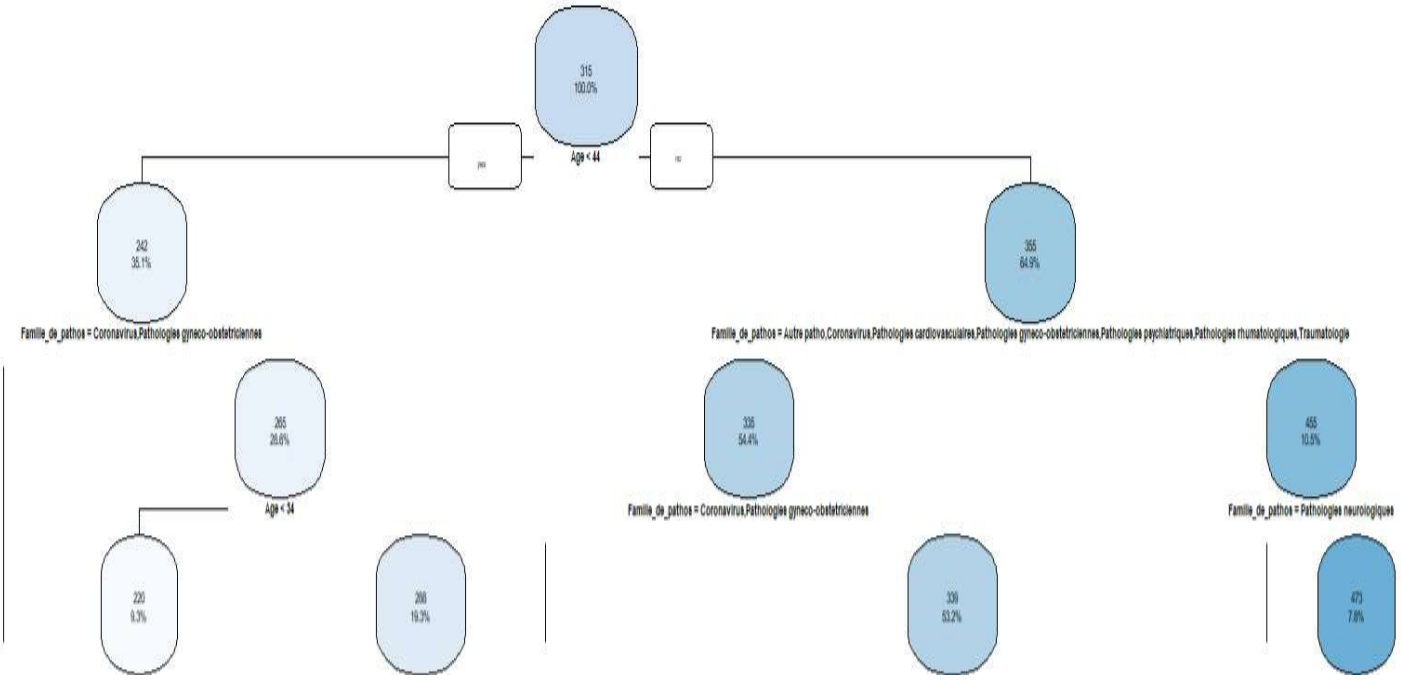
L'algorithme CART s'effectue en 3 étapes :

- Création de l'arbre maximal
- Élagage de l'arbre maximal
- Choix de l'arbre optimal parmi les arbres élagués.

Une première phase est la construction d'un arbre maximal, qui correspond à l'arbre le plus développé. Pour cela, nous devons définir un critère permettant d'identifier la meilleure division parmi toutes celles possibles pour les différentes variables ainsi qu'une règle permettant de décider qu'un nom est terminal.

De manière récursive, on construit une suite de partitions de l'espace des observations jusqu'à ce que chacun des éléments de la partition ne comporte qu'une seule observation. La librairie RPART sur R permet de construire cette forme d'arbre. Généralement, ce type d'arbre entraine des problèmes de surapprentissage. C'est pourquoi une seconde phase, dite d'élagage, est souvent nécessaire. Elle construit une suite de sous-arbres optimaux élagués de l'arbre maximal. L'objectif de cette étape est de réduire la complexité de l'arbre évitant ainsi les problèmes de surajustement. Enfin, l'arbre optimal est ensuite choisi parmi les différents arbres élagués.

Application du modèle CART à la base de données :



La figure ci-dessus présente le résultat d'un arbre de régression obtenu pour le risque de maintien en incapacité. Le résultat de cette modélisation CART nous permet de remarquer qu'à priori l'âge de l'assuré ainsi que la famille de pathologie sont les 2 variables qui apparaissent le plus souvent et le plus en amont dans la subdivision des nœuds. Cependant, on ne peut pas réellement en conclure que ces 2 variables sont les plus discriminantes.

En effet, à chaque étape de division des nœuds, nous choisissons la division qui maximise la variation d'hétérogénéité. Il serait possible, qu'une autre division avec une variation d'hétérogénéité légèrement plus faible soit possible avec une autre variable. Ce sont les variables cachées. On va ainsi utiliser l'algorithme des forêts aléatoires pour confirmer les observations.

Les algorithmes CART présentent plusieurs avantages, ils sont faciles à implémenter et sont fonctionnels avec des variables qualitatives ou quantitatives. De plus, la présentation des résultats sous forme d'arbres facilite la compréhension et l'interprétation.

2.5.3 Forêts aléatoires

Les forêts aléatoires (ou Random Forests) introduites par Reiman en 2001, se basent sur un ensemble d'arbres de décision indépendants. Cet algorithme de prédiction repose sur un 2 tirages aléatoires :

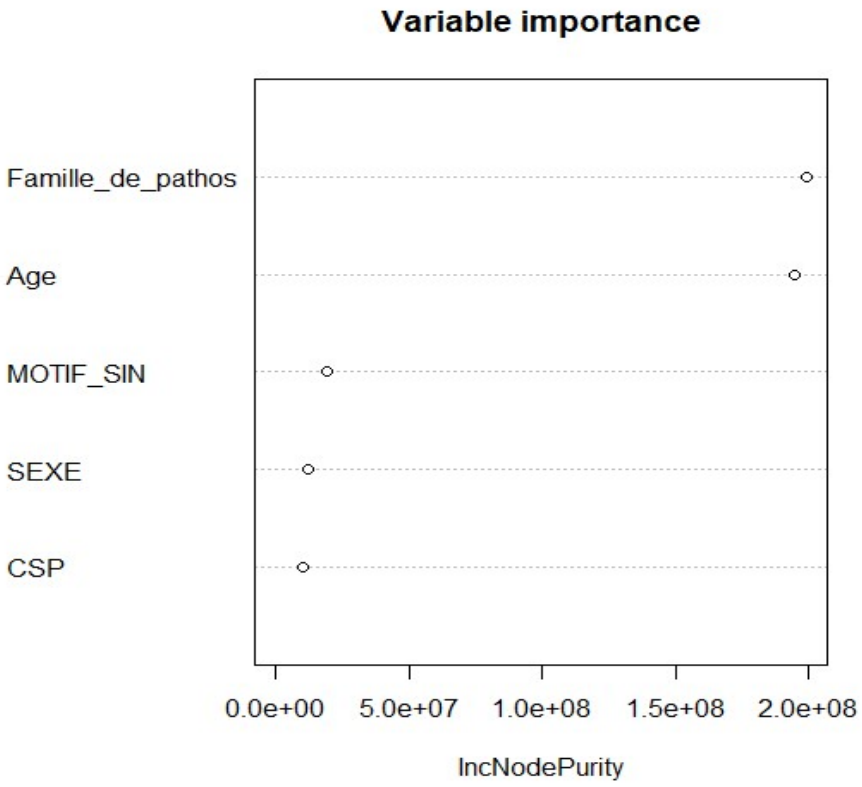
- Un tirage aléatoire sur les variables : c'est le feature sampling
- Un tirage aléatoire (avec remise) sur les observations : c'est le tree bagging

Hormis la prédiction de la variable à expliquer, les algorithmes de forêts aléatoires fournissent d'autres informations comme l'estimation de l'erreur de prédiction (ou erreur Out of bag) ou la mesure de l'importance des variables explicatives.

Estimation de l'importance des variables

Afin de déterminer l'importance d'une variable explicative, les différentes valeurs de cette modalité vont être permutées aléatoirement. Les valeurs prises par les autres variables ne vont pas être modifiées. L'impact de ces permutations sera évalué sur les différents arbres prédits.

L'impact est déterminé en comparant la marge d'erreur de l'arbre avant et après les permutations. Plus l'impact est important, plus la variable sera importante. On obtient les résultats suivants.

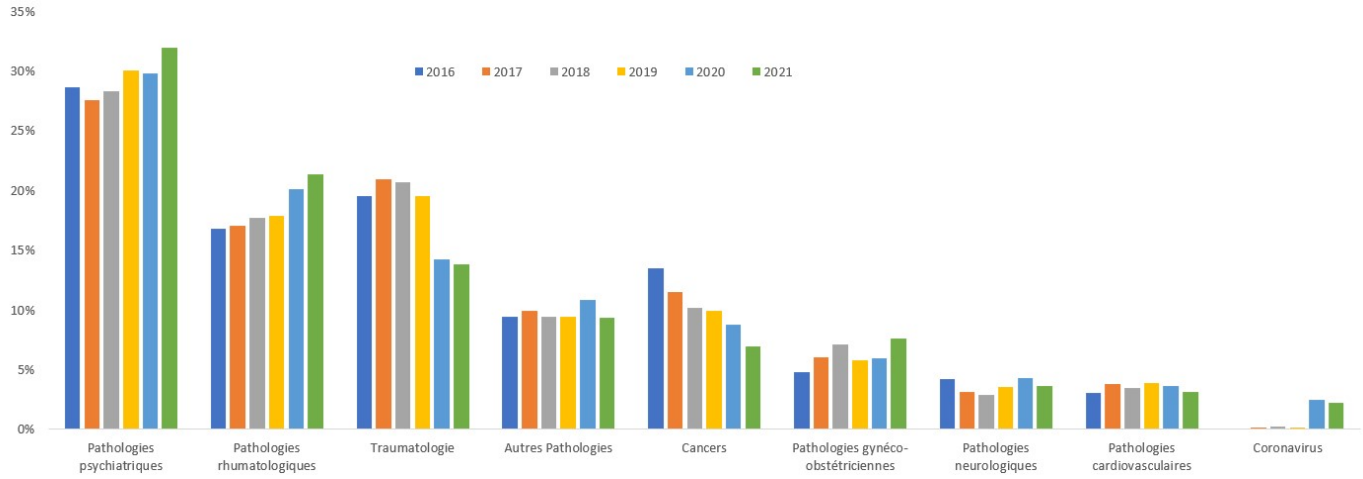


Segmenter notre loi de maintien en incapacité par famille de pathologie semble pertinent.

2.5.4 Analyse de la variable de segmentation

Des statistiques supplémentaires sont effectuées sur les familles de pathologie.

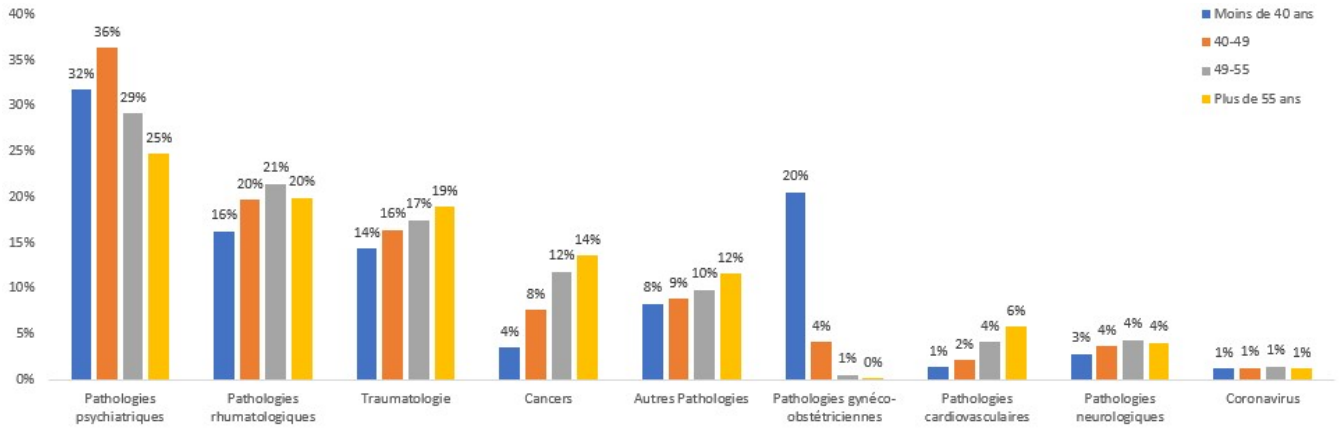
Répartition des pathologies par années de survenance



La part des différentes pathologies (excepté Covid 19) est assez stable dans le temps. Nous remarquons cependant, que la part des pathologies psychiatriques ainsi que la part des pathologies rhumatologiques augmentent légèrement au fur et à mesure des années.

Au contraire, nous constations une décroissance de la portion de notre portefeuille correspondant aux pathologies « Cancers » et « Traumatologie »

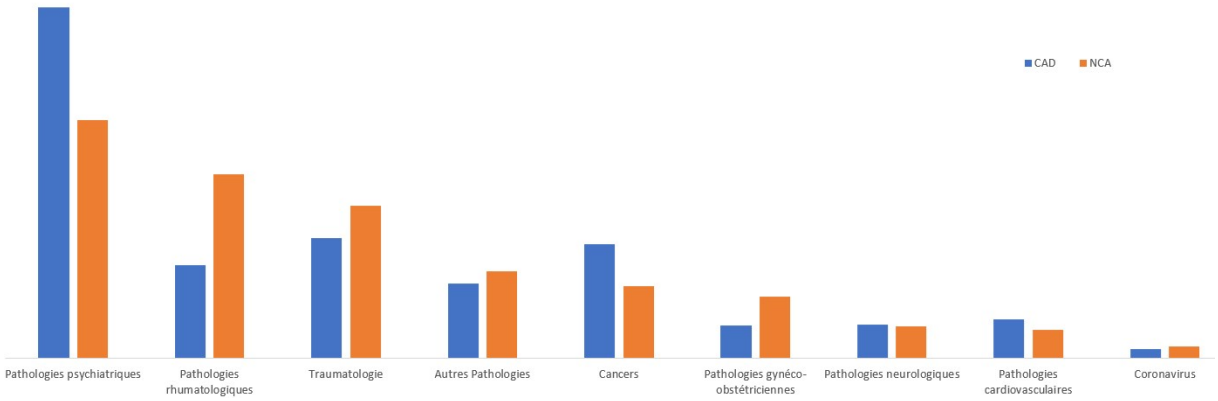
Répartition des familles de pathologie par tranche d'âge



Quelle que soit la tranche d'âge, la part de la famille « Pathologies psychiatriques » est la plus importante. Elle est la plus élevée pour les 40-49 ans.

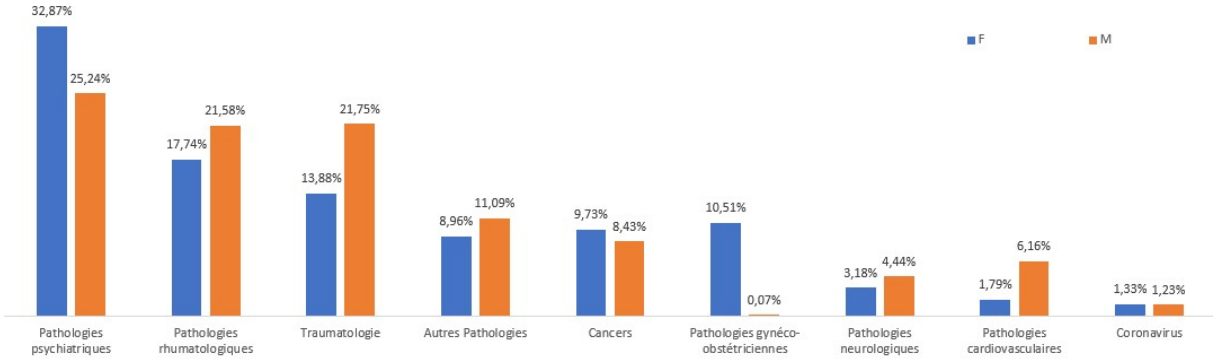
La part des familles « Traumatologie », « Cancers » et « Pathologies cardiovasculaires » tend à augmenter avec l'âge

Répartition des familles de pathologie par CSP



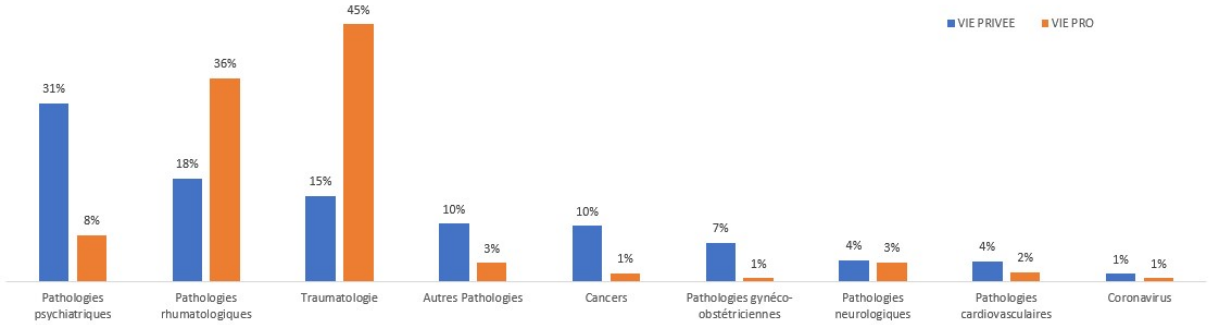
La part des « Pathologies rhumatologiques » est beaucoup plus élevée chez les non-cadres. En revanche, la part des « Pathologies psychiatriques », est particulièrement plus importante chez les cadres. C'est aussi le cas, dans une moindre mesure pour les « Cancers ».

