

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

Par : Élise Santoro

Titre du mémoire : Mise en place d'un outil d'aide au pilotage autour du  
déploiement de la réforme « 100% santé »

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 Actuares :*

\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

*Membres présents du jury  
de la filière :*

\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

*Entreprise :*

*Nom : ADDACTIS France*

*Signature :*

*Directeur du mémoire en entreprise :*

*Nom : Daniel Attias*

*Signature : D. Attias*

*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 :

*D. Attias*

Signature du candidat :

*Elise Santoro*

# RÉSUMÉ

**Mots clés :** assurance santé, réforme « 100% santé », audiologie, dentaire, optique, segmentation, analyse comportementale, apprentissage statistique.

Le système de santé français connaît régulièrement des réformes. Cela lui permet de s'adapter aux nouveaux enjeux économiques, et de conserver sa fonction première de protection des citoyens. En 2018, le Sénat a voté la réforme « 100% santé », afin de réduire le renoncement aux soins en proposant des soins audiologiques, dentaires et optiques remboursés en totalité.

Ce remboursement sera en partie pris en charge par la Sécurité Sociale et les complémentaires santé. Dans ce cadre, les assureurs cherchent à établir des modèles d'estimation d'impact de cette réforme sur leurs prestations. Ce mémoire présente un outil d'aide au pilotage suite au déploiement de cette réforme.

Après la description du fonctionnement du système de santé français, la première partie de ce mémoire propose une étude détaillée de la réforme « 100% santé » pour en comprendre toutes ses spécificités.

La seconde partie propose une étude de l'impact de la réforme sur le montant des prestations versées par l'assureur, aussi bien suite aux modifications propres à la réforme qu'au potentiel changement de comportement des assurés. Cette étude repose nécessairement sur de nombreuses hypothèses, c'est pourquoi plusieurs scénarios sont mis en place et l'impact chiffré de la réforme est présenté sous forme d'intervalle.

Enfin, l'aide au pilotage d'un portefeuille repose également sur une meilleure connaissance de ce dernier. La troisième partie de ce mémoire propose donc une méthode de mise en place et de validation d'une segmentation de portefeuille. Les arbres CART seront utilisés pour la création de la segmentation, alors que des algorithmes tels que le GLM, les forêts aléatoires ou le Gradient Boosting permettront de vérifier la qualité des résultats de l'arbre CART.

Les résultats de cette étude sont à relier à la base de données d'un client d'Addactis et aux hypothèses prises en accord avec ce dernier. La méthodologie utilisée et l'outil créé dans ce mémoire sont néanmoins utilisables sur d'autres bases de données.

# ABSTRACT

**Key words :** Healthcare insurance, « 100% santé » reform, audiology, dental, optics, segmentation, behavioural analysis, statistical learning

The French healthcare system is undergoing regular reforms. This allows it to adapt to new economic challenges and to maintain its primary function of protecting citizens. In 2018, the French Senate voted the « 100% santé » reform in order to reduce the renunciation of healthcare by offering fully reimbursed audiology, dental and optical care.

A part of this reimbursement will be paid by the French social security and another by the healthcare insurers. Therefore, insurers seek to implement models to estimate this reform's impact on their payments. This work therefore presents a tool to help them manage the implementation of this reform.

After describing how the French healthcare system works, the first part of this work proposes a detailed study of the « 100% santé » reform which is necessary in order to understand all its specificities.

The second part proposes a study of the impact of the reform on the amount of payments made by the insurer, both as a result of the modifications specific to the reform and the potential change in the behaviour of the insured. This study is necessarily based on numerous hypotheses, which is why several scenarios are set up and the quantified impact of the reform is presented in the form of an interval.

Finally, assistance in managing a portfolio also depends on a better knowledge of it. The third part of this paper therefore describes a method for setting up and validating portfolio segmentation. CART trees will be used to create the segmentation, while algorithms such as GLM, Random Forests or Gradient Boosting will be used to check the quality of the results of the CART tree.

The results of this study must be linked to the database of a particular client of Addactis and to the hypothesis taken in agreement with this one. The methodology used and the tool created in this work are nevertheless usable on other databases.

# NOTE DE SYNTHÈSE

Les assureurs doivent très régulièrement revoir leurs produits et leurs tarifications suite aux mesures gouvernementales mises en place. Dans le domaine particulier de la santé, l'accès aux soins est une priorité en France. Néanmoins, il s'avère que de nombreux français déclarent encore devoir renoncer à des soins faute de moyens. Face à ce constat, le gouvernement a mis en place la réforme « 100% santé », qui modifie les prises en charge pour les postes audiologie, dentaire et optique et offre un panel de soins sans reste à charge.

La volonté du gouvernement à travers cette réforme est donc de réduire le renoncement aux soins. Toutefois, en parallèle, de nombreuses questions sont soulevées : les tarifs des complémentaires santé vont-ils augmenter pour faire face à une probable hausse des remboursements ? Le comportement des assurés va-t-il réellement changer suite à cette réforme, et dans quelle mesure ? Comment les assureurs peuvent-ils au mieux anticiper les différentes conséquences de cette réforme ?

C'est dans ce cadre d'interrogations que s'inscrit ce mémoire. Il a pour but de proposer aux assureurs un outil d'aide au pilotage de leurs offres de produits santé. Pour cela, il repose sur deux axes. Tout d'abord, l'outil propose une projection des montants de remboursements complémentaires afin de mesurer l'impact chiffré de la réforme lors de la mise en « As-if » des prestations de 2018. Cette évaluation de l'impact leur permettra de connaître une première mesure des conséquences de la réforme en ne prenant en compte que les modifications purement techniques des garanties qu'ils doivent proposer. Ensuite, une étude sur la modification du comportement des assurés suite à la mise en place de la réforme offre à l'assureur un intervalle d'impacts selon différents scénarios (du moins risqué au plus risqué). Enfin, l'outil propose également une segmentation du portefeuille qui peut permettre à l'assureur d'en avoir une meilleure vision et donc de mieux appréhender les conséquences de la réforme sur son propre portefeuille.

## La réforme « 100% santé »

La réforme « 100% santé », votée le 4 décembre 2018, établit trois paniers de soins.





La réforme se décline selon les différents postes de soins. Pourront faire partie du panier « 100% santé » :

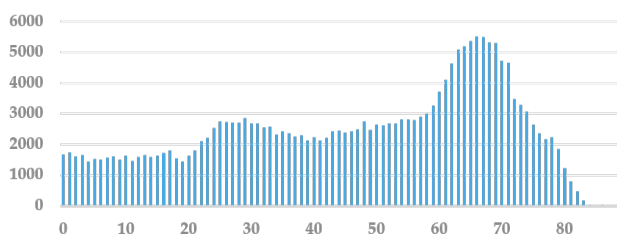
- les prothèses auditives pour le poste audiologie ;
- les prothèses dentaires ainsi que certains soins conservateurs pour le poste dentaire ;
- les équipements (les verres optiques et les montures) pour le poste optique.

De plus, pour chaque acte de ces trois postes, la base de remboursement de la Sécurité Sociale est modifiée selon le panier de soins (elle est généralement augmentée pour le panier « 100% santé » et réduite pour le panier libre). Des prix limites de vente sont également mis en place dans le but de limiter le coût total des soins et donc l'augmentation des remboursements des complémentaires santé.

Par ailleurs, la mise en place de la réforme « 100% santé » se déroule sur plusieurs années afin de laisser le temps aux différents acteurs de s'y préparer. Ainsi, en 2020 les équipements optiques ainsi que les couronnes dentaires et les bridges du panier « 100% santé » seront sans reste à charge pour les assurés et en 2021 ce sera également le cas de tous les actes dentaires et les prothèses auditives. Cependant, dès 2019, les limitations des prix des actes ainsi que la modification des bases de remboursement commencent à être effectives.

## Portefeuille étudié

Le portefeuille étudié est uniquement lié à des contrats de santé individuels, ce qui en fait un portefeuille assez vieillissant :

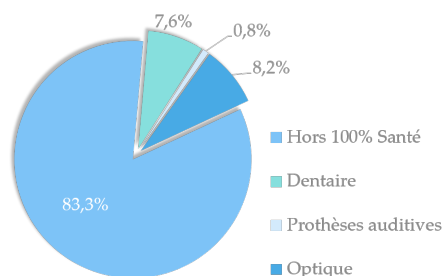


Répartition du portefeuille, en terme d'exposition, selon l'âge des assurés

Par ailleurs, l'offre de l'assureur se décline en trois gammes : Solo, Famille et Senior, chacune développée en trois niveaux de garantie, du plus bas au plus haut. Il s'avère que la gamme Senior représente 42% du portefeuille en termes d'exposition, alors que les gammes Solo et Famille représentent chacune 29%. De plus, la grande majorité des assurés (81%) a souscrit à une offre entrée de gamme, ce qui expose d'autant plus le portefeuille aux conséquences de la réforme « 100% santé ».



Enfin, l'impact calculé dans l'outil ne s'applique que sur les prestations des trois postes concernés par la réforme « 100% santé ». Ainsi, il est à relativiser par rapport à la somme totale des prestations versées par l'assureur en santé. Lors de l'étude du portefeuille il apparaît que les trois postes ne représentent que 17% des remboursements en 2018.



## Impacts de la réforme « 100% santé » sur les prestations de l'assureur

Le calcul des impacts de la réforme « 100% santé » se déroule en deux temps. Tout d'abord, ils sont calculés sans prendre en compte la déformation de consommation de soins des assurés qui devrait accompagner la mise en place du panier « 100% santé ». Ensuite sont ajoutées lesdites déformations afin d'obtenir un impact plus réaliste de la réforme selon le scénario choisi. En effet, les hypothèses à faire pour cette deuxième partie de calcul sont fortes, et il n'est pas raisonnable de ne considérer qu'un seul jeu d'hypothèses, c'est pourquoi plusieurs scénarios seront proposés.

Avant de pouvoir projeter les montants de prestations payés par l'assureur, il est nécessaire de mettre correctement en forme la base de données ainsi que d'y ajouter les informations requises. La base de données initiale comporte les prestations payées par l'assureur en 2018, avec le code de référencement associé à chaque acte remboursé et le montant des bases de remboursements (ou BR) imposé par la Sécurité Sociale. A partir de cette base et grâce au code de référencement des actes, les nouvelles BR modifiées par la réforme (différentes au fil des années de 2019 à 2023) peuvent être ajoutées à la base. En effet, le gouvernement a publié une table reprenant les correspondances entre les anciens tarifs de la Sécurité Sociale et les nouveaux. Le code de référencement a également permis d'ajouter les prix limites de vente (ou PLV), indispensables pour la projection des remboursements.

Une fois la base de données agrémentée de ces informations, il a été nécessaire de poser un certain nombre d'hypothèses afin de mener à bien l'étude.

- Tout d'abord, il n'est pas rare que les assureurs modifient leurs tableaux de garanties pour des raisons autres que l'application d'une nouvelle réforme (comme par exemple pour des questions de rentabilité ou d'ajustement sur le marché). L'étude effectuée dans ce mémoire ayant pour but de mesurer uniquement l'impact de la réforme « 100% santé », elle ne prendra pas en compte de possibles modifications du tableau de garanties autres que celles induites par cette réforme.
- De plus, avec l'accord du client et dans une nécessité de prudence, une hausse proportionnelle à celle des BR est appliquée aux frais payés par l'assuré (appelés frais réels, ou FR). Cette hausse s'appliquera dans la limite des PLV imposés par la réforme.
- Enfin, il est nécessaire de définir l'utilisation des PLV. Après une étude de différents moyens, l'hypothèse choisie est la suivante : la projection des frais réels est égale au

---

minimum entre le PLV et les frais réels de 2018 (eux-mêmes augmentés d'après l'hypothèse précédente).

La projection des remboursements complémentaires peut alors se faire, tout en veillant à ce que, quelque soit le montant des frais réels pour le panier « 100% santé », ils soient remboursés en totalité en 2020 pour l'optique ainsi que les couronnes dentaires et les bridges, et en 2021 pour le reste du panier « 100% santé ». Les résultats de cette projection sont les suivants :

	Entre 2018 & 2019	Entre 2019 & 2020	Entre 2020 & 2021	Entre 2021 & 2022	Entre 2022 & 2023	Entre 2018 & 2023
DENTAIRE	2,60%	9,47%	16,73%	0,49%	0,01%	31,77%
AUDIOLOGIE	2,34%	0,61%	74,98%	0,00%	0,00%	80,16%
OPTIQUE	0,00%	-5,55%	0,00%	0,00%	0,00%	-5,55%
AUTRES POSTES	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
TOTAL	0,22%	0,29%	2,07%	0,05%	0,00%	2,64%

### Impacts de la réforme sur les remboursements complémentaires

Ainsi, l'impact global de la réforme (en prenant en compte les postes non réformés tels que l'hospitalisation ou la pharmacie) résulte en une hausse de 2,64%. Une étude plus détaillée est nécessaire car l'impact diffère fortement selon le poste étudié. En effet, pour l'optique, le montant total des prestations versées par l'assureur diminue de plus de 5% alors que pour le dentaire et l'audiologie les impacts sont positifs (respectivement 32% et 80%). Cette grande hétérogénéité dans les résultats est principalement due aux différences dans la manière de tenter de réduire le renoncement aux soins selon le poste. Néanmoins, les impacts sont aussi dépendants de la base de données utilisée pour faire cette étude ainsi que des hypothèses choisies. Aussi, le portefeuille étant vieillissant et les assurés ayant principalement souscrit à une garantie entrée de gamme, il est normal de trouver en audiologie et en dentaire des impacts relativement importants. Le résultat négatif pour l'optique s'explique par le désengagement quasi total de la Sécurité Sociale sur les équipements du panier libre.

Enfin, les projections effectuées pour les montants des remboursements complémentaires permettent également de mesurer un impact sur le reste à charge des assurés, tout en restant prudent sur l'analyse car le référencement de cette variable présente certains défauts dans la base de données et les hypothèses prises réduisent parfois fortement les frais réels. Il apparaît alors qu'en moyenne le reste à charge des assurés diminue pour tous les postes, même pour l'optique.

### Impacts de la réforme « 100% santé » sur le comportement des assurés

En plus de l'impact purement technique de la réforme « 100% santé », il est fort probable que le comportement des assurés change au fur et à mesure de sa mise en place. Ce changement de comportement aura nécessairement des conséquences sur les montants de prestations des assureurs.

Tout comme pour l'impact « pur » de la réforme, il est nécessaire d'établir plusieurs hypothèses afin d'estimer l'impact de ce changement de comportement sur les prestations.

- Tout d’abord, l’étude du portefeuille s’effectue sur les contrats actifs en 2018, aucun nouveau contrat ne sera donc considéré.
- La déformation des fréquences de consommation est une augmentation et n’est appliquée que sur les actes du panier « 100% santé ». Une exception est faite pour les soins dentaires car, bien qu’ils fassent partie du panier libre dans la définition de la réforme, ils accompagnent presque systématiquement la pose d’une prothèse dentaire.
- Compte tenu de la mise en place et du contenu de la réforme relativement complexes, tous les assurés ne seront au fait de son fonctionnement en même temps. C’est pourquoi, les déformations qui sont appliquées sont pondérées par l’estimation du taux de la population qui est correctement informée sur la réforme « 100% santé ». Ce taux varie entre 36% en 2020 et 100% en 2023.
- Trois scénarios de déformation sont étudiés, un scénario central, basé sur les taux de renoncement aux soins constatés en France, puis déclinés selon la gamme et le niveau de garantie, puis un scénario minorant et un majorant dans le but de proposer à l’assureur un intervalle d’impacts possibles.

Les impacts sur les prestations complémentaires suite à l’application des déformations de fréquence de consommation des assurés sont les suivants :

	<i>Sans déformation</i>	<i>Scénario minorant</i>	<i>Scénario central</i>	<i>Scénario majorant</i>
<b>DENTAIRE</b>	31,8%	60,2%	67%	90,5%
<b>AUDIOLOGIE</b>	80,2%	158,8%	177,5%	214,5%
<b>OPTIQUE</b>	-5,6%	-4,1%	-2%	2,8%
<b>TOTAL</b>	2,6%	5,6%	6,7%	8,9%

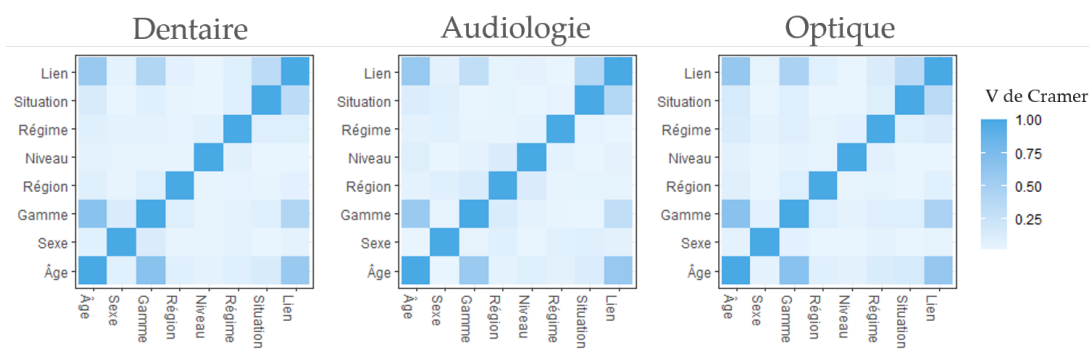
Impacts de la réforme sur les remboursements complémentaires incluant la déformation de fréquence de consommation des assurés

Cela montre que l’impact global de la réforme, déformation de consommation incluse, serait compris entre 5,6% et 8,9%. Nota bene : Il est important de rester prudent sur ce résultat fortement lié à la base de données utilisée ainsi qu’aux hypothèses de déformations choisies.