Répartition des familles de pathologie par sexe



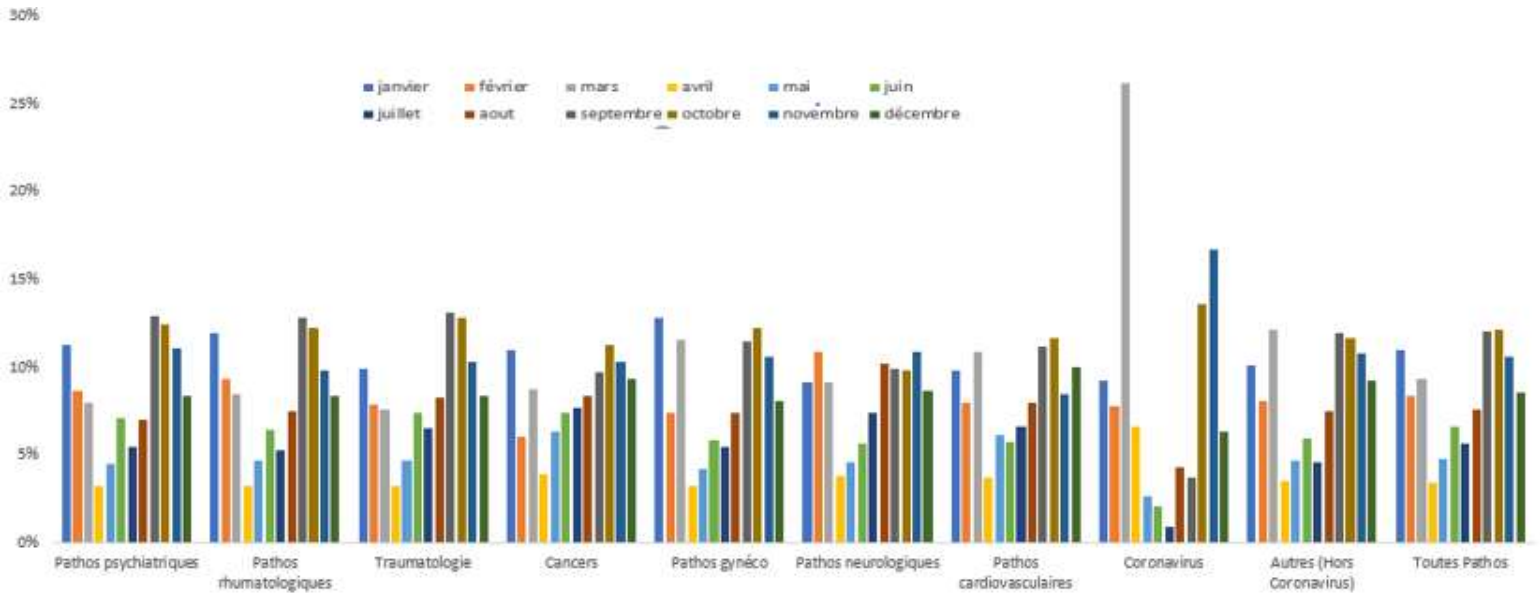
La part des familles « Pathologies psychiatriques » (et « Pathologies gynéco-obstétriciennes ») est plus importante chez les femmes. En revanche, la part des familles « Pathologies rhumatologiques », « Traumatologie » et « Pathologies cardiovasculaires » est plus élevée pour les hommes.

Répartition des familles de pathologie par origine des arrêts



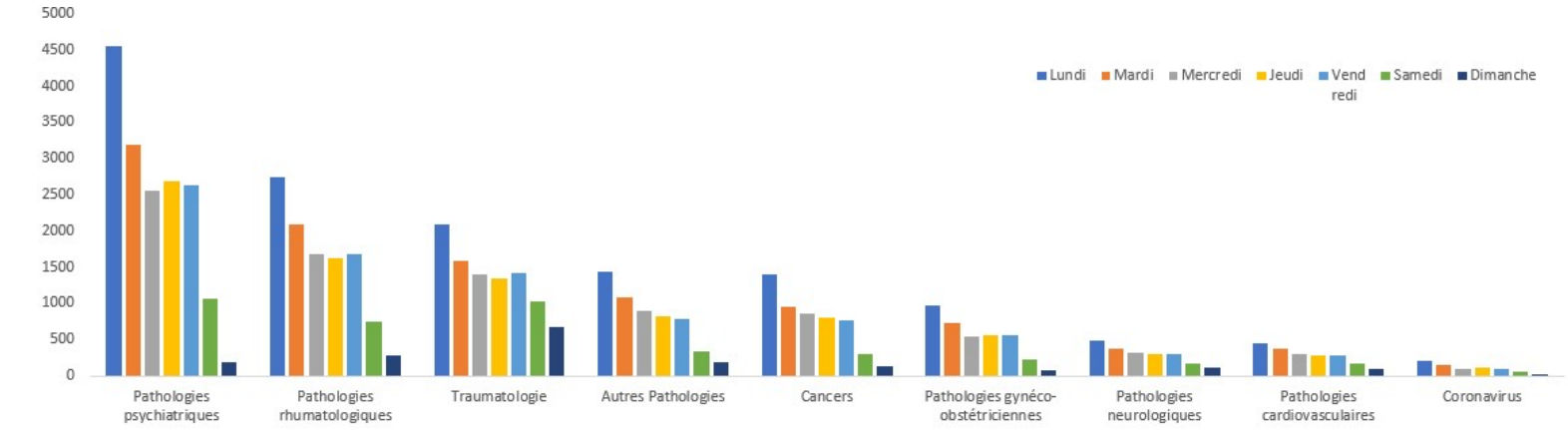
La part des familles « Pathologies rhumatologiques » « Traumatologie » est beaucoup plus élevée pour les arrêts vie professionnelle. En revanche, la part des familles « pathologies psychiatriques »' et « Cancers » est beaucoup plus élevée pour les arrêts vie privée.

Saisonnalité mensuelle



Il n'y a pas de saisonnalité particulière (hors Covid), la proportion du nombre de sinistres déclarés pendant les périodes de vacances reste moins importante par rapport à d'autres périodes de l'année. On observe un pic du nombre d'incapacité pour les mois de septembre et octobre.

Saisonnalité hebdomadaire



Indépendamment de la pathologie, les arrêts de travail sont beaucoup plus fréquents le lundi. 2 explications possibles : la fermeture des cabinets médicaux le week-end et les arrêts de complaisance. Le dimanche est également surreprésenté pour les familles « Traumatologie », « Pathologies neurologiques » (dont AVC) et les « Pathologies cardiovasculaires » : ces pathologies nécessitant souvent les services d'Urgences.

Ainsi, on construira des lois de maintien en incapacité de travail par pathologie afin d'avoir une vision plus segmentée de notre portefeuille

3 Conception des lois de maintien

Nous constatons depuis plusieurs années une hausse du nombre d'arrêts de travail induisant ainsi une forte dérive en prévoyance. Afin de mieux comprendre la dynamique de notre portefeuille, nous proposons de modéliser des lois de maintien en incapacité par pathologie.

Ce chapitre a pour objectif de présenter la construction des estimateurs des lois de maintien en arrêt de travail. Dans un premier temps, nous nous intéresserons aux 2 estimateurs non-paramétriques les plus utilisés : l'estimateur de Kaplan-Meier et l'estimateur de Nelson-Aalen du taux de hasard cumulé. Notons également que les différentes durées modélisées sont des durées totales depuis le début de survenance.

3.1 Contexte de l'étude

3.1.1 Table de maintien en incapacité

Une table de maintien en incapacité est une table donnant pour une durée de 36 mois sur un groupe donné de 10 000 individus d'âge x à l'entrée en incapacité, les effectifs restant en incapacité chaque mois.

3.1.2 Censure

Au moment où les données de l'étude sont fixées, nous avons connaissance de la durée en arrêt de travail pour les personnes qui ne sont plus en incapacité.

Cependant, pour les assurés qui sont toujours en arrêt, nous ignorons la durée restante en incapacité. La seule information dont nous disposons est que la durée de l'arrêt est au moins égale à la durée observée.

La variable n'est pas observée directement, nous savons simplement que celle-ci appartient à un sous-ensemble. C'est le phénomène de censure.

$Soit\ X_1, X_2, \dots, X_n$ un échantillon de durée de survie. Il y a un phénomène de censure s'il existe n variables aléatoires i.i.d $C_1, \dots, C_n$ telles qu'au lieu d'observer l'échantillon de survie $X_1, \dots, X_n$ on observe à la place $(T_1, D_1), \dots, (T_n, D_n)$ avec :

- $T_i = \min(X_i, C_i)$
- $D_i = 1_{\{X_i \leq C_i\}}$

Notons également que les variables $C_1, \dots, C_n$ peuvent être aussi bien aléatoires que déterministes Ainsi :

- $D_i = 1$ si l'évenement est observé. Dans ce cas, $T_i = X_i$ et on observe les durées complètes

- $D_i = 0$ si l'évenement est censuré. Dans ce cas $T_i = C_i$ et on observe des durées incomplètes

Cette notion de censure (et plus tard de troncature) est illustrée à travers un exemple. Imaginons que l'on étudie la durée d'apprentissage d'une tâche chez les enfants. Lors de la fin de l'étude si certains enfants ne savent toujours pas accomplir la tâche alors l'information est censurée à droite par la durée de l'étude C. La seule information que l'on possède sera que la durée d'apprentissage chez ces enfants sera supérieure ou égale à la durée de l'étude.

3.1.3 Troncature

Certaines données sont en revanche déjà présentes au premier jour de notre étude (1er janvier 2016), ces individus sont observables seulement si la durée de leur arrêt est supérieure à une certaine valeur (souvent les franchises contractuelles). On parle alors de troncature.

Reprenons l'exemple précédent. Si au début de l'étude, il apparait que certains enfants ont déjà commencé à apprendre la tâche, alors les données sont tronquées (à gauche).

En somme, les notions de censure et de troncature sont illustrées à travers le schéma suivant :

Schéma récapitulatif des différentes situations

L'analyse de survie permet de traiter notre échantillon présentant à la fois des données complètes et des données censurées pour lesquelles on ne connait qu'un minorant (cas de la censure à droite). La méthode classiquement utilisée en entreprise est celle de Kaplan-Meier qui permet de construire la distribution empirique de la durée.

Dans notre étude, nous sommes en présence de données censurées à droite et tronquées à gauche. En effet, par exemple, si l'assuré est en arrêt de travail pendant quelques jours et qu'il ne dépasse pas la durée de franchise de son contrat, il est probable que celui-ci ne déclare pas son incapacité car dans ce cas l'assureur ne lui versera pas d'indemnités. La durée de cet arrêt de travail ne sera donc pas visible dans les bases.

Nous n'avons dans nos bases que les arrêts pour lesquels Malakoff Humanis a versé une prestation.

3.2 Bases mathématiques pour les estimations de la durée

Définissons quelques fonctions mathématiques qui seront utilisées pour estimer les durées de maintien en incapacité

3.2.1 Fonction de survie :

On définit la probabilité de survie comme la probabilité qu'un événement arrive après un temps t donné

$$S(t) = P(T > t) = 1 - F(t)$$

Lorsque la densité existe, on note

$$f(t) = \frac{d}{dt}F(t) = \lim_{h\to 0}(\frac{P(t < T < t + h)}{h})$$

3.2.2 Fonction de hasard :

On définit la fonction de hasard comme la probabilité de reprendre le travail dans un petit intervalle de temps après t, conditionnellement au fait d'avoir survécu jusqu'au temps t

$$h(t) = \frac{f(t)}{S(t)} = \frac{-S'(t)}{S(t)} = -\frac{d}{dt}ln(S(t))$$

On obtient ainsi une nouvelle expression de la fonction de survie associée à la fonction de hasard ; S(t)= $e^{-\int_0^t h(s)ds}$

En notant H(t)= -$\int_0^t h(s)ds$; la fonction de hasard cumulée, on obtient directement S(t)= $e^{-H(t)}$

3.3 Méthode de construction de Kaplan-Meier et propriétés

Il y a 3 grands types d'estimateurs pour modéliser une courbe de survie : non paramétriques, semi-paramétriques et paramétriques. Si notre loi de survie semble être distribuée selon une loi à priori connue, un modèle paramétrique sera utilisé. Les distributions les plus couramment utilisées pour les modèles paramétriques sont : la loi exponentielle, la loi de Weibull et la loi de Pareto.

Dans certains cas, on peut vouloir ne pas faire d'hypothèse sur la forme de la loi de survie. On estime directement cette fonction dans un espace infini. On utilise alors des estimateurs non-paramétriques. L'estimateur non-paramétrique le plus connu et le plus utilisé est celui de Kaplan-Meier.

Il existe également des estimateurs dits semi paramétriques. La loi de survie n'a pas de forme autodéterminée, mais est construite selon les différentes données que l'on possède. Le modèle de Cox est un modèle semi-paramétrique fréquemment utilisé en analyse de survie.

3.3.1 Présentation de la méthode de Kaplan- Meier

Le principe de l'estimateur de Kaplan- Meier repose sur l'idée que la probabilité t de survivre au-delà de t_2>t_1 peut s'écrire :

$$S(t_2) = P(T > t_2|T > t_1)\, P(T > t_1) = P(T > t_2|T > t_1)\, S(t_1)$$

En répétant ces opérations, on peut forcer l'apparition de terme de la forme $P(T > t_i|\, T > t_{i-1})$

En effet, si l'on choisit intelligemment les instants de conditionnements en prenant les instants ou l'on a soit une sortie soit une censure, cela revient à estimer des probabilités de la forme :

$$\mathrm{p_i} = \mathrm{P(T > t_i|\, T > t_{i-1})}$$

p_i est la probabilité conditionnelle d'être encore en arrêt de travail entre $T_{(i-1)}\, et\, T_{(i)}$ tout en sachant que l'on était en arrêt à l'instant $T_{(i-1)}$

En posant t_i correspondant aux différentes dates de sortie du portefeuille

La probabilité de survivre au-delà de $t_m > t_{m-1} > \ldots > t_1 > t_0$ peut s'écrire :

$$\begin{aligned}\mathrm{S(t_m)} &= \mathrm{P(T > t_m|T > t_{m-1})S(T > t_{m-1})}\\ &= \mathrm{P(T > t_m|T > t_{m-1})P(T > t_{m-1}|T > t_{m-2})\, S(T > t_{m-2})}\\ &\ldots..\\ &= \mathrm{P(T > t_m|T > t_{m-1})\ldots P(T > t_2|T > t_1)\, S(t_1)}\end{aligned}$$

On a $\mathrm{p_i = P(T > t_i|\, T > t_{i-1})}$

On définit $\mathrm{q_i} = 1 - \mathrm{p_i}$ représentant pour un assuré en arrêt de travail à l'instant $\mathrm{t_{i-1}}$ la probabilité de reprendre l'activité professionnelle en i

Un estimateur de $\mathrm{p_i}$ est $1 - \mathrm{q_i}$ avec $\widehat{\mathrm{q_i}} = \frac{\mathrm{d_i}}{\mathrm{n_i}}$

On pose $\mathrm{d_i}$ le nombre de sortie et $\mathrm{c_i}$ le nombre de censure et $\mathrm{tr_i}$ le nombre de troncature gauche, On a : $\mathrm{n_i} = \mathrm{n_{i-1}} - \mathrm{d_i} - \mathrm{c_i} - \mathrm{tr_i}$

Par récurrence, on se ramène à : $\widehat{\mathrm{S}}\,(\mathrm{t}) = \prod_{\mathrm{t_i<t}}\, (1 - \frac{\mathrm{d_i}}{\mathrm{n_i}})$

Cette expression correspond à l'estimateur de survie de Kaplan-Meier noté dorénavant $\widehat{S_{KM}}\,(t)$.

En résumé la méthode de Kaplan- Meier consiste en une succession d'estimations de survie (ici le maintien en arrêt de travail) qui est représentée par une courbe de maintien en arrêt en fonction du temps. L'avantage principal de la méthode de Kaplan- Meier est qu'elle prend en compte les données censurées dans l'estimation de survie, c'est-à-dire dans le cas de notre étude, les arrêts toujours en cours à la fin de la période d'observation (31 mai 2022). De plus la méthode de Kaplan- Meier permet de modéliser la durée totale (prenant en compte la franchise) et non la durée indemnisée.

Par convention, on suppose qu'en présence d'ex æquo les observations non censurées précèdent les observations censurées

<u>Exemple de construction pour l'estimateur de Kaplan-Meier</u>

Durée t_i	Etat	Nombre de sinistres d_i	Base restante à t-1 n_i	Probabilité de sortie	Loi de survie en t S(t)
1	Arrêt Sorti	50	10 000	$\frac{50}{10\,000} = 0{,}005$	0,995
2	Arrêt sorti	40	9 950	$\frac{40}{9\,950} = 0{,}0041$	$0{,}995 \times (1 - 0{,}0041)$ = 0,991
3	Arrêt sorti	40	9 910	0 ,0040	0,987
3	Arrêt Censuré	1	9 870	0,0001	
4	Arrêt sorti	29	9 869	0,0029	$0{,}987 \times (1 - 0{,}0029)$ = 0,984
5	Arrêt Sorti	20	9 840	0,0020	0,982

3.3.2 Propriétés de l'estimateur de Kaplan- Meier

L'estimateur de Kaplan -Meier est convergent, sans biais (si la dernière observation est non censurée), cohérent, asymptotiquement gaussien. De plus si aucune donnée n'est censurée, l'estimateur est égal à la fonction de survie empirique.

- **Cohérence :**

L'estimateur de Kaplan-Meier est cohérent. En effet, sa fonction de survie empirique estimée peut s'écrire :

$$\hat{\mathrm{S}}(\mathrm{t}) = \frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}} 1_{\{\mathrm{T_i}>\mathrm{t}\}} + \sum_{\mathrm{i}=1}^{\mathrm{n}} 1_{\{\mathrm{T_i}<\mathrm{t},\mathrm{D_i}=0\}} \frac{\hat{\mathrm{S}}(\mathrm{t})}{\hat{\mathrm{S}}(\mathrm{T_i})}$$

D_i Caractérise l'état de l'assuré ; Si $D_i = 1$, l'assuré a repris son activité professionnelle, si $D_i = 0$,l'observation est censurée

- **Biais de l'estimateur de Kaplan- Meier**

L'estimateur de Kaplan- Meier comporte un biais positif :

$$E[\,\hat{S}_{KM}(t) - S(t)] \geq 0$$

D'après les travaux de Fleming et Harrington en 1991, pour tout t tel que S(t) >0 :

$$\mathrm{E}[\,\hat{\mathrm{S}}_{\mathrm{KM}}(\mathrm{t}) - \mathrm{S}(\mathrm{t})] = \mathrm{E}[1_{\mathrm{t_n}<\mathrm{t}} \frac{\hat{\mathrm{S}}_{\mathrm{KM}}(\mathrm{t_n})\,(\hat{\mathrm{S}}_{\mathrm{KM}}(\mathrm{t_n}) - \mathrm{S}(\mathrm{t_n}))}{\mathrm{S}(\mathrm{t_n})}]$$

Cette expression vaut 0 si $t < t_n$ (L'indicatrice vaut 0 dans ce cas) ou si la dernière information est une donnée non censurée. Dans le cas contraire, l'estimateur est biaisé.

- **Asymptotiquement Gaussien**

En posant S la fonction de survie théorique, $\hat{S}$ la fonction de survie empirique et n le nombre d'arrêts de travail ; l'estimateur de Kaplan- Meier est asymptotiquement gaussien c'est-à-dire :

Si les fonctions de répartition de la survie et de la censure n'ont aucune discontinuité commune, alors

$$\sqrt{\mathrm{n}}\,(\widehat{S} - \mathrm{S}) \xrightarrow{\mathrm{L}} \mathrm{Z}$$

Avec Z un processus gaussien centré de fonction de covariance

$$Cov(Z(s), Z(t)) = S(t)S(s)\int_0^{\mathrm{t\^{}s}} \frac{dF(u)}{S(u)^2\,(1 - G(u))}$$

Avec F et G représentant les fonctions de répartition de la survie et de la censure

En somme, l'estimateur de Kaplan-Meier est cohérent, asymptotiquement gaussien et comporte un biais positif

Intervalle de confiance à 95%

L'estimateur de Kaplan- Meier est asymptotiquement normal de moyenne S(t).Ainsi,

$$\frac{[\,\hat{S}_x(t) - E[\hat{S}_x(t)]}{\sqrt{V(\,\hat{S}_x(t))}} \sim N(0,1)$$

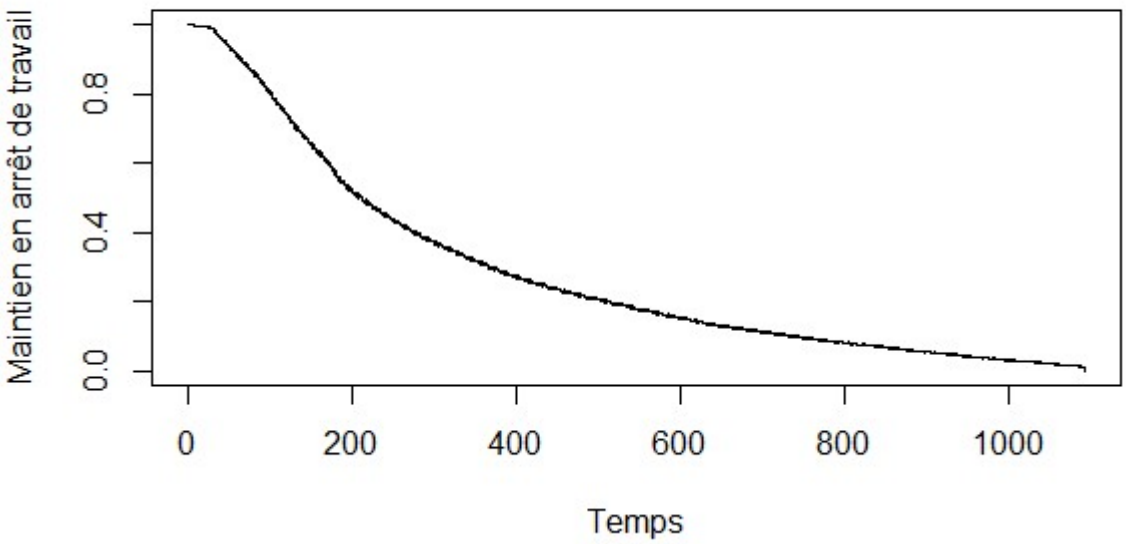
On en déduit ainsi l'intervalle de confiance à 95% de $S_x(t)$:

$$\text{IC} = \hat{S}_x(t) \pm 1{,}96 \times \sqrt{\hat{S}_x(t)}$$

3.3.3 Application de la méthode de Kaplan-Meier

On applique la théorie de Kaplan- Meier pour en déduire la courbe de maintien en arrêt de travail sur notre portefeuille. Indépendamment de la pathologie et pour l'ensemble de données (complètes ou censurées), on obtient le résultat suivant.

Loi de maintien globale

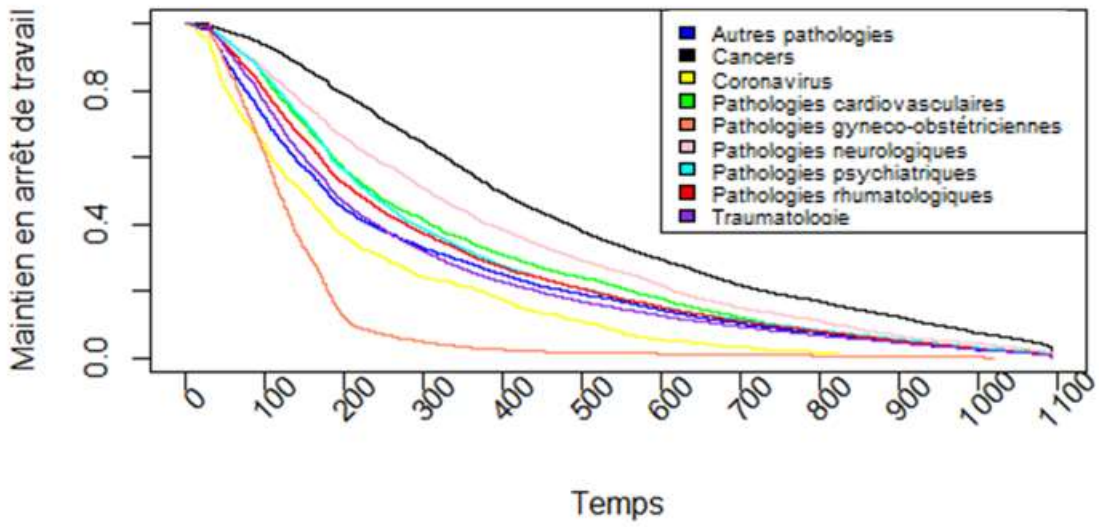


La courbe semble plutôt lisse avec cependant un saut correspondant à la marque des 1095 jours soit la durée totale de l'incapacité.

On peut constater que la fonction de survie décroit très rapidement et principalement lors de la première année. Ainsi la probabilité pour un assuré de se maintenir en incapacité plusieurs mois après le début de son indemnisation est relativement faible.

Afin de capter l'impact des différentes pathologies, nous segmentons dans un second temps notre loi de maintien par famille de pathologie. Nous obtenons le résultat suivant.

Loi de maintien par pathologie



Il apparait que les arrêts de travail associés à la pathologie « Cancers » sont plus longs. En revanche les arrêts de travail associés aux pathologies gynéco-obstétriciennes et dans une moindre mesure les pathologies « Coronavirus » sont plus courtes.

3.4 Estimateur de Nelson Aalen

3.4.1 Présentation de l'estimateur de Nelson-Aalen

Il existe un deuxième estimateur non paramétrique très populaire et très utilisé en entreprise : l'estimateur de Nelson- Aalen

L'estimateur dit de « Nelson-Aalen » est un estimateur non paramétrique qui peut être utilisé pour estimer la fonction de taux de hasard cumulé H. Cet estimateur prend parfaitement en compte les données censurées. Le taux de risque instantané h peut être estimé par la fonction suivante :

$$\hat{h}(t_i) = \frac{d_i}{n_i}$$

On en déduit ainsi l'estimation du taux de hasard cumulé

$$\widehat{H_{NA}}(t) = \sum_{t_{i\leq t}} \frac{d_i}{n_i}$$

On peut estimer la variance de l'estimateur de Nelson- Aalen en utilisant la théorie des processus stochastiques ainsi qu'une approximation par une loi de Poisson.

$$\widehat{Var}(\hat{H}_{NA}(t) = \sum_{t_i\leq t} \frac{d_i}{n^2{}_i}$$

3.4.2 Propriétés de l'estimateur de Nelson-Aalen

L'estimateur de Nelson-Aalen admet quelques propriétés intéressantes. Il est :

- **Sans Biais**

L'estimateur de Kaplan-Meier est biaisé et sous-estime en moyenne la fonction de hasard cumulée.

- **Asymptotiquement Gaussien**

L'estimateur de Kaplan- Meier est asymptotiquement gaussien.

On a plus précisément le résultat suivant :

S'il n'y a pas de discontinuité entre les fonctions de répartition de la survie et de la censure, alors :

$$\sqrt{n}\left(\hat{H} - H\right) \xrightarrow{loi} Z$$

Avec Z un processus gaussien centré de covariance $Cov\big(Z(s), Z(t)\big) = \int_0^{s\wedge t} \frac{dS_1(u)}{S_c(u)^2}$ avec $S_c(t) = \big(1 - F(t)\big)\big(1 - G(t)\big)$ et $S_1(t) = P(T > t, D = 1)$

En utilisant l'estimateur du taux de hasard cumulé de Nelson -Aalen, on peut se ramener à un estimateur de la fonction de survie en utilisant la relation suivante : $\widehat{S_{FH}}(t) = e^{-\hat{H}_{NA}(t)}$.

Il s'agit de l'estimateur de survie de Fleming-Harrington.

$\forall t > t$

$$\hat{S}_{HF}(t) = \exp\left(-\hat{H}_{NA}(t)\right)$$

$$= \prod_{t_i<t} \exp\left(-\frac{d_i}{n_i}\right)$$

En utilisant la delta méthode, on trouve le résultat suivant sur l'estimateur de la variance de $\hat{S}_{HF}$

$$\widehat{Var}\left(\hat{S}_{HF}(t)\right) = \left(\hat{S}_{HF}(t)\right)^2 Var(\hat{H}_{NA}(t))$$

3.5 Estimateur de Survie de Fleming-Harrington.

3.5.1 Présentation

Reprenons l'estimateur de Kaplan-Meier :

$$\widehat{S_{KM}}(t) = \prod_{t_i<t} (1 - \frac{d_i}{n_i})$$

En appliquant la transformation par le logarithme on obtient

$$ln\,\widehat{S_{KM}}(t) = \sum_{t_i<t} ln(1 - \frac{d_i}{n_i})$$

On définit l'estimateur d'Harrington-Fleming tel que

$$ln\,\widehat{S_{HF}}(t) = -\sum_{t_i} \frac{d_i}{n_i}$$

Ainsi nous obtenons le résultat suivant :

$$ln\,\widehat{S_{KM}}(t) - ln\,\widehat{S_{HF}}(t) = \sum_{t_i<t} ln(1 - \frac{d_i}{n_i}) + \frac{d_i}{n_i})$$

En posant x= $\frac{d_i}{n_i}$, il est possible de se ramener à une fonction du type f(x) = ln(1-x) + x qui est toujours négative. D'où :

$$\widehat{S_{KM}}(t) \leq \widehat{S_{HF}}(t)$$

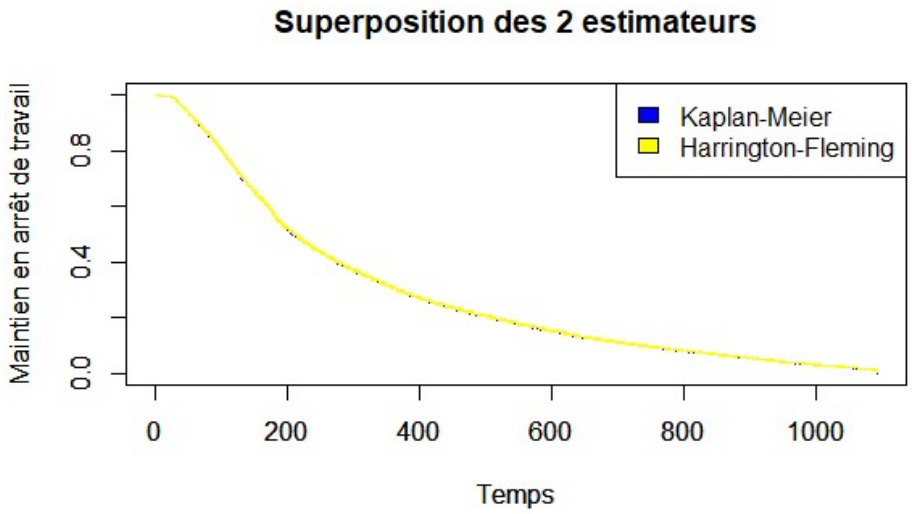
3.5.2 Construction et application

Exemple de construction pour l'estimateur d'Harrington-Fleming

Durée t_i	Etat	Nombre de sinistres d_i	Base restante à t-1 n_i	Probabilité de sortie $\frac{d_i}{n_i}$	Loi de survie en t S(t) $e^{-\frac{d_i}{n_i}}$
1	Arrêt Sorti	50	10 000	$\frac{50}{10\,000} = 0{,}005$	$e^{-0{,}005}$= 0,995
2	Arrêt sorti	40	9 950	$\frac{40}{9\,950} = 0{,}0041$	$0{,}995 \times (e^{-0{,}0041})$ = 0,991
3	Arrêt sorti	40	9 910	0 ,0040	0,987
3	Arrêt Censuré	1	9 870	0,0001	
4	Arrêt sorti	29	9 869	0,0029	$0{,}987 \times (e^{-0{,}0029}) = 0{,}984$
5	Arrêt Sorti	20	9 840	0,0020	0,982

L'estimateur de survie d'Harrington-Fleming est toujours supérieur à l'estimateur de Kaplan- Meier. L'estimateur d'Harrington-Fleming est intéressant si on veut une vision plus prudente de notre table d'expérience. Cette relation peut être vérifiée sur nos données

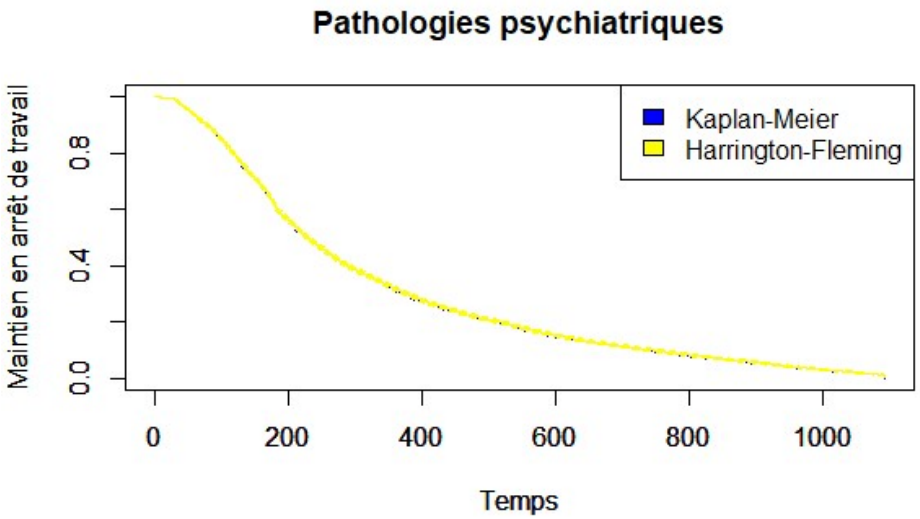
Comparaison des lois de survie de Kaplan Meier et d'Harrington-Fleming



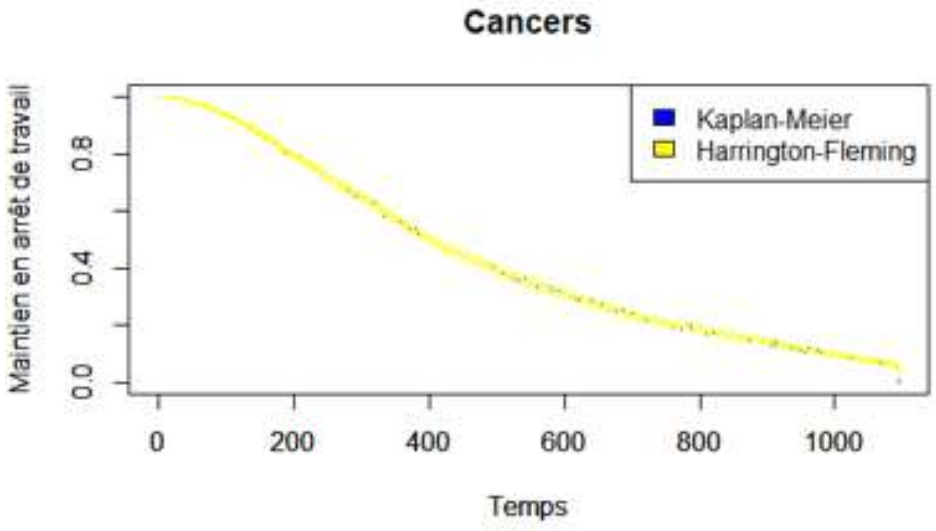
Les 2 fonctions de survie sont quasiment superposées. Cependant, il est possible de distinguer sur la courbe quelques points bleus en dessous des points jaunes.

L'estimation de survie de Kaplan Meier semble être supérieure à l'estimation de survie d'Harrington-Fleming. Cependant, qu'en est-il pour notre portefeuille segmenté par pathologie ?

Comparaison des lois de survie de Kaplan Meier et d'Harrington-Fleming pour les assurés présentant la pathologie psychiatrique comme motif d'incapacité.



Comparaison des lois de survie de Kaplan Meier et d'Harrington-Fleming pour les assurés présentant la pathologie « Cancers » comme motif d'incapacité.



Les résultats entre les 2 estimations sont quasiment équivalents. Notons remarquons cependant qu'à chaque instant et indépendamment de la pathologie l'estimateur de survie d'Harrington-Fleming semble légèrement supérieur à l'estimateur de survie de Kaplan-Meier

4 Prise en compte de variables explicatives par le modèle de Cox

Les statistiques descriptives précédemment effectuées ont permis de mettre en lumière l'influence de certaines variables explicatives (l'âge, la catégorie socioprofessionnelle, le sexe…) sur la durée de l'incapacité de travail.

Afin de prendre en considération l'hétérogénéité du portefeuille, on utilise généralement le modèle de régression de Cox (que l'on peut également appeler modèle à hasard proportionnel). Ce modèle permet de mesurer l'impact des covariables sur la loi de la variable aléatoire X pouvant être censurée. A la différence d'un modèle de régression linéaire classique, les covariables ne sont pas reliées directement à la variable à expliquer, mais ont un effet multiplicatif par rapport à la fonction de hasard.

4.1 Le modèle à hasard proportionnel

4.1.1 Théorie du modèle de Cox

Afin d'étudier l'influence de p vecteurs de variables explicatives sur la loi de la variable aléatoire qui peut potentiellement être censurée par une variable C, on pose $X_i = (X_{i_1}, X_{i_2}, \dots, X_{i_p})$ les valeurs des différentes covariables pour l'assuré i.

La fonction de risque du modèle de Cox s'écrit

$$h(t|X_i) = h_0(t)\exp\left(\beta_1 X_{i_1} + \cdots + \beta_p X_{i_p}\right)$$

$$= h_0(t)\exp(\ X_i\ .\ \beta)$$

$h_0(t)$ est la fonction de risque instantané de base (c'est-à-dire la fonction de risque obtenue en l'absence de variables explicatives).