## Outil de segmentation du portefeuille

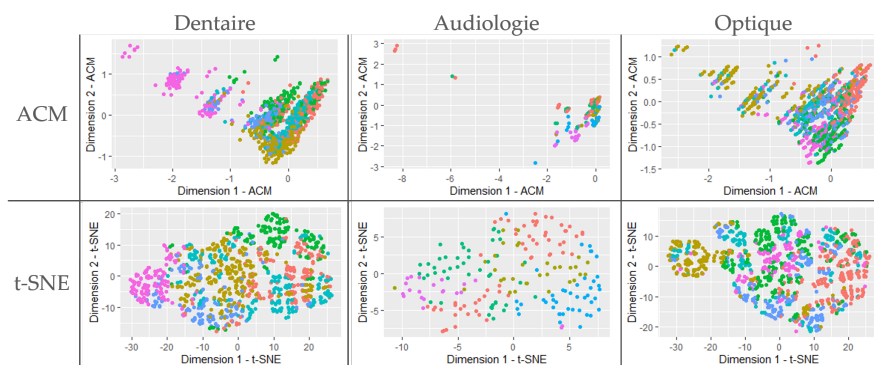
La mesure de l’impact de la réforme « 100% santé » est une considération primordiale dans l’aide au pilotage des assureurs. Pour aller plus loin, il est possible de leur proposer une segmentation plus précise de leur portefeuille, basée sur la totalité des variables dont ils disposent. En effet, cela leur permettra de mieux comprendre leur portefeuille, et donc de mieux appréhender un quelconque changement de comportement de leurs assurés.

Pour ce faire, une première étude non supervisée (c’est-à-dire qui ne repose pas sur une variable cible autour de laquelle les individus seraient regroupés) est mise en place. Cette étude a pour but de vérifier les liens intrinsèques des variables entre elles, mais aussi de tenter d’observer une première segmentation non supervisée du portefeuille. Ainsi, un test de Cramer est effectué sur les variables mises à disposition par l’assureur :



Tests de Cramer pour les trois postes de l'étude

Ce test montre que la majorité des variables sont décorrélées malgré certaines exceptions comme l'âge et la gamme de garantie ou le lien familial. Ensuite, l'algorithme PAM est mis en place dans le but d'observer la répartition des individus et peut-être des profils qui se distingueraient. Pour visualiser les résultats de l'algorithme PAM, l'ACM et l'algorithme t-SNE sont utilisés :



Visualisations des résultats des algorithmes PAM

Aucun profil ne se dégage de cette étude, ce qui montre que sans variable cible, il n'est pas possible de regrouper les individus autour des caractéristiques qui les composent. Néanmoins, cette étude permet de mieux connaître les variables qui vont être utilisées dans la suite de ce mémoire, il est en effet nécessaire que les variables sélectionnées ne présentent pas une corrélation trop importante.

C'est pour cette raison qu'une étude supervisée est menée. Elle a pour but de proposer à l'assureur une segmentation de son portefeuille autour du montant des prestations versées. Cela lui permettra ainsi de mieux identifier les contrats qui impliquent pour lui des remboursements élevés et ceux qui, pour le moment, impliquent des remboursements plus bas.

Pour cette étude, la capacité de segmentation des arbres CART est utilisée. Néanmoins, le manque de robustesse de cette méthode implique la nécessité de valider les résultats par l'utilisation d'autres modèles. Le GLM, les forêts aléatoires ainsi que l'algorithme XG Boost sont de bons candidats. Ils permettent de vérifier que la segmentation est cohérente et surtout qu'elle peut être retenue par l'assureur comme outil de pilotage. Pour ce faire, deux métriques de validations sont utilisées (le coefficient de  $R^2$  et l'indice de Gini). Ainsi, si la

---

modélisation CART obtient un trop mauvais score pour ces deux métriques par rapport aux autres modèles, cela veut dire qu'elle ne peut pas être considérée comme pertinente.

Les montants de prestations sont très différents selon les postes, c'est pourquoi trois études sont menées. Finalement, pour chacun des trois postes étudiés, la segmentation obtenue suite à la modélisation CART est satisfaisante dans le sens où elle conforte la plupart du temps les hypothèses prises dans la partie précédente d'étude de la déformation des fréquences de consommation. Néanmoins, en audiologie elle laisse supposer un renoncement aux soins chez les moins de 65 ans qui n'avait pas été anticipé.

Les assurés les plus enclins à modifier leur consommation suite à la réforme « 100% santé » sont ceux ayant souscrit à une garantie entrée de gamme. Une fracture se fait également entre les assurés ayant moins de 65 ans et les autres : pour l'audiologie, les assurés de moins de 65 ans semblent être plus sujets à un renoncement aux soins, alors que c'est l'inverse pour l'optique (quel que soit le niveau de garantie souscrit). Alors que les assurés peu enclins à modifier leur consommation de soins sont ceux ayant souscrit à des garanties haut de gamme, ce qui est cohérent avec le propos même de la réforme.

De plus, bien que les segmentations obtenues n'apportent que peu d'informations supplémentaires sur les possibles motifs de renoncement aux soins chez les assurés, elles ont permis de vérifier les hypothèses émises lors de l'étude de la déformation du comportement des assurés et d'en remettre en question certaines, notamment en audiologie. Elles offrent également à l'assureur une connaissance précise de son portefeuille, au vue des données disponibles. Ce dernier point est primordial pour lui permettre d'appréhender le mieux possible les impacts de la réforme.

## **Avertissement**

Il n'est pas rare qu'une étude dépende principalement de la qualité des données mises à disposition, mais cela n'a pas empêché l'utilisation de cet outil comme outil de vérification d'hypothèses et d'aide à une meilleure connaissance du portefeuille étudié. Toutefois, l'outil de création et de validation d'une segmentation de portefeuille mis en place dans ce mémoire a vocation à être utilisé chez d'autres assureurs. Ainsi, il est possible que ces derniers puissent l'utiliser et en tirer de meilleures conclusions en fonction de la qualité de leurs données, ainsi qu'une segmentation qui apporterait bien plus d'informations sur le portefeuille. D'ailleurs, cet outil pourrait même être utilisé hors du cadre de la réforme « 100% santé », il suffit de choisir une autre variable autour de laquelle la segmentation doit être faite. Ces méthodes de Machine Learning très peu exploitées en santé peuvent aussi s'avérer être des outils très intéressants en termes de pilotage de portefeuille.

---

# EXECUTIVE SUMMARY

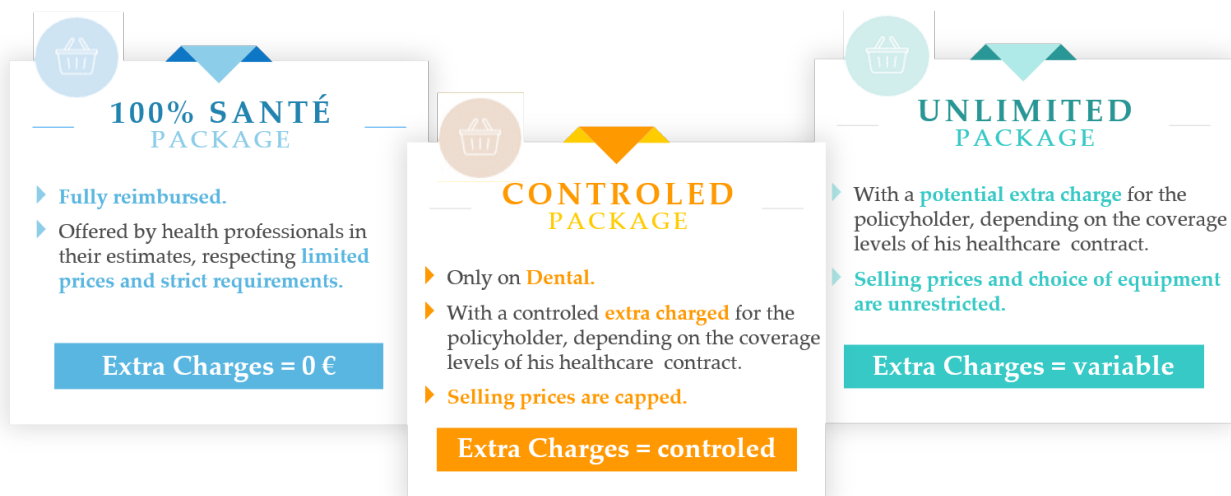
Insurers must regularly review their products and pricing following the government measures. In the specific field of healthcare, the access to care is an established political priority in France. Nevertheless, it turns out that many French people still report that they have to give up some basic cares due to a financial problem. Facing this situation, the French government has introduced the « 100% santé » reform, which modifies the reimbursement of audiology, dental and optical expenses and offers a range of healthcare services with no extra charges.

The will of the government through this reform is therefore to reduce the renunciation of healthcare. However, at the same time, many questions are being raised : will the prices of complementary health care be increased to deal with a probable increase in reimbursements? Will the behavior of the insured really change as a result of this reform, and to what extent? How can insurers best anticipate the various consequences of this reform?

It is within this questioning framework that this work is written. Its aim is to provide insurers a tool to help them manage their health products offers. To do so, it is based on two axes. Firstly, this work proposes a projection of the amounts reimbursed by insurers in order to measure the impact of the reform in figures from 2018 onwards. This impact analysis will enable them to obtain an initial measure of the consequences of the reform by taking into account only the purely technical modifications to the guarantees they have to offer. Next, a study on the change in policyholders' behaviour following the reform's introduction offers the insurer a range of impacts according to different scenarios (from the least risky to the most risky). Finally, this work also proposes a segmentation of the portfolio that can allow the insurer to have a better vision of it and thus to better understand the consequences of the reform on its own portfolio.

## The « 100% santé » reform

The « 100% santé » reform, voted on December 4, 2018, establishes three sets of healthcare packages.



---

The reform is organized according to the type of care.

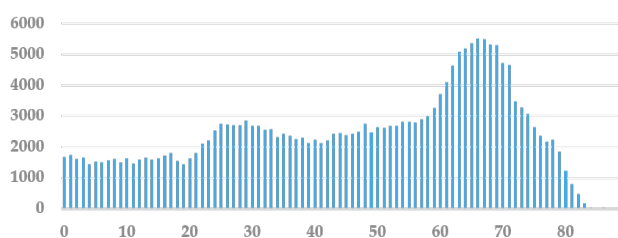
- For audiology, hearing aids may be part of the package « 100% santé ».
- For dentistry, dental prostheses as well as some conservative care may be part of the package « 100% santé ».
- For optics, equipment (optical lenses and frames) may be part of the package « 100% santé ».

Moreover, for each act of these three items, the national social security reimbursement base is modified according to the package of care (it is generally increased for the « 100% santé » package and reduced for the unlimited package). Sales price caps are also put in place in order to limit the total cost of care and thus the increase in health insurance reimbursements.

In addition, the establishment of the « 100% santé » reform is taking place over several years in order to give the various involved parties time to prepare for it. Thus, in 2020 the optical equipment as well as dental crowns and bridges in the « 100% santé » package will be free of charge for the insured and in 2021 this will also be the case for all dental acts and hearing aids. However, by 2019, the limitations on the prices of the procedures and the modification of the bases of reimbursement will begin to take effect.

## The portfolio used in the study

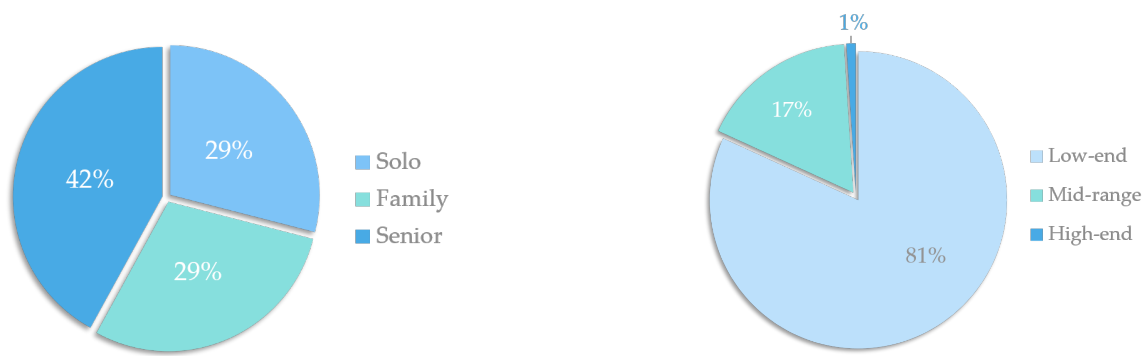
The portfolio studied is only linked to individual healthcare contracts, which makes it a fairly aging portfolio :



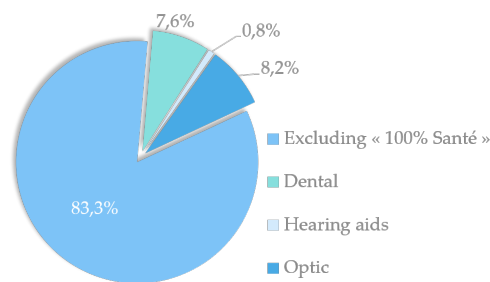
Distribution of the portfolio, in terms of exposure, by age of policyholders

In addition, the insurer's offer is divided into three product categories : Solo, Family and Senior, each developed in three levels of coverage, from the lowest to the highest. The Senior category represents 42% of the portfolio in terms of exposure, while the Solo and Family categories each represent 29%. Also, the vast majority of policyholders (81%) have subscribed to an low-end offer, which further exposes the portfolio to the consequences of the « 100% santé » reform.





Finally, the impact calculated in the tool only applies to reimbursements for the three healthcare categories concerned by the « 100% santé » reform. Thus, it should be put into perspective in relation to the total sum of reimbursements paid by the health insurer. When studying the portfolio it appears that the three items represent only 17% of reimbursements in 2018.



## Impacts of the « 100% santé » reform on the insurer's reimbursements

The calculation of the impacts of the « 100% santé » reform takes place in two steps. Firstly, they are calculated without taking into account the distortion in the consumption of healthcare by the insured, which should come with the introduction of the health package. Then, the said distortions are added in order to obtain a more realistic impact of the reform according to the chosen scenario. Indeed, the assumptions to be made for this second part of the calculation are strong, and it is not reasonable to consider only one set of assumptions, which is why several scenarios will be proposed.

Before projection of the amounts of payments made by the insurer, it is necessary to correctly format the database and add the required information. The initial database contains the payments made by the insurer in 2018, with the reference code associated with each reimbursed procedure and the amount of the reimbursement basis (or RB) imposed by the French social security system. Based on this database and thanks to the reference code of the acts, the new RBs modified by the reform (different over the years from 2019 to 2023) can be added to the database. Indeed, the government has published a table showing the correspondences between the old social security prices and the new ones. The referencing code has also made it possible to add the limit selling prices (or LSP), which are essential for the projection of reimbursements.

Once the database had been completed with this information, it was necessary to formulate a number of hypotheses in order to carry out the study.

- First of all, it is not uncommon for insurers to modify their tables of guarantees for other reasons than the adoption of a new reform (for example, for reasons of profitability or market adjustment). Since the study carried out in this paper aims to measure only the impact of the « 100% santé » reform, it will not take into account possible changes to the table of guarantees other than those induced by this reform.
- Furthermore, with the client’s agreement and in a need for prudence, an increase proportional to the RB is applied to the charges paid by the insured (called actual charges). This increase will apply within the limit of the LSPs imposed by the reform.
- Finally, it is necessary to define the use of LSPs. After a study of different ways, the hypothesis chosen is as follows : the projection of actual charges is equal to the minimum between the LSP and the actual charges in 2018 (themselves increased according to the previous hypothesis).

The insurer’s reimbursements can then be projected, while ensuring that, whatever the actual charges for the health basket, they will be reimbursed in full in 2020 for optics as well as dental crowns and bridges, and in 2021 for the rest of the health package. The results of this projection are as follows :

	Between 2018 & 2019	Between 2019 & 2020	Between 2020 & 2021	Between 2021& 2022	Between 2022 & 2023	Between 2018 & 2023
<b>DENTAL</b>	2,60%	9,47%	16,73%	0,49%	0,01%	31,77%
<b>AUDIOLOGY</b>	2,34%	0,61%	74,98%	0,00%	0,00%	80,16%
<b>OPTIC</b>	0,00%	-5,55%	0,00%	0,00%	0,00%	-5,55%
<b>OTHERS</b>	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
<b>TOTAL</b>	0,22%	0,29%	2,07%	0,05%	0,00%	2,64%

#### Impacts of the reform on insurers’ reimbursements

Therefore, the overall impact of the reform (taking into account unreformed categories such as hospitalisation or pharmacy) is an increase of 2.64%. A more detailed study is necessary because the impact differs greatly according to the category studied. Indeed, for optics, the total amount of payments made by the insurer drops by more than 5%, while for dentistry and audiology the impacts are positive (32% and 80% respectively). This great heterogeneity in the results is mainly due to differences in the way in which attempts are made to reduce the renunciation of healthcare according to category. Nevertheless, the impacts are also dependent on the database used to carry out this study as well as on the hypotheses chosen. Also, as the portfolio is ageing and the policyholders have mainly subscribed to low-end guarantee, it is normal to find relatively significant impacts in audiology and dental. The negative result for optics can be explained by the almost total withdrawal of social security from the equipment in the unlimited package.