Les coefficients β associés aux différentes variables explicatives sont déterminés par maximum de vraisemblance partielle. Nous testerons par la suite leur non-nullité à travers des tests statistiques.

Le modèle de Cox est dit « à hasard proportionnel ». En effet, le rapport des fonctions de risque instantané de 2 assurés est constant pendant toute la durée d'observation.

Considérons 2 assurés i et j.

$\forall\ i, j$

$$\frac{h(t|X_i)}{h(t|X_j)} = \frac{h_0(t)\exp(\ X_i\ .\ \beta)}{h_0(t)\exp(\ X_j\ .\ \beta)} = \exp(\ [X_i - X_j]\ .\ \beta)$$

Les risques des 2 assurés sont proportionnels.

Il s'agit de l'une des hypothèses fortes du modèle de Cox. Le rapport des risques instantanés de 2 assurés est indépendant du temps. C'est l'hypothèse des risques proportionnels.

Outre l'hypothèse des risques proportionnels, le modèle de Cox nécessite aussi la validation de l'hypothèse dite de log-linéarité :

$$\log(\ h(t|X_i\ ) = \log(\ h_0(t)\ ) + \beta X_i$$

Le logarithme de la fonction de risque instantanée est donc une fonction linéaire des X_i. Ces 2 hypothèses fondamentales devront être respectées afin d'appliquer notre modèle de Cox à nos données. Le rapport des différents risques instantanés nous permet également de mieux interpréter notre modèle et de mieux comprendre les spécificités de notre portefeuille.

Distinguons 2 cas de figure.

Si notre variable d'intérêt est qualitative (ou quantitative et discrète) :

Prenons par exemple la variable "sexe" valant 1 si l'assuré en arrêt de travail est un homme, et 0, s'il s'agit d'une femme

$$X_{i_1} = \begin{cases} 1 \text{ si l'assuré i est un homme} \\ 0 \text{ si l'assuré i est une femme} \end{cases}$$

Le rapport de risque entre un homme et une femme est :

$$\frac{h_0(t)\exp\left(\beta_1\times 1 + \cdots + \beta_p X_{i_p}\right)}{h_0(t)\exp\left(\beta_1\times 0 + \cdots + \beta_p X_{i_p}\right)} = \exp(\beta_1)$$

Toutes choses étant égales par ailleurs :

- Si $\beta_1 < 0\ alors\ \exp(\beta_1) < 1$ et le risque pour l'assuré de mettre fin à son état d'incapacité est plus élevé chez les femmes que chez les hommes.
- Si $\beta_1 > 0\ alors\ \exp(\beta_1) > 1$ et le risque pour l'assuré de mettre fin à son état d'incapacité est plus élevé chez les hommes que chez les femmes.
- Si $\beta_1 = 0\ alors\ \exp(\beta_1) = 1$ et le risque pour l'assuré de mettre fin à son état d'incapacité est le même chez les hommes et chez les femmes.

Si, en revanche, notre variable d'intérêt est continue :

Prenons par exemple la variable 'âge' correspondant à l'âge de l'assuré au moment de son incapacité.

On interprète le rapport de risque lorsque la variable augmente d'une unité (ici d'une année). Le rapport de risque est :

$$\frac{h_0(t)\exp\left(\beta_1\times(X_{i_1}+1)+\cdots+\beta_pX_{i_p}\right)}{h_0(t)\exp\left(\beta_1\times X_{i_1}+\cdots+\beta_pX_{i_p}\right)} = \exp(\beta_1)$$

Toutes choses étant égales par ailleurs :

- Si $\beta_1 < 0\ alors\ \exp(\beta_1) < 1$ et le risque de mettre fin à son état d'incapacité augmente quand l'âge de l'assuré diminue.
- Si $\beta_1 > 0\ alors\ \exp(\beta_1) > 1$ et le risque de mettre fin à son état d'incapacité augmente quand l'âge de l'assuré augmente.
- Si $\beta_1 = 0\ alors\ \exp(\beta_1) = 1$ et la variable âge n'a pas d'impact sur le risque instantané.

Le modèle de Cox permet également d'obtenir les différents coefficients β_i. Ces coefficients peuvent être obtenus par maximum de vraisemblance.

Fonction de vraisemblance

Soit X une variable aléatoire continue de densité f dépendant d'un paramètre θ. La fonction de vraisemblance notée L s'écrit :

$$L(\beta|x) = f_\theta(x).$$

Si l'on considère non plus une observation mais un ensemble d'observations indépendantes d'une même variable aléatoire X. Par exemple, on peut prendre la variable "âge" pour l'ensemble des assurés de notre portefeuille. Lorsque l'on a plusieurs observations indépendantes, la vraisemblance est le produit des vraisemblances individuelles, c'est-à-dire le produit des vraisemblances de chaque observation. On peut également définir la fonction de log-vraisemblance

$$L\left(\beta\middle|x_{1,}x_{1,}\dots,x_n\right) = \prod_{i=1}^{n} L(\beta\,|\,x_i)$$

On peut également définir la fonction de log-vraisemblance :

$$l(\theta) = \ln\left(L\left(\beta\middle|x_{1,}x_{1,}\dots,x_n\right)\right) = \sum_{i=1}^{n} \ln\, f_\beta(x_i)$$

Notre fonction de vraisemblance ne prend pas en compte les informations liées à la censure qui sont fondamentales dans le cas d'une analyse de survie.

C'est pourquoi, nous allons déterminer la forme générale de la vraisemblance dans le cadre d'un modèle présentant des données censurées (à droite). On écrit la contribution de l'individu i à la vraisemblance L_i .

On rappelle que D est notre variable de censure :

$$D_i = \begin{cases} 1 \text{ si } \mathrm{X}_i \leq C_i \\ 0 \text{ si } \mathrm{X}_i > C_i \end{cases}$$

et

$$T_i = \min(\mathrm{X}_i, C_i)$$

$$L_i(\beta) = P(T_i \in [t_i, t_i + dt_i], D_i = d_i)$$

d_i ne pouvant prendre que 2 valeurs : 0 ou 1, on a

$$\begin{aligned} L_i(\beta) &= P(T_i \in [t_i, t_i + dt_i], D_i = 1) \\ &= P(\ min(X_i, C_i) \in [t_i, t_i + dt_i], X_i \leq C_i) \\ &= P(X_i \in [t_i, t_i + dt_i], t_i < C_i) = f_X(\beta, t_i)\, S_C(\beta, t_i)\, dt_i \end{aligned}$$

De manière analogue :

$$\begin{aligned} P(T_i \in [t_i, t_i + dt_i], D_i = 0) &= P(\min(X_i, C_i) \in [t_i, t_i + dt_i], X_i \geq C_i) \\ &= P(C_i \in [t_i, t_i + dt_i], X_i \geq t_i) = S_X(t_i, \beta)\, f_c(t_i, \beta)\, dt_i \end{aligned}$$

On peut résumer les 2 cas precedents à travers l'expression suivante :

$$P(T_i \in [t_i, t_i + dt_i], D_i = d_i) = [f(\beta, t_i)\, S_C(\beta, t_i)]^{d_i}\, S(\beta, t_i)\, f_c(\beta, t_i)]^{1-d_i}$$

Par hypothèse, la censure est non informative, c'est-à-dire que la loi de censure est indépendante du paramètre β. On peut réécrire la vraisemblance sous la forme suivante.

$$L(\beta) = \prod_{i=1}^{n} [f(\beta, t_i)]^{d_i}\, S(\beta, t_i)]^{1-d_i}$$

En remarquant que $f(t) = S(t) \times h(t)$, on a :

$$L_i(\beta) = [S(\beta, t_i)\, h(\beta, t_i)]^{d_i}\, S(\beta, t_i)]^{1-d_i} = h(\beta, t_i)^{di}\, S(\beta, t_i)$$

La vraisemblance totale du modèle peut ainsi s'écrire :

$$L(\beta) = \prod_{i=1}^{n} S(\beta, t_i)[h(\beta, t_i)]^{d_i}$$

L'estimation des coefficients passe par la détermination d'une vraisemblance partielle. La vraisemblance partielle du modèle est :

$$L_{COX}(\beta) = \prod_{i=1}^{n} [\frac{h(t_i)}{\sum_{j \epsilon R_i} h(t_j)}]^{d_i}$$

Etant dans un modèle à risques proportionnels, on peut remplacer la fonction de risque par son expression pour obtenir :

$$L_{COX}(\beta) = \prod_{i=1}^{n} [\frac{\exp\,(\,X_i\,.\beta)}{\sum_{j\epsilon R_i} \exp\,(\,X_j\,.\beta)}]^{d_i}$$

La vraisemblance partielle est indépendante du risque de base $h_0(t)$ et par conséquent, ses paramètres ne peuvent alors pas être estimés par cette méthode.

Le passage au log donne :

$$\log(L_{COX}) = \sum_{i=1}^{n} [\beta X_i - \log\,(\sum_{j\epsilon R_i} \exp(\,\beta X_j)]$$

L'estimateur du maximum de vraisemblance est donc :

$$\hat{\beta} = argmax_B \log\,(L_{COX})$$

A la différence de la régression linéaire, nous ne pouvons pas trouver de solution analytique. On utilise donc un algorithme d'optimisation (l'algorithme de Newson-Raphson) à partir des équations de score et de la matrice d'information de Fisher.

On rappelle que la fonction Score est définie par $S(\beta) = \frac{\partial \log L(\beta)}{\partial \beta}$

$\forall\, i, j$, on appelle matrice d'information de Fisher, la matrice d'information constituée des éléments suivants :

$$I(\beta) = \left[\frac{-\delta^2 \log\,(L_{cox}(\beta)}{\delta\beta_i\beta_j}\right]_{i,j}$$

Cette méthode de résolution nécessite la validité d'une hypothèse forte, il ne doit pas y avoir d'événements simultanés ce qui en pratique n'est pas toujours le cas. Ainsi pour une date donnée, un seul assuré doit sortir de son état d'incapacité.

Si cette hypothèse n'est pas valide, l'estimation de la vraisemblance doit être corrigée en utilisant la méthode dite de Breslow (méthode de référence utilisée sur le logiciel SAS). Par ailleurs d'autres méthodes ont également été développées par la suite (par exemple la méthode d'Efron en 1977)

Afin de traiter le cas des ex- æquo, Breslow a mis en place en 1974 une approximation de la fonction de log vraisemblance :

$$\log\big(L(\beta)\big) = \sum_{i=1}^{T} \beta \sum_{j=1}^{d_i} X_j - d_i \log[\sum_{j\epsilon R_i} \exp\big(\beta X_j\big)]$$

Avec T le nombre de dates différentes
Et d_i le nombre d'observations à l'instant i

4.1.2 Tests sur les paramètres

Une des principales caractéristiques de la méthode du maximum de vraisemblance est qu'il existe des statistiques de tests faciles à mettre en place. Les valeurs des paramètres β précédemment estimés peuvent être testées.
Un test statistique est très utile lorsqu'on veut prendre une décision entre 2 hypothèses statistiques.
L'hypothèse énoncée est appelée hypothèse nulle H_0.
H_1 est appelé l'hypothèse alternative.

$$H_0: \beta = \beta_0$$
$$H_1: \beta \neq \beta_0$$

3 tests (équivalents asymptotiquement) sont généralement utilisés : le test du maximum de vraisemblance, le test de Wald et le test du Score.

- **Test du maximum de vraisemblance**

Le test du rapport de vraisemblance se base sur le rapport de vraisemblance entre les 2 modèles. La statistique de ce test est : $\chi_R{}^2 = 2 \times [\, log(\, L(\beta_0)) - log(L(\beta))]$

La statistique du rapport de vraisemblance entre les 2 modèles suit approximativement une distribution du Khi-2 de degré p égal à la différence du nombre de paramètres estimés entre les 2 modèles.

$$\chi_R{}^2 \sim \chi_p{}^2$$

- **Test du Score**

Le test du score prend en considération la pente du log vraisemblance en β. La pente correspond à la fonction score $S(X) = \frac{\partial \log L(\beta)}{\partial \beta}$.

La statistique du test est : $\chi_S{}^2 = \big(S(\beta_0)\big)' I(\beta_0)^{-1} \big(S(\beta_0)\big)$

Comme la statistique du rapport de vraisemblance, la statistique du test du score suit également une distribution du Khi-2 de degré p égal à la différence du nombre de paramètres estimés entre les 2 modèles.

$$\chi_S{}^2 \sim \chi_p{}^2$$

- **Test de Wald**

Le test de Wald se concentre sur la différence $(\hat{\beta} - \beta_0)$ entre l'estimateur du maximum de vraisemblance, et la valeur que l'on cherche à tester.

La statistique du test de Wald est :

$$\chi_w{}^2 = \big(\hat{\beta} - \beta_0\big)' I\big(\hat{\beta}\big)\big(\hat{\beta} - \beta_0\big)$$

La statistique du test du score suit également une distribution du Khi-2 de degré p égal à la différence du nombre de paramètres estimés entre les 2 modèles.

$$\chi_W{}^2 \sim \chi_p{}^2$$

Sur notre base de données, il sera nécessaire de tester la non-nullité ($H_0 : \beta = 0$) de nos coefficients, estimés par le maximum de vraisemblances partielles, via ces 3 tests statistiques.

4.1.3 Validation des hypothèses du modèle de Cox

L'Hypothèse des risques proportionnels est l'hypothèse fondamentale du modèle de Cox. Cette hypothèse stipule que le rapport des fonctions de hasard entre 2 individus est constant dans le temps.

$$\frac{h(t|X_i)}{h\big(t\big|X_j\big)} = \frac{h_0(t)\exp(\, X_i \,.\, \beta)}{h_0(t)\exp(\, X_j \,.\, \beta)} = \exp(\, [X_i - X_j] \,.\, \beta)$$

En pratique, plusieurs méthodes nous permettent de valider cette hypothèse.

- Graphiquement, on peut vérifier cette hypothèse.
  En se basant sur la fonction de hasard, pour différentes valeurs des variables, on trace les nuages de points :
  $$\left(t, \left[\log(h(t|X_i)) - \log(h(t|X_j)\right]\right)$$
  L'hypothèse est vérifiée si les points sont alignés sur une droite parallèle à l'axe des abscisses.

- On peut également utiliser un test basé sur les résidus de Schœnfeld
  Pour calculer le résidu de Schœnfeld que l'on notera r_i d'une covariable X, on effectue la différence entre sa valeur à la date i pour l'assuré i et la valeur moyenne de l'ensemble des individus encore à risques à cet instant. A noter que l'on ne calcule le résidu de Schœnfeld que sur des individus non censurés.
  Soit r_{ik} le k^{eme} résidu de Schœnfeld pour la i^{eme} covariable.
  $$r_{jk} = \delta_k \, (x_{ik} - a_{ik})$$

  Avec :
  x_{ik} représente la valeur de la i^{eme} covariable à la date k
  a_{ik}représente la valeur moyenne des individus encore à risques à l'instant k

  $$a_{ik} = \frac{\sum_{m \in R_k} x_{im} \exp\,(x_{im}\widehat{\beta})}{\sum_{m \in R_k} \exp\,(x_{im}\widehat{\beta})}$$

  δ_k est l'indicateur de censure

L'hypothèse des risques proportionnels est validée si le graphe des résidus de Schœnfeld forme une marche aléatoire de moyenne nulle. En effet les résidus ne feront ainsi pas apparaitre de dépendance par rapport à la variable temporelle.

L'hypothèse de log linéarité se vérifie à l'aide des résidus de martingale.
Les résidus de martingale correspondent, pour un individu i, à la différence entre l'observation de l'évènement δ_i 0 si l'observation est censurée, 1 dans le cas contraire et le nombre de sortie théoriques évaluées par le modèle.

$$r_m = \delta_i - \varphi(t_i|X_i)$$

Les résidus de martingale sont d'espérances nulles. De plus, si les observations sont indépendantes et identiquement distribuées (i.i.d), les résidus de martingale sont entre 2 sujets non corrélés.

4.2 Application à nos données et estimation des paramètres

4.2.1 Estimation des paramètres

Après avoir étudié la théorie du modèle de Cox, nous allons l'implémenter à l'aide du logiciel R sur nos données afin de déterminer si le modèle est adapté.
En prenant en compte les variables explicatives utilisables à notre disposition notre 1er modèle s'écrit :

$$h(t|x) = h_0(t)\exp(\ \beta_1 \times\ age + \ \beta_2\ \times SexM + \ \beta_3\ \times Motif\_SinVie\_pro + \ \beta_4\ \times CSPNCA\ )$$

La variable "âge" est étudiée sous forme continue. Nous n'avons pas créé de tranches d'âge afin d'une part de ne garder qu'un seul degré de liberté et d'autre part, ne pas perdre d'information à la suite de la discréditation de la variable.
Les variables associées au sexe (homme/femme), à la catégorie socioprofessionnelle (cadre/non-Cadre) et au motif du sinistre (Vie professionnelle/ Accident de la vie privée) étant binaires, on a :

$$SexM = \begin{cases} 1 \text{ si l'assuré i est un homme} \\ 0 \text{ si l'assuré i est une femme} \end{cases}$$

$$Motif\_SinVie\_pro = \begin{cases} 1 \text{ si le motif de l'incapacité est professionnel} \\ 0 \text{ sinon} \end{cases}$$

$$CSPNCA = \begin{cases} 1 \text{ si l'assuré i n'est pas un cadre} \\ 0 \text{ si l'assuré i est un cadre} \end{cases}$$