Finally, the projections made for the insurer’s reimbursement amounts also make it possible to measure an impact on the remaining amount to be paid by policyholders, while remaining prudent in the analysis because the referencing of this variable has certain defects in the database and the assumptions made sometimes significantly reduce actual charges. On average, it appears that the amount payable by policyholders falls for all categories, even for optical products.

#### Impacts of the « 100% santé » reform on the behaviour of policyholders

Besides the purely technical impact of the « 100% santé » reform, it is very likely that the behaviour of the insured will change as it is introduced. This change in behaviour will necessarily have consequences on the amounts of payments made by insurers.

As with the « pur » impact of the reform, several assumptions need to be made in order to estimate the impact of this change in behaviour on the insurer's payments.

- First of all, the portfolio review is based on contracts active in 2018, so no new contracts will be considered.
- The deformation of the frequencies of consumption is an increase and is only applied on the acts of the package « 100% santé ». An exception is made for dental health-care because, although they are part of the unlimited package in the definition of the reform, they almost always come with the fitting of a dental prosthesis.
- Given that the process and content of the reform is relatively complex, not all insured persons will be aware of its functioning at the same time. This is why the distortions that are applied are weighted by the estimated rate of the population that is correctly informed about the reform « 100% santé ». This rate varies between 36% in 2020 and 100% in 2023.
- Three scenarios of distortion are studied, a central scenario, based on the rates of renunciation of healthcare observed in France, then broken down according to the type and level of guarantee, followed by a minor and major scenario with the aim of proposing a range of possible impacts to the insurer.

The impact on the insurer's payments after applying distortions in the frequency of consumption by policyholders is as follows :

	<i>Without deformation</i>	Minoring scenario	Central scenario	Majoring scenario
DENTAL	31,8%	60,2%	67%	90,5%
AUDIOLOGY	80,2%	158,8%	177,5%	214,5%
OPTIC	- 5,6%	-4,1%	-2%	2,8%
TOTAL	2,6%	5,6%	6,7%	8,9%

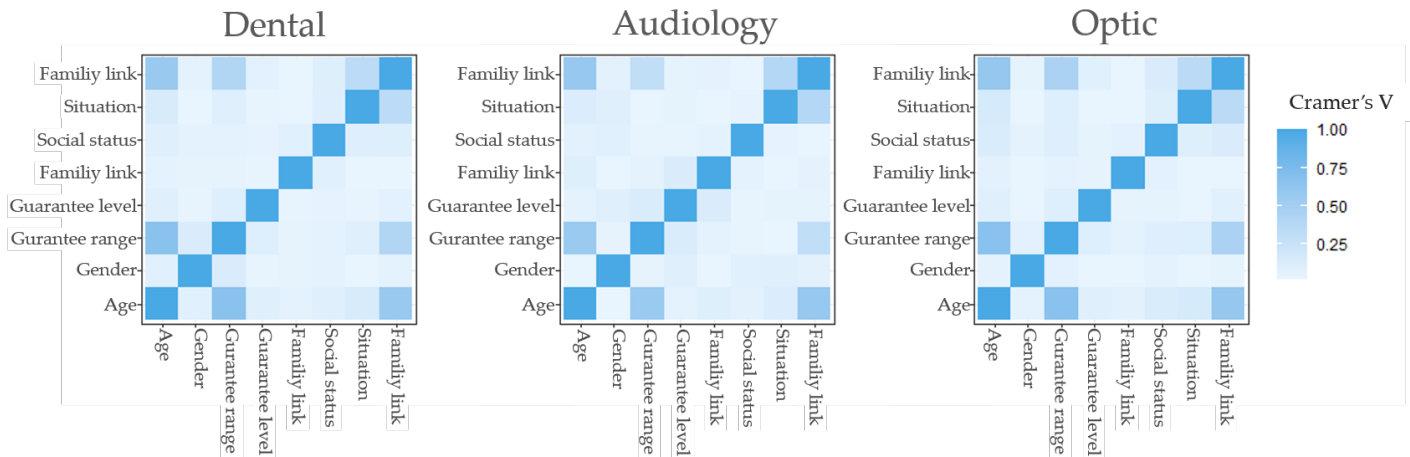
Impacts of the reform on the insurer's reimbursements, including the distortion of the frequency of consumption by policyholders

This shows that the overall impact of the reform, including the distortion of consumption, would be between 5.6% and 8.9%. Nota bene : It is important to remain cautious about this result, which is strongly linked to the database used as well as to the hypothesis of distortions chosen.

## Portfolio segmentation tool

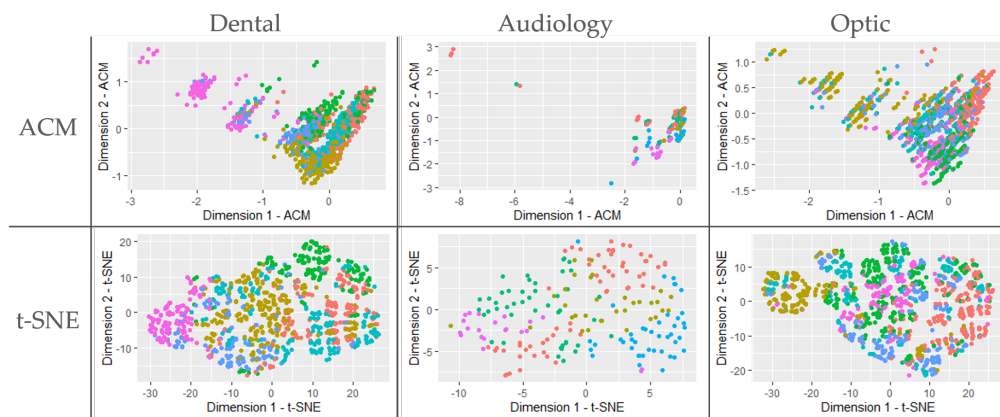
Measuring the impact of the « 100% santé » reform is a primary consideration in helping insurers to manage their business. To go further, it is possible to offer them a more precise segmentation of their portfolio, based on all the variables at their disposal. Indeed, this will enable them to better understand their portfolio, and therefore to better apprehend any change in the behaviour of their policyholders.

To do this, an first unsupervised study (i.e. not based on a target variable around which individuals would be grouped) is set up. The aim of this study is to verify the inherent links between variables, but also to try to observe a first unsupervised segmentation of the portfolio. Thus, a Cramer's test is carried out on the variables available from the insurer :



Cramer's test

This test shows that the majority of the variables are uncorrelated despite some exceptions such as age and type of guarantee or family relationship. Next, the PAM algorithm is implemented in order to observe the distribution of individuals and perhaps profiles that would be distinct. To visualise the results of the PAM algorithm, the MCA and the t-SNE algorithm are used :



Visualisations of the results of PAM algorithms

No profile emerges from this study, which shows that without a target variable, it is not possible to group people around their characteristics. Nevertheless, this study allows a better knowledge of the variables that will be used in the rest of this study, it is indeed necessary that the selected variables do not show too high a correlation.

It is for this reason that a supervised study is being conducted. Its purpose is to propose to the insurer a segmentation of its portfolio based on the amount of payments paid. This will enable the insurer to better identify the contracts which involve high reimbursements and those which, for the time being, involve lower reimbursements.

---

For this study, the CART tree segmentation capacity is used. Nevertheless, the lack of robustness of this method implies the need to validate the results by using other models. GLM, Random Forests and the XG Boost algorithm are good candidates. They make it possible to verify that the segmentation is coherent and, above all, that it can be used by the insurer as a management tool. To do so, two validation metrics are used (the coefficient of  $R^2$  and the Gini index). Therefore, if the CART model scores too badly for these two metrics compared to other models, it means that it cannot be considered relevant.

The amounts of payments vary greatly from one healthcare category to another, which is why three studies have been carried out. Finally, for each of the three categories studied, the segmentation obtained following the CART modelling is satisfactory because, for the most part, it confirms the assumptions made in the previous part of the study concerning the distortion of consumption frequencies. Nevertheless, in audiology, it suggests that the under-65s have renounced healthcare which had not been anticipated.

The policyholders most willing to change their consumption following the « 100% santé » reform are those who have subscribed to low-end coverage. There is also a gap between policyholders under 65 and the others : for audiology, policyholders under 65 seem to be more likely to renounce healthcare, whereas the opposite is true for optics (whatever the level of cover subscribed). The policyholders who are reluctant to change their consumption of healthcare are those who have subscribed to high-end guarantees, which is in line with the very purpose of the reform.