La population de référence correspond donc aux femmes cadres en incapacité pour cause personnelle. En implémentant ce modèle de Cox avec le logiciel R à l'aide de la fonction *coxph* du package *survival*, on obtient l'estimation des différents paramètres du modèle.

```
Call:
coxph(formula = Surv(duree, Censure) ~ Age + SEXE + MOTIF_SIN +
    CSP, data = data4)

  n= 56489, number of events= 54074

                     coef exp(coef)  se(coef)       z Pr(>|z|)
Age              -0.021448  0.978780  0.000415 -51.686  < 2e-16 ***
SEXEM            -0.057381  0.944234  0.008920  -6.433 1.25e-10 ***
MOTIF_SINVIE PRO -0.263315  0.768500  0.018459 -14.265  < 2e-16 ***
CSPNCA           -0.005237  0.994776  0.011016  -0.475    0.634
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Et comme toutes choses étant égales par ailleurs :

- La catégorie socioprofessionnelle n'influe pas sur le risque instantané.
- Le risque instantané diminue de 6% si l'assuré est un homme.

4.2.2 Sélection des variables pertinentes

Nous sommes maintenant en mesure de déterminer les coefficients β de notre modèle de Cox. Les coefficients ont été estimés sur l'ensemble des covariables dont nous disposions. Cependant, certaines variables ne sont pas forcément pertinentes.
De plus la variable relative à la catégorie socioprofessionnelle n'est pas significative. Regardons si notre modèle peut être amélioré par minimisation de l'AIC.
Le critère d'information d'Akaike (ou critère AIC) est un moyen permettant de mesurer la qualité d'un modèle statistique. C'est un moyen efficace de sélection de modèles.
Le Critère d'Akaike est la vraisemblance du modèle pénalisé par le nombre de paramètres à estimer.

$$AIC = 2k - 2\ln(L)$$

3 stratégies de sélection sont possibles :

- Méthode AIC forward

La méthode de sélection ascendante (ou AIC forward) consiste à démarrer avec notre modèle nul (un modèle avec une seule variable et le Critère AIC le plus faible) puis d'y ajouter les variables une par une. La variable ajoutée au fur et à mesure et celle qui diminue le plus le critère d'AIC. La principale limite de cette méthode est que lorsqu'on introduit une variable dans le modèle, on ne peut plus la supprimer.

- Méthode AIC backward

La méthode descendante (AIC backward) consiste à partir du modèle complet et d'y éliminer pas à pas les variables de manière à avoir le critère d'AIC le plus faible. Le principal problème de cette méthode est qu'il n'y est plus possible de réintroduire une variable dans le modèle une fois qu'elle a été supprimée.

- La méthode AIC stepwise

Cette méthode est une amélioration de la méthode ascendante. En effet, dorénavant à chaque étape, il y a possibilité de supprimer une variable précédemment ajoutée.

Ces stratégies basées sur les critères d'Akaike nous permettent de conclure le modèle optimal composé des variables "â*ge" "sexe"* et "*Motif_SIN".*

$$h(t|x) = h_0(t)\exp(\ \beta_1 \times\ age + \ \beta_2 \times SexM + \ \beta_3 \times Motif\_SinVie\_pro)$$

Les coefficients associés au modèle sont :

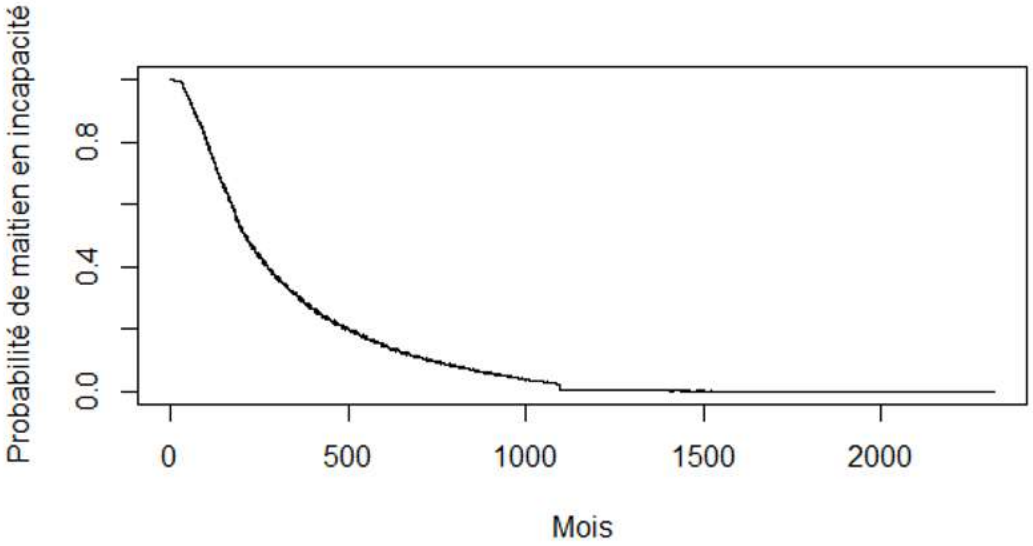
```
  n= 56489, number of events= 54074

                      coef  exp(coef)   se(coef)       z Pr(>|z|)
Age              -0.0214262  0.9788017  0.0004124 -51.954  < 2e-16 ***
SEXEM            -0.0570440  0.9445526  0.0088915  -6.416  1.4e-10 ***
MOTIF_SINVIE PRO -0.2637503  0.7681653  0.0184359 -14.306  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Les coefficients de ce modèle sont tous significatifs. De plus, on a également la valeur des différents tests.

```
Concordance= 0.578  (se = 0.001 )
Likelihood ratio test= 3015  on 3 df,   p=<2e-16
Wald test            = 3055  on 3 df,   p=<2e-16
Score (logrank) test = 3079  on 3 df,   p=<2e-16
```

On remarque que pour l'ensemble des 3 tests la valeur de la probabilité (p-value) est inférieure au seuil des 5%.
L'hypothèse de non-nullité des coefficients est donc vérifiée pour l'ensemble des coefficients β .
La forme de la fonction de survie pour ce modèle de Cox est :

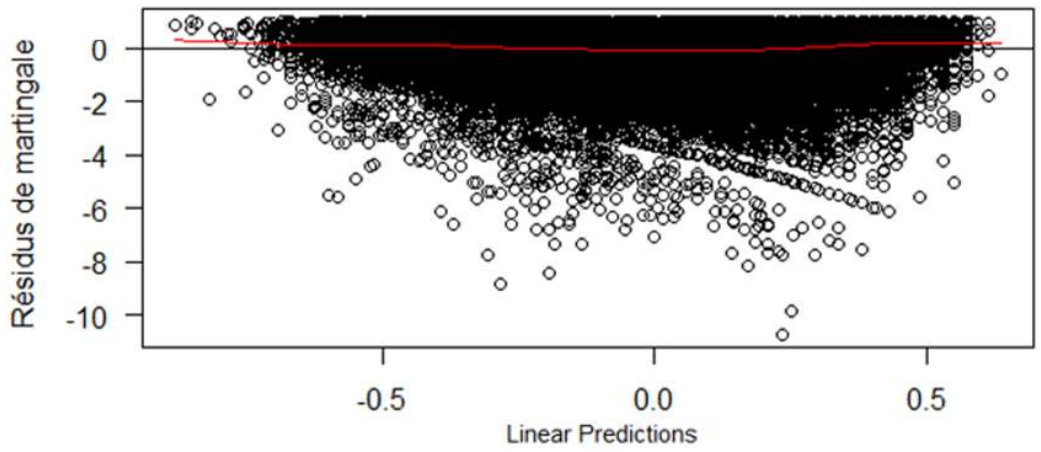
Loi de maintien pour le modèle de Cox



4.3 Validation du modèle

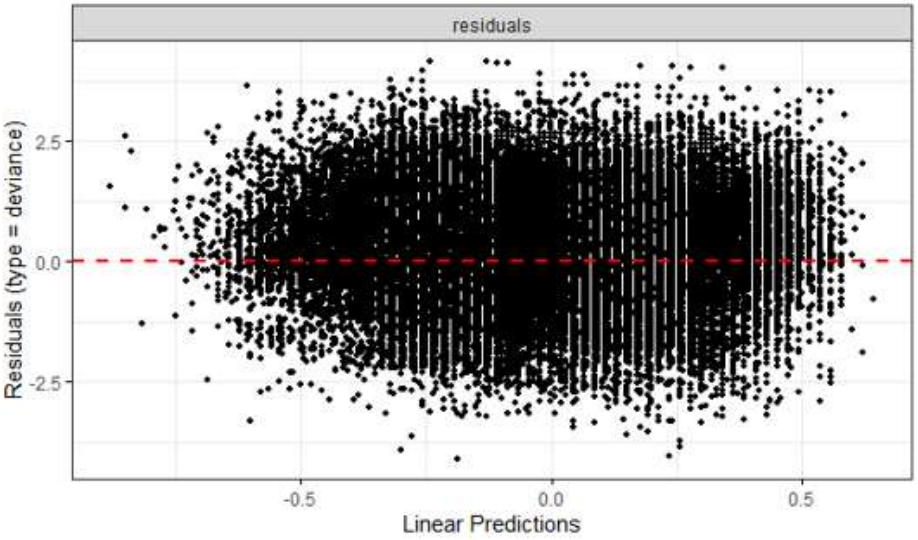
Pour que le modèle soit utilisable, il faut valider l'hypothèse de log-linéarité ainsi que l'hypothèse de proportionnalité.
Représentons, les résidus de martingale

Résidus de martingale



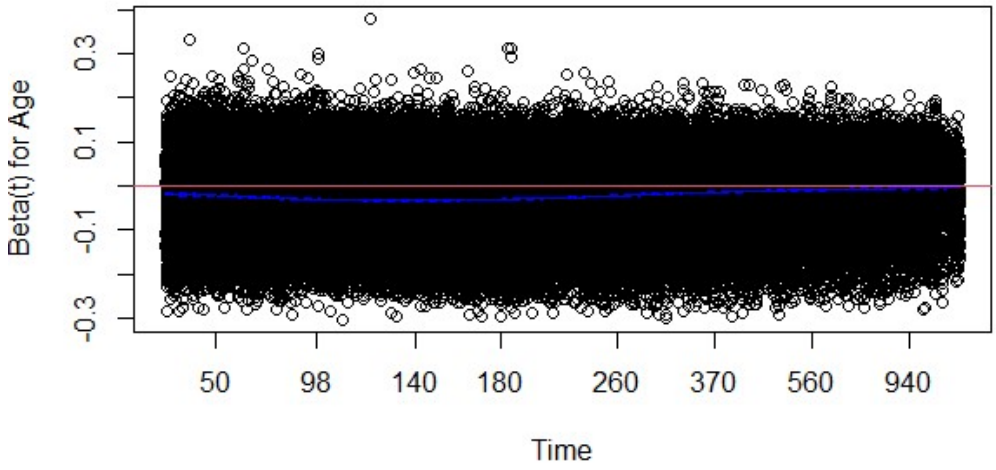
L'espérance des résidus de martingale semble être nulle. L'hypothèse de log linéarité est validée.
A partir des résidus de martingale, nous pouvons détecter certaines valeurs aberrantes en utilisant les résidus de déviance. Le résidu de déviance est une transformation normalisée du résidu de martingale.
Un graphique montrant trop de valeur aberrante peut traduire un mauvais modèle. Les résidus de déviance doivent être distribués à peu près symétriquement autour de l'axe des abscisses. Les valeurs positives correspondent aux assurés étant sortis de leur état d'incapacité trop tôt par rapport aux durées modélisées.
Les valeurs négatives correspondent aux assurés étant en incapacité trop longtemps.
Les points avec un résidu de déviance très éloigné des autres résidus sur l'axe des ordonnées correspondent à des valeurs aberrantes. Représentons les résidus de dévianc

Résidus de déviance

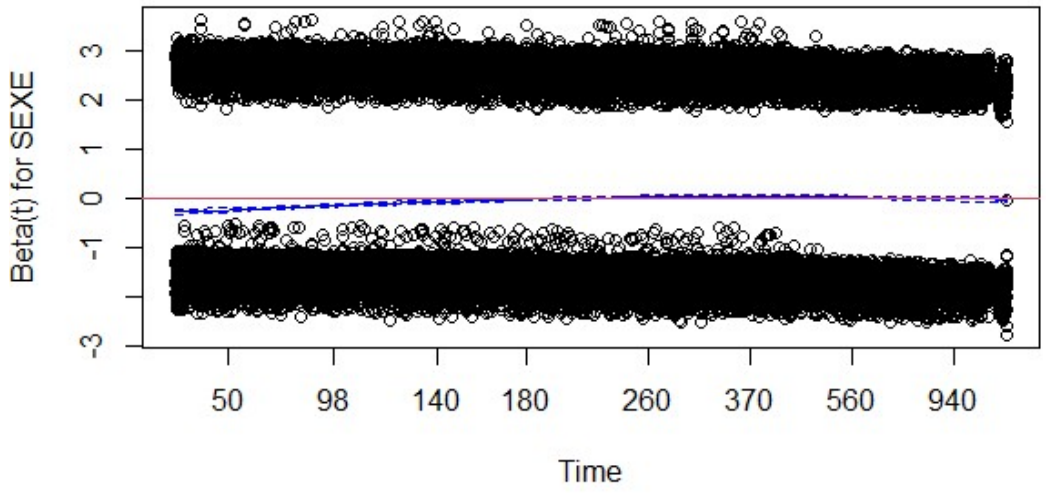


Le modèle semble être adapté.
L'hypothèse la plus importante dans le modèle de Cox est l'hypothèse des risques proportionnels que nous allons évidemment tester.
La validation de cette hypothèse peut s'effectuer par l'analyse des résidus de Schœnfeld Ces résidus doivent être indépendants de la durée de l'arrêt de travail.
La fonction *cox.zph* sur R vérifie que les variables explicatives sont effectivement bien indépendantes du temps en se basant sur l'étude des résidus de Schœnfeld.

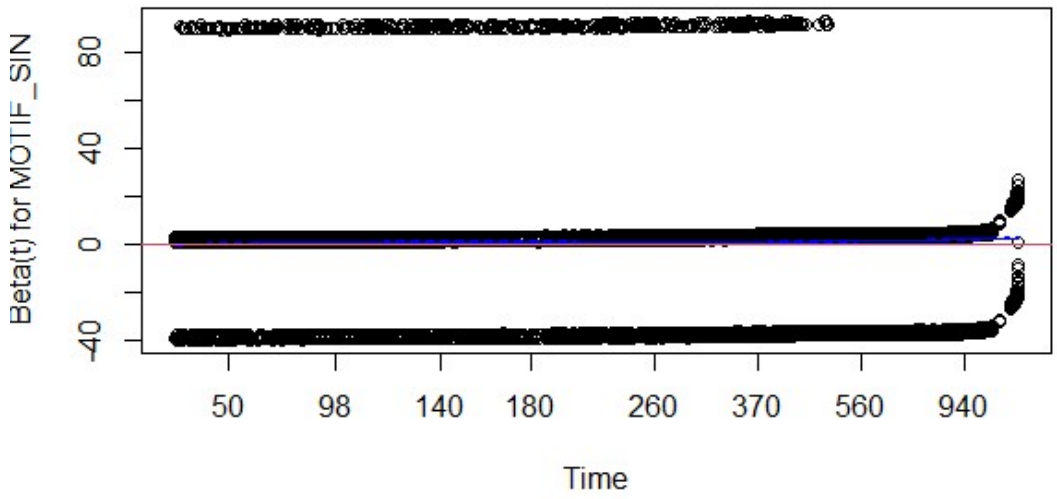
Résidus de Schœnfeld pour la variable âge



Résidus de Schœnfeld pour la variable sexe



Résidus de Schœnfeld pour la variable motif_sin



Le logiciel d'étude R trace implicitement un lissage ajusté au tracé par des splines. Il est difficile d'en tirer un jugement. Les résidus de Schœnfeld semblent être des droites horizontales mais on remarque quelques petits sauts particulièrement pour les variables Age et Sexe. Le test graphique n'est pas suffisant pour conclure de la validité de l'hypothèse. Pour vérifier si notre modèle est utilisable, un test statistique sur les résidus de Schœnfeld est mis en place.
Afin de montrer que les résidus sont stables avec le temps, nous posons l'hypothèse suivante pour chaque variable explicative :

$$H_0 : \beta_j(t) = \beta_j$$
$$H_1 : \beta_j(t) \neq \beta_j$$

La significacité du modèle global sera traduite par l'hypothèse suivante :

$$H_0 : \beta(t) = \beta$$
$$H_1 : \beta(t) \neq \beta$$

La sortie suivante est obtenue.

	chisq	df	p
Age	381.3	1	< 2e-16
SEXE	112.1	1	< 2e-16
MOTIF_SIN	15.9	1	6.6e-05
GLOBAL	478.0	3	< 2e-16

Les variables relatives à l'âge au sexe et au motif du sinistre ont une p-value inférieure à 5% ce qui implique que l'hypothèse nulle est rejetée en faveur de l'hypothèse alternative pour chaque variable explicative ainsi que pour le modèle global. Les résidus de Schœnfeld sont dépendants du temps.

L'hypothèse fondamentale des risques proportionnels n'est pas validée pour notre modèle. Par conséquent, le modèle de Cox n'est pas adapté à notre étude.
Pour la suite de l'étude, nous nous concentrerons sur les estimateurs de Kaplan- Meier et de Nelson-Aalen.

5 Lissage des taux bruts

Pour les estimateurs de Kaplan- Meier et de Nelson- Aalen, nous avons construit les lois de survie. Cependant, cette construction s'est basée sur des taux de sortie qui peuvent s'avérer très erratiques. C'est pourquoi, il est fondamental d'implémenter des méthodes précises et fiables nous permettant de lisser les taux bruts de sortie.
2 familles de lissage peuvent être citées :

- Des méthodes paramétriques qui évaluent les différents paramètres du modèle à partir des observations. Les méthodes de Makeham ou de Weibull sont les plus utilisées généralement.
- Des méthodes non paramétriques telles que les méthodes de Whittaker-Henderson ou les méthodes de lissage par moyenne mobile.

Dans le cadre de notre étude de la modélisation des lois de maintien en incapacité, nous nous intéresserons aux méthodes de lissage non paramétriques qui vont nous permettre de lisser nos taux bruts sans hypothèse quelconque sur la forme du modèle.

5.1 Méthode par moyenne mobile

5.1.1 Présentation

Une première méthode dite non paramétrique se base sur l'utilisation d'une moyenne mobile. Cette méthode, principalement utilisée dans le cadre d'études de séries temporelles, permet de lisser les taux bruts en atténuant les fluctuations irrégulières et significatives tout en gommant les possibles saisonnalités. On préférera lisser à l'aide d'un ordre impair pour éviter des phénomènes de déphasage de la série lissée.

5.1.2 Implémentation

Pour autant que l'on se restreigne aux moyennes mobiles d'ordre p impair (p=2n+1) on a :

$$L^*_k = \frac{1}{p} \sum_{i=-n}^{n} \widehat{L_{k+i}}$$

Ainsi à chaque taux brut k, le taux lissé est égal à la moyenne arithmétique des p=2n+1 taux bruts décomposés comme :

- Les valeurs des n taux bruts suivant le point k
- Le taux brut en k
- Les valeurs des n points précédents le taux brut en k

C'est ainsi que le principal problème posé par cette méthode se trouve sur les valeurs extrêmes. La méthode des moyennes mobiles n'est pas applicable aux bords.

5.2 Méthode de Whittaker-Henderson

5.2.1 Présentation

La méthode de Whittaker-Henderson est une méthode de lissage non paramétrique utilisable aussi bien dans un cas unidimensionnel que dans un cas bidimensionnel. Cette méthode de base sur la minimisation de 2 critères : le critère de fidélité et le critère de régularité.

5.2.2 Méthode de Whittaker-Henderson dans le cas unidimensionnel

Implémentons cette méthode de lissage dans le cas unidimensionel. On pose :

$\hat{L} = \hat{L}_{1\leq i\leq n}$ le vecteur de dimension n contenant l'ensemble des taux bruts qui doivent être lissés.

$L^* = {L^*}_{1\leq i\leq n}$ le vecteur de dimension n contenant l'ensemble des taux lissés.

W = $diag(w_i)_{1\leq i\leq n}$ la matrice diagonale des poids atribués à chacun des taux bruts.

Z : un parametre de notre modele que l'on choisi et qui permet d'améliorer la régularité des taux

Le critère de fidélité F se définit de la façon suivante. :

$$F = \sum_{i=1}^{n} w_i \left({L_i}^* - \hat{L}_i\right)^2 = \left(L^* - \hat{L}\right)' W \left(L^* - \hat{L}\right)$$

Le critere de régularité se définit comme :

$$S = \sum_{i=1}^{n-z} (\Delta^z {L_i}^*)^2 = (\Delta^z L^*)'(\Delta^z L^*)$$

Avec Δ l'opérateur différence tel que $\Delta L_i = L_{i+1} - L_i$ et ${\Delta L_i}^z$ l'opérateur différence composé z fois tel que ${\Delta L_i}^z = \sum_{j=0}^{z} \binom{z}{j} (-1)^{z-j} L_{i+j}$

Nous cherchons à minimiser le critère M correspondant à une combinaison linéaire de critère de fidélité et du critère de régularité. On note également h un 2e paramètre.

$$M = F + h \times S$$

En introduisant les formules explicites de la fidélité et de la régularité, on a :

$$M = \left(L^* - \hat{L}\right)' W \left(L^* - \hat{L}\right) + h \times (\Delta^z L^*)'(\Delta^z L^*)$$

En posant K_z la matrice $(n - z) \times n$ telle que $K_z L = \Delta^z L$, on obtient :

$$M = \left(L^* - \hat{L}\right)' W \left(L^* - \hat{L}\right) + h\, L^{*\prime} K_z{}' K_z L^*$$

$$M = L^{*\prime} W L^* - 2 L^{*\prime} W \hat{L} + \hat{L}' W \hat{L} + h L^{*\prime} K_z{}' K_z L^*$$

Minimiser M revient à résoudre le probleme d'optimisation suivant.

$$\frac{\partial M}{\partial L^*} = 0$$

En application des règles de dérivation usuelles :

$$\frac{\partial M}{\partial L^*} = 2WL^* - 2W\hat{L} + 2hK_z{}'K_zL^*=0$$

$W + hK_z{}'K_z$ est inversible car c'est une matrice définie positive.

Ainsi :

$$L^* = (W + hK_z{}'K_z)^{-1}W\hat{L}$$

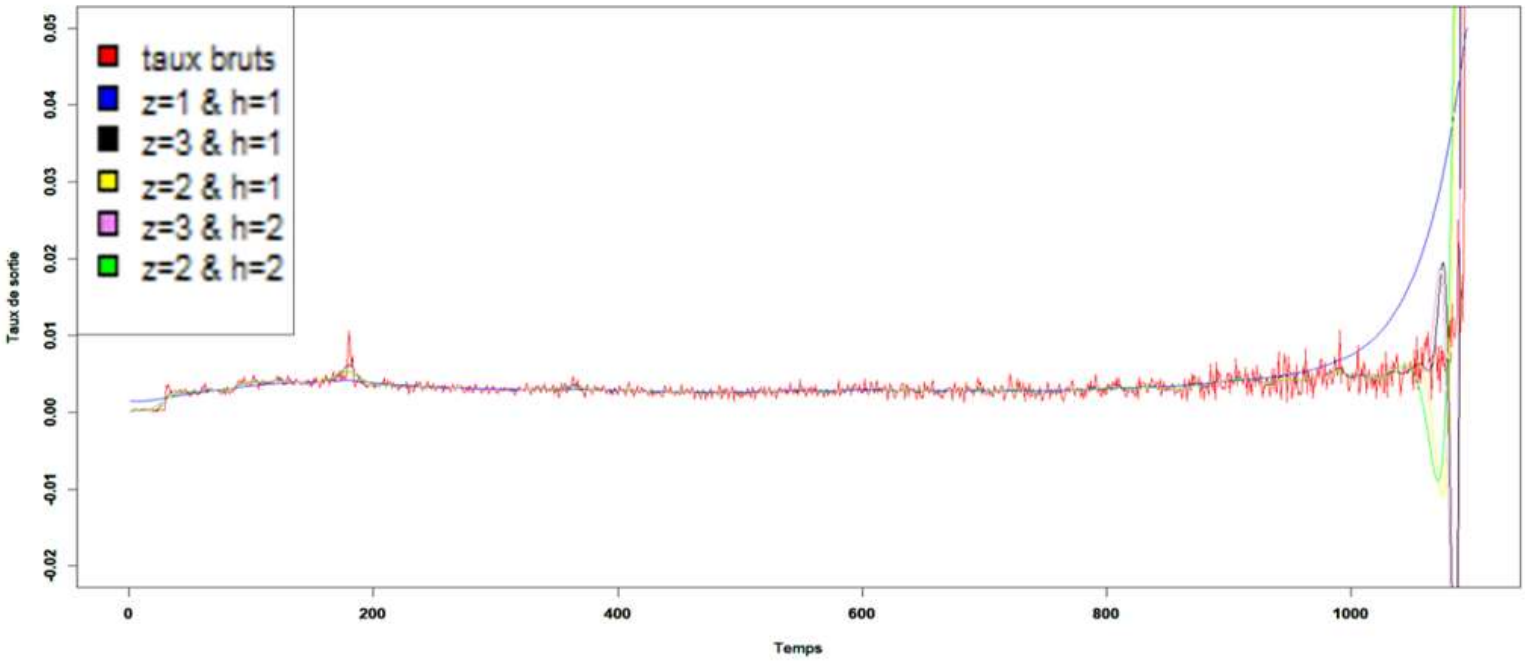
5.3 Application sur nos données

L'étape de lissage a été réalisée sur le logiciel R pour l'ensemble des méthodes.

Méthode de Whittaker-Henderson

La méthode de Whittaker-Henderson nécessite de choisir un vecteur poids. 2 principaux choix peuvent être effectués à ce stade : nous pouvons opter pour des poids identiques pour l'ensemble de nos données ou nous pouvons opter pour une répartition des poids prenant en compte les effectifs de maintien en incapacité. Ainsi, les anciennetés très élevées dont les taux peuvent être moins fiables auront un poids et donc un impact très faible dans le lissage.
Les valeurs des 2 paramètres z et h sont déterminées de manière visuelle. Le paramètre z traduit la précision du lissage tandis qu'un paramètre h élevé implique une courbe plus lisse. Nous testerons plusieurs valeurs de z et de h afin de choisir les paramètres les plus satisfaisants.

Lissage des taux bruts par la méthode de Whittaker-Henderson

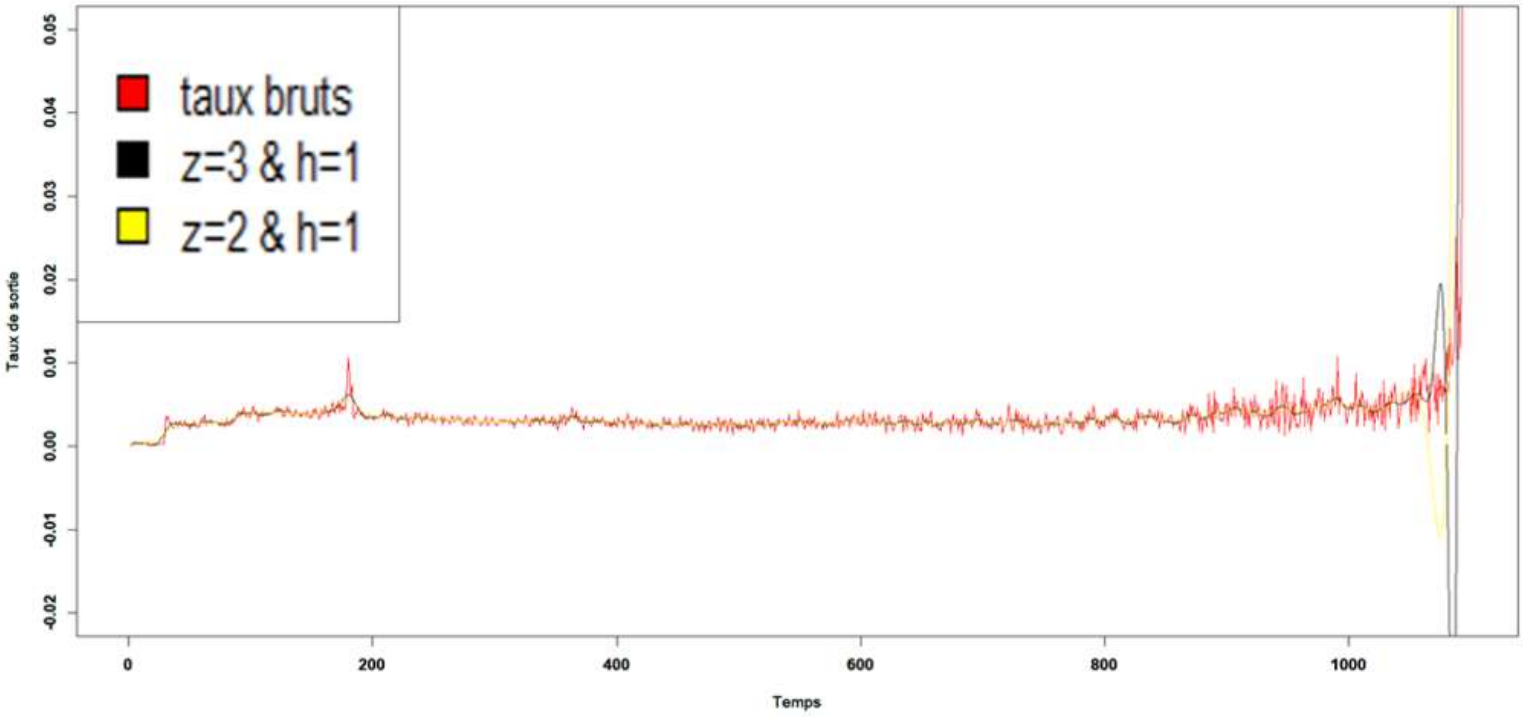


Après avoir réalisé plusieurs essais avec la méthode de lissage de Whittaker-Henderson, nous retiendrons les 2 combinaisons de valeurs z=3, h =1 et z=2, h =1 qui sont les paramètres qui allient le mieux, précision (surtout pour les durées supérieures à 1000 jours) et lissage.

- z=3 et h=1
- z=2 et h=1

Nous représentons ces tracés sur un nouveau graphique.

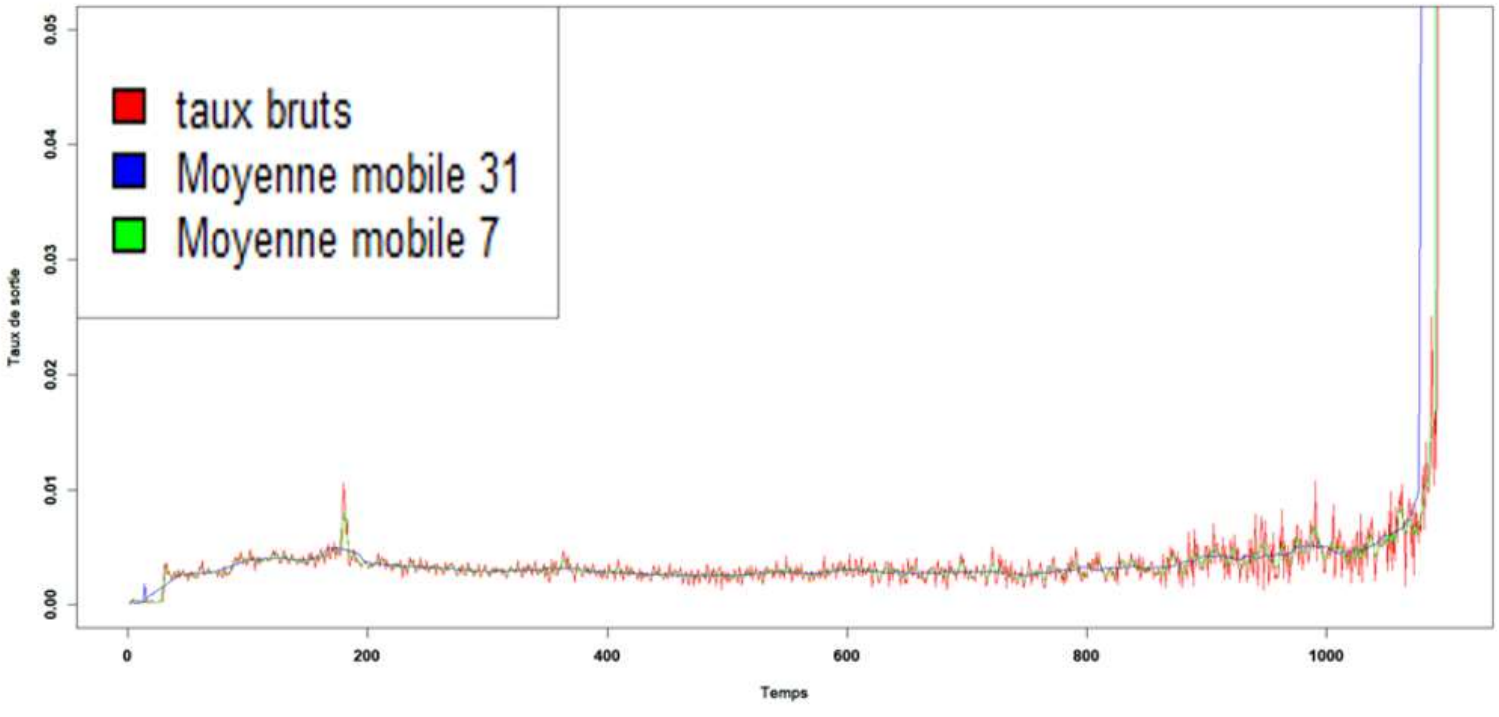
Lissage des taux bruts par la méthode de Whittaker-Henderson



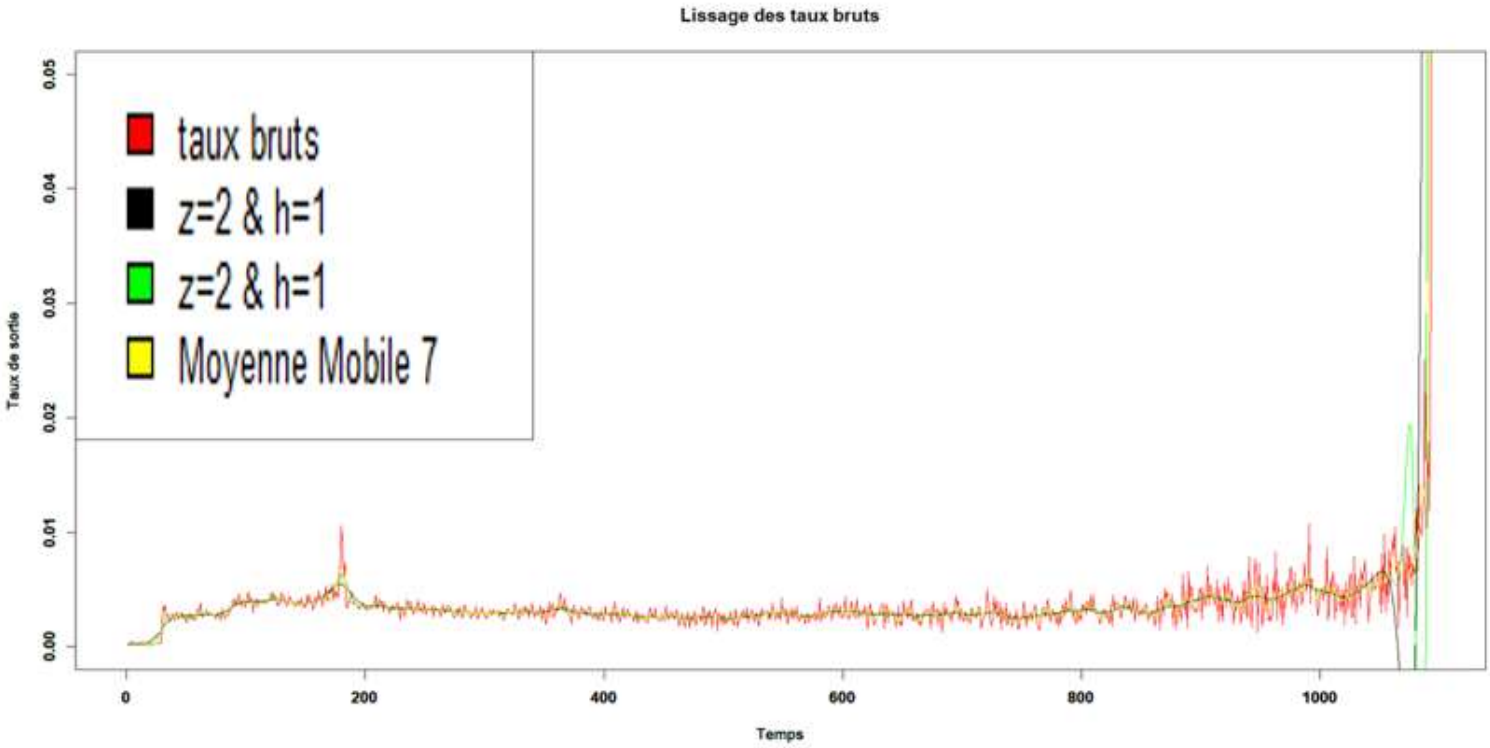
Méthode des moyennes mobiles :

La méthode des moyennes mobiles a été implémentée directement depuis EXCEL. Plusieurs valeurs d'ordres ont été testées dans le paramétrage de notre modèle de lissage. Arbitrairement, nous avons choisi de représenter 2 lissages d'ordres respectifs 7 et 31 permettant d'observer l'influence de l'ordre sur le lissage. On obtient :

Lissage des taux bruts par la méthode des moyennes mobiles



On vérifie bien que plus la moyenne mobile est de période élevée et plus sa courbe sera lissée. En comparant la méthode des moyennes mobiles avec un ordre 7 et la méthode de Whittaker-Henderson de paramètres z= 3 et h=1 :



La moyenne mobile d'ordre 7 semble la plus adaptée à nos taux bruts
Les taux révisés ont l'air cohérents et satisfaisants. Dans l'ensemble, les résultats sont très proches pour la majorité des points. Cependant sur les durées très élevées et proche de 1095, la méthode des moyennes mobiles parait la plus satisfaisante.

La méthode des moyennes mobiles d'ordre 7 lisse particulièrement bien les données tout en restant proche des estimations.