Moreover, although the segmentations obtained provide little additional information on the possible reasons for the renunciation of healthcare by policyholders, they have made it possible to verify the hypotheses put forward in the study of the deformation of policyholders' behaviour and to challenge some of them, particularly in audiology. They also provide the insurer with a precise knowledge of its portfolio given the available data. This last point is essential for enabling it to understand the impact of the reform as well as possible.

## Disclaimer

It often happens that a study depends on the quality of the data available, but this has not prevented the use of this work as a hypothesis testing tool and as an assistance to a better understanding of the portfolio. However, the protocol of segmentations building and validation introduced in this work is intended to be used in other studies. Therefore, it is possible that the latter could use it and draw better conclusions based on the quality of their data, as well as a segmentation that would provide much more information on the portfolio. Moreover, this work could even be used outside the framework of the « 100% santé » reform, just choose another variable around which the segmentation should be made. These Machine Learning methods, which are very little used in healthcare insurance, can also prove to be very interesting tools in terms of portfolio management.

# REMERCIEMENTS

Je souhaite tout d'abord remercier Daniel ATTIAS, mon tuteur de stage, pour m'avoir permis de rejoindre son équipe, pour sa confiance et le temps qu'il a su trouver afin de répondre à mes questions et de partager avec moi ses connaissances ainsi que son expérience. Un grand merci à Nabil RACHDI, responsable des recherches en data, pour m'avoir aidée à interpréter mes résultats et à les analyser correctement du point de vue métier.

Je remercie également tous les membres de l'équipe Actuarial Consulting pour leurs contributions à mon stage. Notamment Khedija ABDELMOULA CLAVERIE, responsable de l'équipe, pour m'avoir donné l'opportunité d'effectuer mon alternance au sein du cabinet, Aude SERVIER, pour le temps qu'elle a passé à m'expliquer ses missions et leurs différents aspects techniques, ainsi que Toufik DALI, Fabian GUILLEMIN et Cassandra BELOT pour leur patience et leur aide sur la réforme « 100% santé ».

J'aimerais aussi adresser mes sincères remerciements à toute l'équipe d'ADDACTIS France pour son accueil, sa convivialité, sa bienveillance et les partages d'expériences. Je tiens également à mentionner les collègues suivants pour leur soutien, leur bonne humeur et leur aide au moment de la rédaction de ce mémoire : Bryan GAUTIER, Floriane BUFFET et Yannick PERUS.

Enfin, je remercie les enseignants de l'Institut de Statistique de l'Université de Paris pour la qualité des enseignements dispensés, les connaissances qu'ils ont pu m'apporter et leur accompagnement tout au long de ma formation, et tout particulièrement mon tuteur académique Olivier LOPEZ.

# GLOSSAIRE

BR	Base de remboursement de la Sécurité Sociale
PLV	Prix limite de vente
HLF	Honoraire limite de facturation
RC	Remboursement complémentaire
FR	Frais réels
LPP	Liste des produits et prestations
CCAM	Classification commune des actes médicaux
CSBM	Consommation de soins et de biens médicaux
DCSi	Dépense courante de santé au sens international
MSA	Mutuelle Sociale Agricole
RSI	Régime Social des Indépendants
TNS	Travailleurs non salariés
DRESS	Direction de la Recherche, des Études, de l'Évaluation et des Statistiques
IRDES	Institut de Recherche et Documentation en Économie de la Santé
UNOCAM	Union Nationale des Organismes d'Assurance Maladie Complémentaire
CNSD	Confédération Nationale des Syndicats Dentaires
GIFO	Groupe des Industriels et Fabricants de l'Optique
ANI	Accord national interprofessionnel
NPA	Norme de pratique actuarielle
yc	y compris
Mise en « As-if »	Technique de projection des coûts : pour une prestation versée à l'année N, les paramètres permettant de calculer le coût de cette prestation (comme les BR, les PLV et les garanties) sont ajustés à leur valeur lors de l'année N+1 afin de recalculer la prestation « comme si » l'acte médical et le remboursement avaient été effectués à l'année N+1. Cette technique ne se restreint en réalité pas à l'année N+1, il est possible de faire le même calcul pour n'importe quelle année postérieure à N.

# TABLE DES MATIÈRES

<b>INTRODUCTION</b>	<b>1</b>
<b>I PRÉSENTATION DE LA RÉFORME ET DES DONNÉES D'ÉTUDE</b>	<b>3</b>
<b>1 LE SYSTÈME DE SANTÉ EN FRANCE</b>	<b>4</b>
1 La Sécurité Sociale . . . . .	5
2 Les complémentaires santé . . . . .	6
3 Les différents contrats et le remboursement des soins . . . . .	7
4 Les réformes de l'assurance santé . . . . .	8
<b>2 LA RÉFORME « 100% SANTÉ »</b>	<b>13</b>
1 Le renoncement aux soins . . . . .	13
2 Les principales mesures de la réforme . . . . .	16
3 Les étapes de la mise en place de la réforme . . . . .	20
4 Les conséquences pour les contrats des complémentaires santé . . . . .	21
<i>En résumé</i> . . . . .	22
<b>3 LA BASE DE DONNÉES</b>	<b>23</b>
1 Description de la base de données . . . . .	23
2 Création de variables . . . . .	24
3 Qualité des données . . . . .	26
4 Statistiques descriptives . . . . .	26
<i>En résumé</i> . . . . .	31
<b>II IMPACTS DE LA RÉFORME</b>	<b>32</b>
<b>1 TRAVAIL PRÉPARATOIRE SUR LA BASE DE DONNÉES</b>	<b>33</b>
1 Tables de Bases de Remboursement et de Prix Limites de Vente . . . . .	33
2 Agrégation et identification des prestations . . . . .	34
3 Création de la Table finale de prestations . . . . .	36
<b>2 IMPACTS SUR LES COÛTS DES TROIS POSTES CONCERNÉS</b>	<b>37</b>
1 Hypothèses . . . . .	37
2 Projections . . . . .	38
3 Résultats . . . . .	38
4 Conséquences pour l'assureur . . . . .	42
5 Conséquences pour l'assuré . . . . .	45
<i>En résumé</i> . . . . .	49
<b>3 IMPACT DE LA RÉFORME SUR LE COMPORTEMENT DES ASSURÉS</b>	<b>50</b>
1 Hypothèses . . . . .	50
2 Mise en place des taux de déformation de la fréquence de consommation . . . . .	52
3 Résultats . . . . .	52



4	Approche rentabilité assureur . . . . .	53
	<i>En résumé</i> . . . . .	56
<b>III</b>	<b>SEGMENTATION DU PORTEFEUILLE</b>	<b>57</b>
<b>1</b>	<b>ÉTUDE NON SUPERVISÉE</b>	<b>58</b>
1	Présentation théorique des modèles . . . . .	58
2	Variables sélectionnées pour l'étude . . . . .	61
3	Le test de Cramer . . . . .	62
4	L'algorithme PAM . . . . .	62
	<i>En résumé</i> . . . . .	65
<b>2</b>	<b>ÉTUDE SUPERVISÉE</b>	<b>66</b>
1	Présentation théorique des modèles . . . . .	66
	<i>En résumé</i> . . . . .	72
2	Variables sélectionnées pour l'étude . . . . .	73
3	Méthode de l'étude . . . . .	74
4	Segmentation et analyse pour le poste dentaire . . . . .	74
5	Segmentation et analyse pour le poste audiologie . . . . .	85
6	Segmentation et analyse pour le poste optique . . . . .	91
	<i>En résumé</i> . . . . .	100
	<b>CONCLUSION</b>	<b>101</b>
	<b>BIBLIOGRAPHIE</b>	<b>103</b>
	<b>ANNEXES</b>	<b>A1</b>
A	Listes A et B prothèses auditives . . . . .	A2
B	Description des variables composant la base de données de l'étude . . . . .	A3
C	Analyse des correspondances multiples . . . . .	A6
D	Algorithme t-SNE . . . . .	A8
E	Analyse complète pour le poste audiologie . . . . .	A10
F	Analyse complète pour le poste optique . . . . .	A17

---

# INTRODUCTION

En France, le système de santé tend à réduire au mieux les inégalités sociales. Cette tendance trouve ses racines dans l'Histoire. Pendant la Seconde Guerre mondiale, les résistants rêvaient d'un monde « d'après » plus juste, d'un projet politique pour la France qu'ils appelèrent « Les Jours heureux », dans le programme desquels figurait la création de la Sécurité Sociale. L'idée était de permettre à tous de vivre sans avoir peur de la maladie, de l'accident de travail, de la vieillesse... À la Libération, en 1945, le Conseil national de la Résistance mit en place un système global de protection sociale, la Sécurité Sociale, obligatoire et universelle, qui reconnaissait déjà le rôle complémentaire des mutuelles.