6 Provisionnement

D'après le code des assurances et l'article A331-22, les provisions techniques des prestations d'incapacité et d'invalidité sont la somme :

- Des provisions correspondant aux prestations d'incapacité de travail à verser après le 31 décembre de l'exercice, au titre des sinistres en cours à cette date majorées des provisions dites pour rentes en attente relatives aux rentes d'invalidité susceptibles d'intervenir ultérieurement au titre des sinistres d'incapacité en cours au 31 décembre de l'exercice ;

- Des provisions correspondant aux prestations d'invalidité à verser après le 31 décembre de l'exercice au titre des sinistres d'invalidité en cours à cette date

Le calcul des provisions techniques de prestations d'incapacité de travail et d'invalidité est effectué à partir des éléments suivants :

- Des lois de maintien en incapacité de travail et invalidité développées par le BCAC.
  Toutefois, il est possible pour une institution d'utiliser une loi de maintien établie par ses soins et certifiée par un actuaire agréé à cet effet par l'une des associations d'actuaires reconnues par la commission de contrôle mentionnée à l'article L. 510-1 du code de la mutualité ;

- Un taux d'actualisation qui ne peut excéder 75 % du taux moyen des emprunts de l'Etat français calculé sur base semestrielle, sans pouvoir dépasser 4,5 %

Rappelons que les provisions pour sinistres en cours correspondent à la valeur actuelle des engagements probables futurs.

6.1 Coefficients de provisionnement

6.1.1 Calculs des coefficients de provisionnements

Les tables de maintien en incapacité précédemment construites peuvent être utilisées afin de provisionner les prestations servies en cas d'incapacité.
Dans un premier temps, nous nous intéresserons à la provision mathématique unitaire c'est-à-dire le montant de la provision mathématique pour un euro de rente mensuelle versée en début de période.
Notons :

- i correspondant au taux d'actualisation annuel
- a correspondant à l'ancienneté de l'assuré en incapacité

La provision mathématique PM unitaire pour un assuré ayant une ancienneté a s'écrit :

$$PM_{a,deb} = \sum_{k=a+1}^{36} \frac{l_k}{l_a} \times \left(\frac{1}{(1+i)^{\frac{k-a}{12}}} \right)$$

Pour le versement d'un euro de rente mensuelle en fin de période, la formule devient :

$$PM_{a,fin} = \sum_{k=a}^{35} \frac{l_k}{l_a} \times \left(\frac{1}{(1+i)^{\frac{k-a}{12}}} \right)$$

Le coefficient de provisionnement préconisé par le BCAC considère la moyenne des flux versés en début de période et de ceux versés en fin de période. Nous obtenons l'égalité ci-dessous. Sous l'hypothèse que les sorties ont lieu en milieu de période.

$$PM_{a,BCAC} = \sum_{k=a}^{35} \frac{1}{2 \times l_a} \times \left[\frac{l_k}{(1+i)^{\frac{k-a}{12}}} + \frac{l_{k+1}}{(1+i)^{\frac{k+1-a}{12}}} \right]$$

Les différents coefficients de provisionnement prennent en considération les probabilités de maintien en incapacité à travers le rapport $\frac{l_k}{l_a}$.

Nous utiliserons pour la suite de ce mémoire la troisième formule.

On note également que coefficient de provisionnement est décroissant par rapport au taux d'actualisation i. En effet, $\forall\, i_2 > i_1 > 0$, on a $PM_a(i_2) < PM_a(i_1) < PM_a(0)$.

En multipliant le coefficient de provisionnement par l'indemnité mensuelle IT, on obtient la formule de la provision pour rente d'incapacité en cours.

6.1.2 Esperance résiduelle

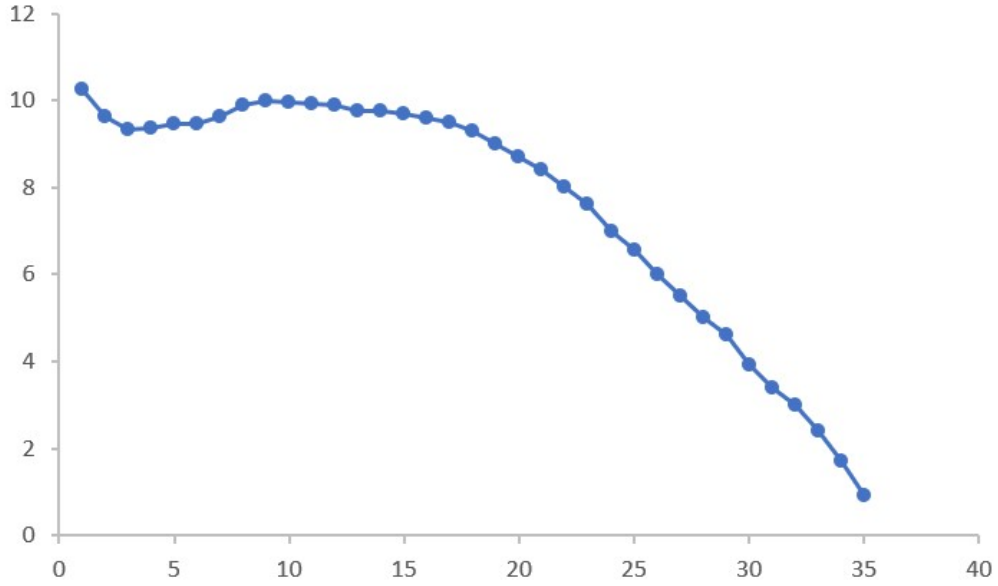
Afin de neutraliser l'effet de taux technique i, on considère ce dernier nul. On obtient alors l'espérance résiduelle notée arbitrairement E.

$$E_a = \frac{1}{2} \sum_{k=a}^{35} \frac{l_k + l_{k+1}}{l_a}$$

α represente l'ancienneté de l'incapcité

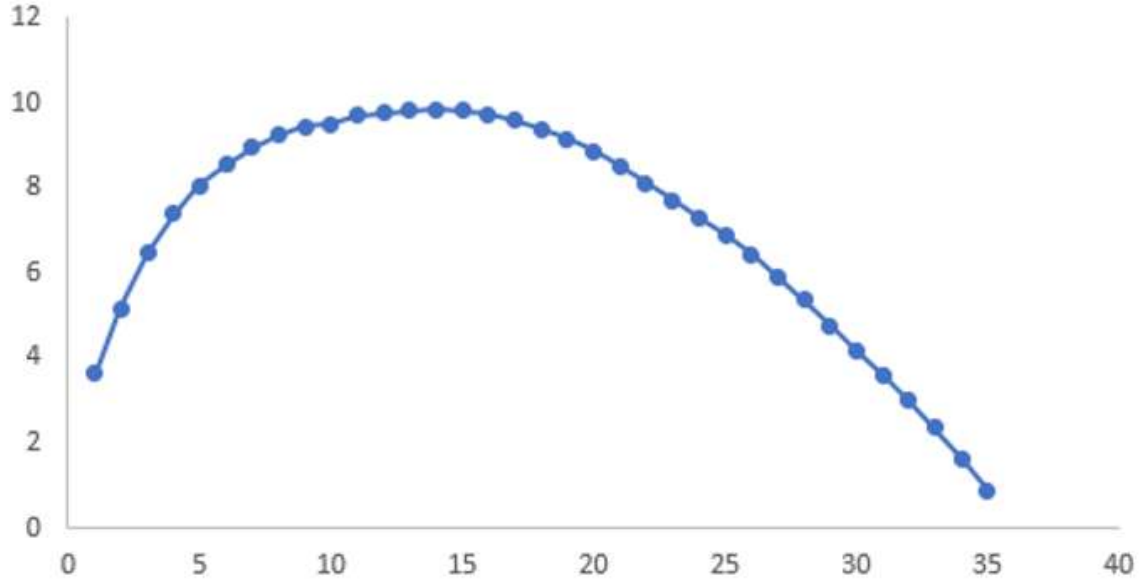
Ainsi en prenant la loi de maintien estimée par la méthode de Kaplan-Meier et lissée par la méthode des moyennes mobiles

Espérance résiduelle à 3 ans de la loi expériementale



A titre comparatif, en prenant la Loi BCAC 2013 construite bi-dimensionnellement avec des âges arbitraires.
On obtient de manière équivalente, une forme de courbe totalement différente surtout

Espérance résiduelle à 3 ans d'après les tables du BCAC



Les 2 courbes sont sensiblement différentes sur les premiers mois d'ancienneté.
Cela est du à la construction de notre portefeuille par pathologie. En effet, la majoritité des retours des assurés indiquant les pathologies concerne les arrêts de travail supérieur à 35 jours. Nous n'avons par conséquent que très peu d'arrêt de travail de moins d'un mois alors qu'ils représentent en réalité l'immense majorité. La loi de maintien en incapacité de travail est surestimée à partir de données provenant d'un panel d'organismes d'assurance.

Ainsi, nous observons dans un premier temps énormément de sorties (ce qui n'est pas le cas de notre base de données) et donc des probabilités de maintien qui augmentent considérablement avant de se stabilliser au bout de quelques mois entrainant une diminution de l'espérance résiduelle .

Pour le calcul des provisions par les compagnies d'assurance, le taux utilisé que l'on appelle taux technique est reglementé. Il est déterminé en fonction du Taux Moyen des emprunts d'état (TME)
En prenant comme taux d'actualisation, le taux technique égal à 75% du TME Moyen sur 24 mois. (taux maximum)
Il faut ainsi prendre 75% de 0.88 (taux au 31 decembre 2022)

6.2 Etude des Boni-mali sur une période d'un an

6.2.1 Calcul des Boni mali

Par construction, notre loi de maintien étant très particulière, il est nécessaire de vérifier rétroactivement que celle-ci est bien cohérente avec notre portefeuille année par année. C'est pour cela que nous allons réaliser un backtesting de la table en calculant les boni-mali sur une période d'un an.
Nous choisissons arbitrairement une date d'évaluation N (comprise en 2016 et 2021), au 1er janvier de l'année N, nous sélectionnons les sinistres en cours et nous estimons les montants à verser sur un an jusqu'au 31/12/N (en partant du principe que l'on indemnise 1€ de rente par mois)
Le calcul des Boni-Mali se fait ensuite en soustrayant ce qui a réellement été versé au cours de cette période.

BONI-MALI = $\frac{Esperance\ résiduelle\ sur\ un\ an - Nombre\ de\ jours\ indémnisés}{Nombre\ de\ jours\ indémnisés}$

Nos données sont principalement concentrées sur les dernières années (2019-2020-2021), c'est pourquoi il est plus judicieux ici d'utiliser l'espérance résiduelle sur 1 an. Nous nous intéresserons aux années 2019,2020 et 2021.

L'espérance résiduelle sur 1 an est définie par le coefficient de provisionnement unitaire suivant :
$\forall k \in [0; 35]$,

$$E_a = \frac{1}{2} \sum_{k=a}^{\min(35, a+11)} \frac{l_k + l_{k+1}}{l_a}$$

Cette formule permet de calculer l'espérance résiduelle pour une ancienneté entière (en mois). Dans le cas d'ancienneté non entière, nous utiliserons une extrapolation linéaire afin d'avoir des résultats précis.

On pose :

- a : l'ancienneté exacte
- a_{sup} : l'ancienneté entière supérieure à l'ancienneté exacte
- a_{inf} : l'ancienneté entière inférieure à l'ancienneté exacte
- $E_{sup}\ et\ E_{inf}$: les espérances résiduelles avec des anciennetés entières

$$E_a = (a_{sup} - a) \times E_{inf} + (a - a_{inf}) \times E_{sup}$$

6.2.2 Comparaison

En calculant les provisions pour l'année, on remarque que tous les résultats sont assez proches. En effet, comme nous avons vu précédemment, les estimateurs de Kaplan-Meier et de Nelson Aalen sont asymptotiquement équivalents.
L'estimateur de Kaplan-Meier nous apparait cependant légèrement plus prudent. Par ailleurs, le lissage n'a pas un énorme impact sur les résultats ce qui est cohérent avec l'analyse précédente. A chaque instant t, $\hat{S}_{FH}(t) \geq \hat{S}_{KM}(t)$

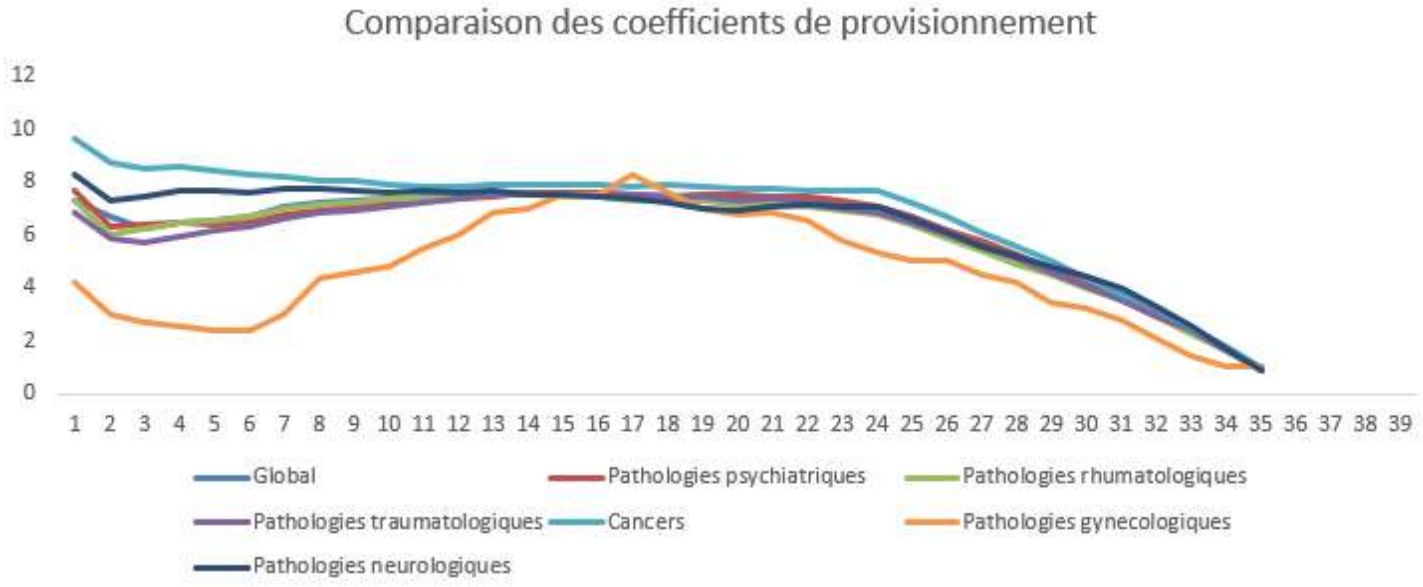
Nous choisirons pour la suite l'estimateur de Kaplan-Meier lissé par la méthode des moyennes mobiles d'ordre 7

	Provision 2019	Provision 2020	Provision 2021
Portefeuille	2 764 965	3 838 989	4 978 622
Kaplan-Meier + MM7	2 903 037	3 894 373	5 123 670

6.2.3 Impact de la segmentation

Nous avons construit nos lois de maintien par pathologie. Afin de déterminer l'impact potentiel de la segmentation sur nos lois de durée.

Toujours en utilisant l'estimateur de Kaplan-Meier associé à un lissage des moyennes mobiles d'ordre 7, calculons nos coefficients de provisionnement par pathologie.



Les familles de pathologie impactent grandement le provisionnement.

On remarque que les coefficients de provisionnement à un an associés à la pathologie "cancers" sont globalement les plus élevés alors que les coefficients de provisionnement à un an associés aux pathologies gynécologiques sont les plus faibles. Cela est cohérent par rapport aux études précédentes.

	Provision 2019	Estimation 2019	Boni-mali 2019 v	Provision 2020	Estimation 2020	Boni-mali 2020	Provision 2021	Estimation 2021	Boni-mali 2021
Pathologies psychiatriques	830 495	873 457	+1,05%	1 195 013	1 210 135	+1,27%	1 599 255	1 757 486	+10,76%
Pathologies rhumatologiques	504 736	500 893	-0,76%	738 214	772 785	+4,68%	1 016 096	1 067 899	+5,09%
Pathologies traumatologiques	491 843	482 075	-1,98%	572 435	604 568	+5,61%	658 499	701 537	+6,53%
Cancers	406 109	359 124	+11,57%	497 404	565 712	+13,73%	594 988	624 001	+4,88%
Pathologies gynécologiques	109 518	111 478	+1,78%	157 346	150 683	-4,23%	182 745	189 452	+3,67%
Pathologies neurologiques	104 511	113 455	+8,56%	172 425	164 580	-3,42%	239 627	233 058	-2,74%
Pathologies cardiovasculaires	75 047	81 241	−2,58%	102 795	113 858	+10,76%	160 552	161 957	+0,8%

Sur la période 2019-2021, les estimations obtenues avec les lois lissées dégagent des bonis sur les 3 années de survenance. La table d'expérience construite est prudente.

La segmentation par pathologie en revanche présente quelques irrégularités. Cependant on remarque que plus la volumétrie est importante, plus il tend à avoir des bonis.

7 Estimation du gain des contrôles médicaux sur le risque arrêt de travail

L'objectif de cette étude est d'aider la direction médicale dans le contrôle des arrêts de travail (prévoyance collective) en améliorant la détection des dossiers à forte probabilité de fraude.

Avec un nombre de salariés arrêtés supérieur à 40% chaque année depuis 2016, l'absentéisme maladie reste un problème majeur. Dans ce contexte de hausse constante du nombre d'incapacités, Malakoff Humanis a développé ces dernières années un processus de contrôle médical afin de mieux comprendre les raisons de ces arrêts de travail.

Cette partie vise premièrement à présenter le dispositif et les différents axes permettant d'améliorer la rentabilité économique des contrôles en réduisant les coûts d'expertises.
Pour cela, un meilleur ciblage sera visé en augmentant le nombre de nouveaux abus détectés et en diminuant les prestations liées à l'arrêt des paiements indus.
Nous calculerons ensuite le gain brut des contrôles médicaux sur l'année 2021.

7.1 Fonctionnement du dispositif de contrôle des arrêts de travail

7.1.1 Présentation du dispositif

Le contrôle médical est effectué sur la base d'un certificat médical d'incapacité de travail (CMIT), adressé au salarié en arrêt de travail d'au moins 30 jours ou à période de franchise atteinte, selon les contrats, qu'il devra compléter ainsi que son médecin et retourner à l'attention du médecin conseil au service médical. Ce document est couvert par le secret médical.

Compte tenu des coûts de gestion, il n'est pas intéressant d'un point de vue financier de déclencher les contrôles en dessous du 31ème jour.

Sur la base du CMIT complété, la cellule médicale juge si des documents complémentaires et/ou une expertise médicale sont nécessaires. Le SMC effectue un contrôle des arrêts de travail et les arrêts de travail non justifiés sont classés en 3 catégories :

- non médicalement justifié
- exclusion médicale contractuelle
- fausse déclaration

Des contrôles médiaux sont aussi possibles sur :

- les risques décès : pour connaître notamment la pathologie et respecter la contrainte de portabilité (recherche d'antériorité).
- les risques arrêt de travail : pour contrôler si l'arrêt est ou non justifié, pour une recherche d'antériorité, pour une orientation éventuelle vers le service ARE, pour l'identification des cas de recours contre tiers, des « fausses » rechutes, et des fausses déclarations (à l'adhésion) entrainant la nullité du contrat.

Nous nous focaliserons ici bien évidemment sur les contrôles médicaux liés au risque arrêt de travail.

On note que le CMIT est rempli par l'assuré. En effet, dans ce cas de figure, l'ordre des médecins conseille aux praticiens d'être prudents en laissant les patients transmettre leurs propres informations. Le médecin traitant n'a pas à remplir, signer, apposer son cachet ou contresigner un questionnaire de santé.

De plus, le service médical ne peut en aucun cas, contacter ou demander des informations au médecin traitant ni obtenir la première page de l'arrêt de travail ou figure les éléments médicaux.

Ce type de fonctionnement peut pour l'assureur engendrer des difficultés voir des erreurs dans la classification par pathologie. Par exemple dans notre base de données, un nombre non négligeable d'arrêts de travail très longs sont référencés avec une pathologie COVID.

Cela pose la question de la fiabilité des pathologies indiquées sur le CMIT et donc de la fiabilité de nos données par pathologie. Dans ce contexte, les arrêts de plus de 60 jours seront systématiquement expertisés.

7.1.2 Expertise médicale

A la suite d'un contrôle médical, le service médical de Malakoff Humanis peut demander une expertise médicale.

Dans le cas d'une garantie incapacité, l'assureur versant des indemnités peut en effet demander une expertise médicale afin d'évaluer l'état de santé de l'assuré en arrêt de travail.

Le principal objectif de cette démarche est de fournir aux gestionnaires des informations supplémentaires, telles que la pathologie associée à l'arrêt de travail, la durée restante ou les différents traitements pris par l'assuré.

Cette expertise médicale réalisée par un médecin tiers, permettra au groupe de délibérer sur le caractère abusif ou non de l'arrêt de travail.

Le coût moyen d'une expertise médicale pour Malakoff Humanis étant de 400 €, il est nécessaire de bien cibler les arrêts de travail à contrôler.

Les différents critères permettant au service médical de demander une expertise médicale sont

- Durée longue, au regard de la pathologie
- Absence de date de reprise prévisionnelle
- Pathologies psychiatriques et rhumatologiques
- Indemnités journalières élevées

A réception des CMIT dûment complétés par les assurés en incapacité, les gestionnaires du SMC les analyseront et pourront, après consultation du médecin conseil, demander une expertise médicale pour les arrêts de travail jugés douteux ou anormalement longs.

7.1.3 Périmètre soumis au contrôle médical

Le contrôle est une pratique de gestion qui peut être mise en œuvre même si les dispositions contractuelles ne prévoient pas explicitement le principe des contrôles.

En revanche, excepté le cas des contrôles à la suite d'une sélection médicale à l'adhésion, la suspension des paiements liée au contrôle n'est possible que si les dispositions contractuelles le permettent.

S'il y a eu sélection médicale à l'adhésion

- Le service gestionnaire bloque la prestation immédiatement : aucun paiement avant rendu de l'avis du SMC (contrôle médical)

S'il n'y a pas eu de sélection médicale à l'adhésion,

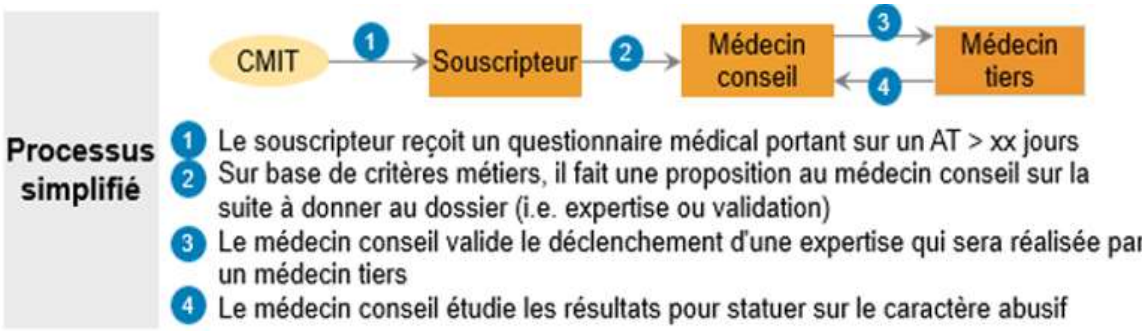
- Le paiement des prestations aux entreprises et le contrôle médical, qui donnera éventuellement lieu à suspension des paiements, sont enclenchés.

Le contrôle est donc déclenché parallèlement à la mise en gestion du sinistre.

La suspension du paiement à la suite d'un contrôle est conditionnée par la présence dans les dispositions contractuelles applicables (conditions générales) au dossier concerné de clauses prévoyant la possibilité de suspendre le paiement si le contrôle montre que l'arrêt de travail n'est pas justifié et ce, même si la CPAM poursuit le paiement de ses prestations.

Notons que si l'assuré répond tardivement à la demande de pièces complémentaires et que le paiement avait été suspendu par le SMC, le paiement sera rétroactif si acté par l'avis médical.

En résumé, les différentes étapes du processus du contrôle médical :



Source : Malakoff Humanis

7.2 Calcul de l'efficacité des contrôles médicaux en 2021

Dans le cadre de notre mémoire, le calcul de l'efficacité sera effectué sur l'année 2021, année où les données sont les plus exhaustives. Ce calcul reposera sur l'hypothèse qu'il n'y a de gain que sur les

arrêts expertisés qui ont été déclarés non justifiés et pour lesquels il y a eu cessation de paiement de prestations. Le gain brut résultant du contrôle médical correspond au montant de prestations économisé du fait du raccourcissement de la durée en arrêt de travail par la cessation de paiement des prestations incapacités.

Le nombre de jours économisés est défini comme la différence entre le nombre de jours théoriques d'un arrêt expertisé et la durée moyenne observée de la pathologie à l'issue du délai d'expertise pour les dossiers non justifiés.

Nous estimerons ainsi le nombre de jours économisés par les contrôles médicaux sur le risque incapacité. Pour chaque arrêt non justifié, la durée de l'arrêt (de la date de survenance à la date de cessation de paiement) sera comparée à la durée théorique modélisée grâce à l'estimateur de Kaplan- Meier.

7.2.1 Gain brut pour un arrêt expertisé non justifié :

On pose le gain brut comme :

$$\textbf{Gain brut : nombre de jours économisés} \times \textbf{IJ moyenne}$$

On en déduit le gain net en retranchant au gain brut les coûts de l'expertise

$$\textbf{Gain net} = \textbf{Gain brut} - \textbf{coût d'une expertise}$$

On note **ROI** le retour sur investissement (Return on Investment). Il s'obtient en calculant :

$$\frac{\sum gains\ bruts}{\sum coûts\ des\ expertises\ SMC}$$

La somme des coûts des expertises concerne tous les arrêts expertisés, pas seulement ceux qui sont déclarés non justifiés par le SMC

Le caractère dissuasif des expertises n'est pas pris en compte, puisqu'il semble difficile de distinguer les expertises "réellement dissuasives" (c'est bien l'expertise qui déclenche un retour au travail) des expertises "faussement dissuasives" (il y a eu expertise, suivie d'une reprise du travail, mais qui était en fait prévue).

7.2.2 Exemple fictif

Considérons 4 incapacités déclarées non justifiés :

Assuré	Famille de pathologie	Durée de l'arrêt (A)	Délai d'expertise (B)	IJ versée (C)	Espérance à la date d'expertise (D)	Gain en jours (E)=(B)+(D)-(A)	Gain en € (E) * (C)
1	Psychiatrique	267	185	42	245	163	6 846
2	Neurologique	224	190	64	325	291	18 624
3	Cancer	628	275	64	309	0	0
4	Traumatologie	327	432	59	292	397	23 423

L’espérance à la date d’expertise est calculée à partir de notre étude sur les durées par famille de pathologie. Le gain brut peut ainsi être calculé

7.3 Gains obtenus

Sur l’année 2021, 241 arrêts expertisés ont été déclarés non justifiés. Ces arrêts de travail représentent un gain brut de 2 343 224 €.

Parmi ces résultats :

- Près de 2/3 des arrêts non justifiés (150) concernent les pathologies psychiatriques
- Le gain sur les pathologies psychiatriques représente à lui seul 1 560 948 €
- Seul un arrêt présentant la pathologie « Cancers » a été déclaré non justifié.

Pour les 4 principales familles de pathologie.

Nombre d’arrêts non justifiés	Famille de pathologie	Gain brut
150	Psychiatrique	1 560 948
33	Rhumatologique	209 129
19	Traumatologie	192 378
1	Cancers	51 585

Il est plus avantageux d'orienter les expertises sur les pathologies psychiatriques.

Avec un gain brut en 2020 de 870 K€ et 80 arrêts déclarés non justifiés. Ce chiffre est nettement supérieur aux autres pathologies. Les arrêts d’origine psychiatrique sont les plus expertisés.

Les arrêts en traumatologie avec les arrêts d’origine psychiatrique sont le plus souvent classés comme non justifiés à la suite d’une expertise.

Ce calcul de gain peut être influencé à la hausse. En effet l’effet dissuasif du contrôle médical n’est pas pris en compte (c'est à dire le fait que l'envoi du certificat raccourcisse la durée totale de l'arrêt de travail). Ce dernier ne pourra être évalué qu’avec un recul d’au minimum 2 ans mais les statistiques en la matière montrent des pourcentages d’environ 10%. Nous pourrons le mesurer si la durée moyenne de l’arrêt de travail par pathologie diminue.

7.4 Amélioration du dispositif

Afin de maximiser les gains, il est proposé de se concentrer quasiment exclusivement sur les pathologies psychiatriques en essayant de réduire les délais d’expertise.
Cependant des pathologies indiquées dans le CMIT par l’assuré (sans nécessité de fournir un justificatif) pourraient se révéler inexactes. C’est pourquoi le questionnaire médical doit être le plus explicite possible.

De plus une augmentation des questionnaires médicaux permettrait d’avoir un historique s’assurant ainsi de la cohérence de la pathologie indiquée et permettrait également dans certain cas une identification de la cause du premier arrêt (information importante dans le cadre d’un litige avec un

autre assureur conséquence de la contrainte de portabilité (Loi EVIN) pour connaître la date de survenance : le fait générateur est-il avant ou après le changement d'assureur ?)

Le développement du service ARE – Mon Accompagnement Reprise – qui prévoit un suivi des individus au-delà du domaine purement médical serait également une solution–pertinente pour limiter la dérive de la sinistralité du risque AT. Cet accompagnement pourrait être utilisé afin d'aider les assurés en arrêt de travail pour cause de pathologies psychiatriques à mieux appréhender, préparer et envisager leur retour au travail réduisant par la même occasion le risque de rechute.

Conclusion

Les certificats médicaux d'incapacité de travail (CMIT) envoyés par Malakoff Humanis aux assurés en incapacité de travail permettent de recueillir des informations parmi lesquelles les pathologies médicales à l'origine des arrêts de travail.

Dans ce cadre, l'objectif de ce mémoire était de construire des lois de maintien en incapacité par famille de pathologie. Pour cela, un travail conséquent de nettoyage et de réconciliation a été effectué sur les données fournies par le SMC et les différentes extractions réalisées sous SAS.

Pour l'estimation de la fonction de survie, les méthodes non paramétriques de Kaplan-Meier et de Fleming-Harrington ont été étudiées. Afin de prendre en considération certaines variables pouvant également avoir un impact sur la durée de maintien, nous avons testé un modèle de Cox. La non-vérification de l'hypothèse du hasard proportionnel nous a obligé à ne pas appliquer ce modèle.

Les taux bruts de sortie précédemment estimés ont été ensuite lissés par la méthode des moyennes mobiles et de Whittaker-Henderson. Les résultats étant très proches, nous avons retenu le modèle de Kaplan-Meier associé à la méthode de lissage des taux bruts par moyenne mobile.

L'analyse des boni-mali effectuée sur le portefeuille global puis sur la loi segmentée par famille de pathologie s'est montrée satisfaisante. L'étude visait à comparer les indemnités versées par Malakoff Humanis sur les 3 dernières années aux indemnités estimées par notre modèle retenu. Ce backtesting a permis de conclure à une relative prudence de la table d'expérience globale construite car pour les années de survenance étudiées, on aboutit à un « sur-provisionnement » de la table d'expérience

Enfin, ces résultats ont été appliqués dans le but de calculer l'efficacité des contrôles médicaux. En s'appuyant sur la liste des arrêts de travail expertisés et déclarés comme non justifié, nous avons comparé la durée de l'arrêt de travail par rapport à la durée théorique prévue par le modèle étudié précédemment. Cette étude de gain permettra dans le futur d'orienter la demande d'expertise médicale en ciblant principalement les pathologies anormalement longues, les pathologies psychiatriques ou le risque d'abus est le plus élevé.

Ce dispositif ARE d'aide au retour à l'emploi conçu et mis en place pour encourager une reprise en douceur des salariés en arrêts de travail pourrait se baser sur les résultats de cette étude pour cibler les pathologies nécessitant un réel accompagnement. Une meilleure compréhension des arrêts et des durées moyennes en fonction de la pathologie améliorerait l'efficacité du dispositif proposé par le groupe.

Bibliographie

[1] ELLY ASSURANCES. La Sécurité Sociale

[2] FRANCE ASSUREURS. Le marché des assurances santé et prévoyance en 2020

[3] V. BENOIT. Prévoyance collective. Support de cours ISFA

[4] Légifrance

[5] Portail de la fonction publique

[6] Baromètre Absentéisme Malakoff Humanis

[7] F. PLANCHET. Modèles de durée. Support de cours ISFA

[8] Institut des Actuaires. LIGNES DIRECTRICES DE LA CONSTRUCTION DES LOIS DE MAINTIEN EN INCAPACITÉ ET EN INVALIDITÉ

[9] S. QUANTIN. Modèles semi-paramétriques de survie en temps continu sous R.

[10] O. BOUAZIZ. Analyse de survie : le modèle de Cox

[11] BAGUI.Refonte des lois de maintien en incapacité temporaire de travail. MémoireISFA.2013

[12] E. BORGOMANO. Construction et certification d'une table de maintien en incapacité. Mémoire ISFA.2020

[13] M. SILLE. Conception des lois de maintien d\'expérience en incapacité et lois de passage pour les agents des fonctions publiques territoriale et hospitalière Mémoire ISFA.2018

[14] A. RAHMOUNI. Construction de lois expérimentales en incapacité temporaire de travail. Mémoire ISUP.2020

[15] Z. RAITI. Modélisation de la durée de maintien en arrêt de travail d'une population de travailleurs non-salariés. Mémoire ISUP 2017

Annexe

Extrait d'un CMIT

NOM : ..... **PRENOM :** .....

A REMPLIR PAR L'ASSURE

Date d'arrêt de travail : ..... Accident ☐ Accident du travail ☐ Maladie ☐

S'il s'agit d'une rechute, précisez la date du premier accident ou maladie : .....

Nature des blessures ou de la maladie : .....

.....

En cas d'accident, précisez les circonstances : .....

Quel traitement médical suivez-vous ? (*joindre les copies d'ordonnances)* : .....

.....

A quelle date avez-vous ressenti les premiers symptômes ? .....

Aviez-vous déjà consulté un médecin pour cette affection ? OUI ☐ NON ☐

Si **OUI**, à quelle(s) date(s) ? ...../...../..... ...../...../.....

Quel traitement aviez-vous suivi ? (médicaments, kinésithérapie, autres soins) : .....

.....

Avez-vous été opéré ? OUI ☐ NON ☐

Si **OUI**, précisez la date et la nature de l'intervention chirurgicale (*joindre une copie du compte rendu*) :

.....

Avez-vous été hospitalisé ? OUI ☐ NON ☐

Si **OUI**, à quelle(s) date(s) ? ...../...../..... ...../...../.....

Aviez-vous déjà été en arrêt de travail pour cette affection ? OUI ☐ NON ☐

Si **OUI**, précisez les date(s) et durée(s) : .....

Antécédents Médicaux et Chirurgicaux

Etes-vous suivi pour une autre affection médicale ? OUI ☐ NON ☐

Si **OUI**, laquelle et depuis quand ? ❶ ..... ❷ ..... ❸ .....

❶ ...../...../..... ❷ ...../...../..... ❸ ...../...../.....

Suivez-vous un traitement médical pour cette (ces) autre(s) affection(s) ? OUI ☐ NON ☐

Si **OUI**, lequel (lesquels) et depuis quand ? .....

.....

Avez-vous déjà subi une intervention chirurgicale ? OUI ☐ NON ☐

Si **OUI**, laquelle et à quelle(s) date(s) ? .....

Avez-vous été hospitalisé ? OUI ☐ NON ☐

Si **OUI**, pour quelle(s) affection(s) et à quelle(s) date(s) ? .....

.....

Avez-vous bénéficié d'arrêts de travail ? OUI ☐ NON ☐

Si **OUI**, date(s) durée(s) et motif(s) .....

Avez-vous des séquelles d'accident ou de maladie ? OUI ☐ NON ☐

Si **OUI**, lesquelles et depuis quand ? .....

Etes-vous exonéré du Ticket Modérateur pour une affection de longue durée ? OUI ☐ NON ☐

Pour quelle(s) affection(s) et depuis quand ? .....

Autres Informations

Bénéficiez-vous d'une :	pension d'invalidité	☐	rente accident du travail	☐
	rente maladie professionnelle	☐	allocation tierce personne	☐

Si **OUI**, précisez : Catégorie de la pension ..... Taux de la rente d'incapacité ..... %

A quelle date et pour quel motif avez-vous bénéficié de cette pension, rente ou allocation ? .....

Avez-vous repris une activité professionnelle ? OUI ☐ NON ☐

Si **OUI** : à temps complet ☐ à temps partiel ☐ à mi-temps thérapeutique ☐ à quelle date ...../...../.....

Quelles sont les démarches en cours :	Reclassement professionnel	OUI ☐	NON ☐
	Demande d'allocation Tierce personnes	OUI ☐	NON ☐
	Demande de mise à la retraite	OUI ☐	NON ☐

Précisez : Votre poids actuel ..... Kg Taille : ..... Votre latéralité : Droitier ☐ Gaucher ☐

Je soussigné(e) déclare les réponses ci-dessus complètes et sincères.

Date : ..... **Signature de l'assuré :**