Depuis, les bases de ce système sont régulièrement réformées et modernisées, principalement dans le domaine des retraites et de l'aide sociale. En 2015, l'Assemblée nationale adopte le projet de loi santé instaurant la généralisation du tiers payant à tous les assurés. Toutefois, un nombre non négligeable de citoyens déclare encore renoncer à des soins médicaux faute de moyens [14]. Le gouvernement a donc mis en place en 2018 la réforme « 100% santé » visant à supprimer leur reste à charge pour de nombreux soins dans les postes audiologie, dentaire et optique. Pour cela, un effort commun de réduction des prix de vente et d'augmentation des remboursements est demandé aux praticiens, à l'Assurance Maladie et aux complémentaires santé.

L'impact pour ces dernières est difficile à mesurer, et cela pour trois raisons :

- la mise en place progressive de la réforme sur plusieurs années, qui implique une évolution des montants des prestations versées complexe à analyser et à anticiper ;
- les nombreuses spécificités des remboursements déclinées selon le poste de soins, qui impliquent une revue importante des tarifs de la Sécurité Sociale ainsi que des prix limites de vente et rendent l'anticipation de la sinistralité future délicate ;
- le changement probable dans le comportement des assurés et notamment dans la fréquence de leur consommation des actes de soins qu'il est très difficile d'appréhender.

La première raison comporte un enjeu stratégique pour les complémentaires santé, car elle leur impose une revue annuelle des conséquences de la réforme. Ces changements dans les prestations demandent également un suivi minutieux des montants de remboursements associés, afin d'être réactives en cas de dérive des coûts. Ce premier point demande donc un contrôle et un suivi réguliers, une fois la réforme appliquée.

La deuxième raison se situe en amont de l'application de la réforme. La revue des différentes bases de remboursement, potentiellement avec des périmètres d'actes différents, rend une mise en « As-if » des coûts complexe.

Néanmoins, cette mise en « As-if » ne tient pas compte d'une éventuelle déformation dans la fréquence de consommation des assurés. C'est ce qui constitue la troisième raison : si la réforme a bien les effets escomptés sur la réduction du renoncement aux soins, la consommation des assurés devrait être modifiée. Une bonne connaissance de son portefeuille et des profils de consommation de ses assurés est alors ici primordiale pour l'assureur afin d'évaluer au mieux les hausses de fréquences de consommation.

---

L'objectif de ce mémoire est de proposer à un client d'Addactis une étude ainsi qu'un outil d'aide au pilotage de cette réforme. L'outil présente plusieurs aspects. Tout d'abord, il offre une mesure de l'impact de la réforme sur les remboursements effectués par les complémentaires santé, en prenant en compte les différents moyens mis en œuvre par le gouvernement. Puis, il permet le développement d'une segmentation des assurés selon les remboursements dont ils ont bénéficié en 2018, afin de permettre à l'assureur de mieux anticiper et gérer les potentiels changements de comportement de ces assurés suite à l'application de la réforme, et donc les changements de fréquence des sinistres dans son portefeuille.

Pour ce faire, la première partie de ce mémoire est constituée d'une étude du système de santé français, de la nomenclature utilisée par l'Assurance Maladie, et plus particulièrement de la réforme « 100% santé ».

Puis une première simulation des coûts futurs sera réalisée sur la base des prestations passées. Des hypothèses de hausse du nombre d'actes seront ensuite posées afin de prendre également en compte la déformation des fréquences de consommation des assurés.

Enfin, des méthodes de machine learning seront utilisées dans le but d'obtenir une segmentation plus fine des assurés du portefeuille de l'assureur en fonction de leur profil de consommation. Cela permettra d'étudier les caractéristiques des assurés étant enclins à changer leur fréquence de soins médicaux.

En conclusion, à l'heure où la crise du coronavirus fait voler en éclats les systèmes de santé internationaux, il s'agit de mettre tout en œuvre pour permettre à la spécificité française du « 100% santé » de préserver l'équilibre entre la suppression du reste à charge pour les uns, et l'effort de réduction des prix de vente et d'augmentation des remboursements pour les autres, afin que le concept historique des « Jours heureux » d'après-guerre ne se transforme pas en vaine utopie. Ce mémoire a ainsi pour vocation de constituer une aide au pilotage pour les assureurs, afin qu'ils mesurent le réel impact de la réforme, et qu'ils puissent, avec les outils les plus adaptés, ajuster leur politique tarifaire.

## PREMIÈRE PARTIE

### PRÉSENTATION DE LA RÉFORME ET DES DONNÉES D'ÉTUDE

# 1 | LE SYSTÈME DE SANTÉ EN FRANCE

## Les remboursements de soins

Chacune des notions apparaissant dans ce premier paragraphe seront par la suite développées.

Quelque que soit le type de contrat (individuel ou collectif), le fonctionnement du remboursement des frais de santé reste le même et s'effectue en plusieurs étapes. Tout d'abord, la Sécurité Sociale répertorie chaque acte médical et lui associe une base de remboursement (BR). Elle prend ensuite en charge un certain pourcentage de cette BR, appelé Taux de remboursement (différent en fonction des actes médicaux), laissant ainsi à la charge de l'individu un montant appelé Ticket modérateur. Par ailleurs, de nombreux soins font souvent l'objet de dépassements d'honoraires (c'est à dire que le prix de l'acte est supérieur à la base de remboursement fixée), qui sont également à la charge de l'individu.

C'est à ce titre que la complémentaire santé intervient : elle peut prendre en charge tout ou partie du ticket modérateur ainsi que des dépassements d'honoraires. Les remboursements effectués par les organismes complémentaires sont décrits dans le contrat d'assurance et peuvent se traduire de différentes façons (en euros ou en pourcentage de la BR ou des frais réels par exemple). Il est important de noter que puisque ces remboursements peuvent dépendre de celui effectué par la Sécurité Sociale, la politique sociale mise en oeuvre par le gouvernement joue un rôle important dans la mise en place et le calcul des tarifs des contrats de santé par les organismes complémentaires.

Il reste donc à la charge de l'assuré une éventuelle partie des dépassements d'honoraires, ainsi que la participation forfaitaire de 1€ prélevée sur le remboursement de la Sécurité Sociale, qui vise à réduire l'aléa moral et à responsabiliser les individus dans leurs démarches de santé.

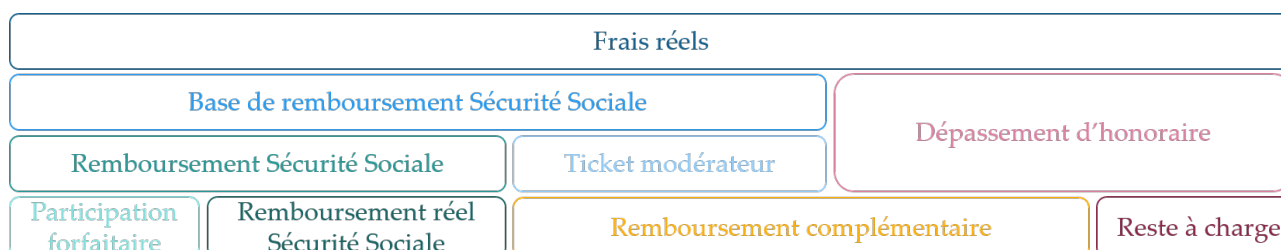


FIGURE 1.1 – Illustration du système de remboursement en France

Honoraires praticiens	Honoraires auxiliaires médicaux	Analyses et examens de laboratoire	Médicaments	Optique
70%	60%	60% à 70%	15% à 100%	60%
Dentaire	Prothèses auditives	Orthopédie	Hospitalisation	Frais de transport
70%	60%	60%	65% à 80%	65%

FIGURE 1.2 – Taux de remboursement de la Sécurité Sociale selon l'acte médical en 2019

---

# 1 La Sécurité Sociale

Dès 1940, un haut fonctionnaire français, Pierre Laroque a mis en œuvre un plan qui a conduit à la création de la Sécurité Sociale. Il cherchait à mettre en place « un plan complet de Sécurité Sociale, visant à assurer à tous les citoyens des moyens d'existence, dans tous les cas où ils sont incapables de se le procurer par le travail, avec gestion appartenant aux représentants des intéressés et de l'État » [23]. Porté par de nombreux résistants, dont notamment le député Ambroise Croizat, le projet d'une Sécurité Sociale voit finalement le jour le 4 octobre 1945 dans une ordonnance qui décrit le rôle et les missions de l'institution, ainsi que différents régimes qui en viennent à coexister.

À l'heure actuelle, quatre types de régimes composent les caisses de la Sécurité Sociale et assurent l'assistance financière de différents risques répertoriés dans cinq branches indépendantes : la maladie, la famille, la retraite, les cotisations et les recouvrements, les accidents du travail et les maladies professionnelles.

- Le régime général regroupe plus de 91% de la population française [26] avec les bénéficiaires de la Protection Universelle Maladie ainsi que les travailleurs salariés du secteur privé.
- La mutuelle sociale agricole (MSA) concerne les exploitants et les salariés agricoles, elle compte 1,2 million de cotisants [26].
- Le régime social des indépendants (RSI) gère la branche maladie des artisans, commerçants et professionnels libéraux, et comptabilise 2,8 millions de cotisations [26].
- Les régimes spéciaux proposent soit des prestations liées à des risques spécifiques pour des métiers particuliers tels que ceux de la fonction publique (militaires, fonctionnaires civils), de la SNCF, ou encore ceux de l'Église par exemple ; soit des régimes particuliers tels que celui des étudiants ou encore celui de l'Alsace Moselle. Le régime Alsace Moselle concerne les habitants de cette région, il date de l'occupation allemande et est très avantageux (par exemple, les médicaments sont mieux remboursés et le forfait journalier hospitalier n'est pas appliqué). Ces avantages s'accompagnent cependant de cotisations plus élevées qu'ailleurs.

Ces différents régimes sont financés par les cotisations sociales payées par les employeurs et les salariés, par la Contribution Sociale Généralisée, ainsi que par des contributions et taxes (comme par exemple la Taxe de Solidarité Additionnelle).

Bien que la Sécurité Sociale prenne en charge une grande partie des frais médicaux, il arrive souvent que les frais soient trop importants pour être totalement remboursés. Une Base de Remboursement de la Sécurité Sociale (BR) est donc mise en place pour chaque acte de soin, en fonction du régime de l'individu. La Sécurité Sociale prend en charge un certain pourcentage de cette BR. Ainsi, les dépenses engagées sont appelées Frais réels et ne sont souvent pas entièrement prises en charge par la Sécurité Sociale. Il est donc possible de compléter les remboursements par la souscription d'un contrat d'assurance santé auprès des organismes complémentaires comme l'explique le paragraphe suivant.

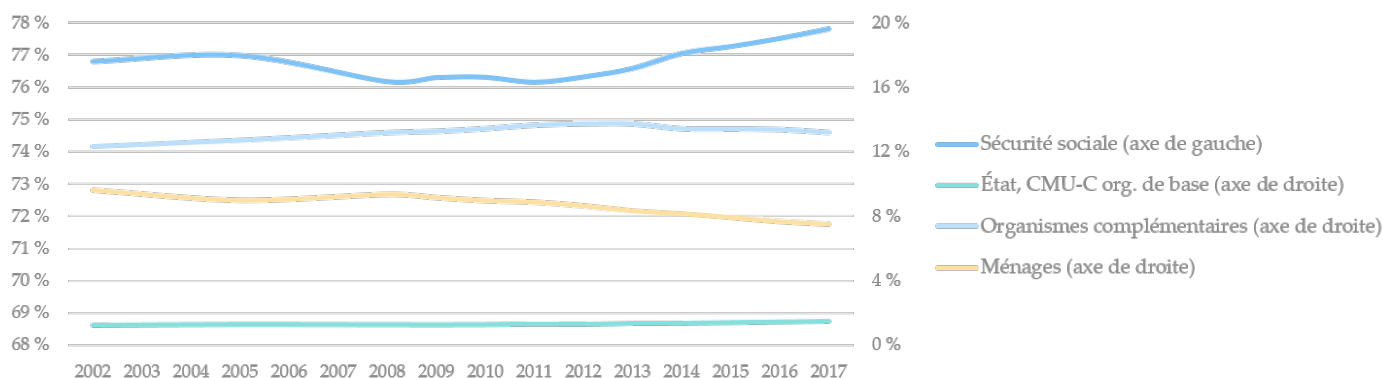


FIGURE 1.3 – Part des financeurs de la consommation de soins et de biens médicaux entre 2002 et 2017, en % de la dépense totale [11]

## 2 Les complémentaires santé

Les complémentaires santé permettent à leurs assurés de se décharger d’une part ou de la totalité du montant qu’il leur reste à payer une fois que la Sécurité Sociale a effectué son remboursement. En contrepartie, chaque assuré verse une cotisation à la complémentaire santé.

En 2014, les organismes complémentaires ont contribué à hauteur de 13,5% au financement de la consommation de soins et de biens médicaux en versant 27,1 milliards d’euros de prestations [29]; en 2016, plus de 95% de la population française [8] était couverte par une complémentaire santé. Par ailleurs, l’Accord National Interprofessionnel (ou ANI, qui sera développé par la suite à la section 4 de ce chapitre) impose à toutes les entreprises de proposer une couverture complémentaire de santé collective à l’ensemble de leurs employés peu importe leur ancienneté, ce qui place d’autant plus les organismes complémentaires au devant de la scène.

Ce secteur d’activité joue donc un rôle de plus en plus important dans l’économie française. Trois grands types d’organismes se partagent le marché de la complémentaire santé : les mutuelles, les compagnies d’assurances et les instituts de prévoyance. La finalité de leur activité est la même : couvrir leurs clients contre les risques liés à la santé. Cependant chaque organisme est affilié à un code juridique différent et possède donc ses propres moyens de fonctionnement.

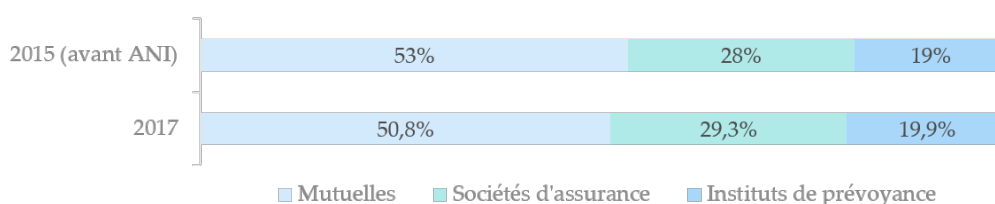


FIGURE 1.4 – Parts de marché de l’assurance santé en 2015 & 2017 [11] et [9]

### Les mutuelles

Il s’agit de sociétés de personnes à but non lucratif basées sur un principe de solidarité, c’est-à-dire que les cotisations de chaque adhérent sont indépendantes de leur risque indivi-

---

duel. Ces cotisations représentent la plus grande partie des revenus de ces sociétés. De plus, leur fonctionnement est égalitaire puisque chaque sociétaire dispose d'une voix lors des Assemblées Générales. Enfin, leurs activités sont réglementées par le Code de la Mutualité. En 2017, les mutuelles restent les acteurs majeurs dans le domaine de la complémentaire santé en France avec 51% du marché [?], représentant 18,3 milliards d'euros de cotisations.

## **Les compagnies d'assurances**

Il s'agit le plus souvent de sociétés anonymes à but lucratif, c'est-à-dire que les bénéfices qu'elles engrangent sont par la suite reversés à leurs actionnaires. Il existe également des sociétés d'assurance mutuelles qui sont à but non lucratif et qui se rapprochent donc du fonctionnement des mutuelles. Cependant, quel que soit le type de société, une compagnie d'assurance est régie par le Code des Assurances. Les compagnies d'assurances prennent de plus en plus de place sur le marché de la santé. En 2017, elles détenaient 31% des parts du marché avec 11 milliards d'euros de cotisations (contre 19% en 2001) [?].

## **Les instituts de prévoyance**

Comme les mutuelles, les instituts de prévoyance sont des sociétés de personnes et de droit privé, à but non lucratif. Cependant ils sont réglementés par le Code de la Sécurité Sociale. Une autre différence avec les mutuelles est que les instituts de prévoyance ont une gestion paritaire. C'est-à-dire que ce sont les partenaires sociaux, comme les syndicats professionnels par exemple, qui gèrent leur fonctionnement. Les institutions de prévoyance ne traitent presque que des contrats collectifs d'assurance de personnes, ce qui réduit la part de marché globale à laquelle elles ont accès. De plus, de la même manière que les mutuelles, elles perdent des parts de marché au profit des compagnies d'assurances. Malgré cela, elles centralisaient encore 18% des parts du marché [?] en 2017, soit environ 6,4 milliards d'euros de cotisations.

# **3 Les différents contrats et le remboursement des soins**

Les organismes complémentaires qui viennent d'être présentés proposent deux types de contrats : les contrats individuels ou les contrats collectifs.

## **Les contrats individuels**

Ils ne comportent que deux signataires : l'assuré et l'assureur. Il est cependant possible de faire profiter des garanties du contrat à d'autres personnes, appelées les ayants-droit. Ces derniers peuvent faire partie du cercle familial de l'assuré (comme le conjoint ou les enfants) ou pas (comme par exemple le second conducteur d'un véhicule). Une même personne peut disposer de plusieurs contrats individuels chez différents organismes d'assurance complémentaire.

Généralement, les tarifs des contrats individuels sont très ciblés : l'assuré doit répondre à un questionnaire qui déterminera par la suite le montant de ses cotisations. La zone géographique ainsi que l'âge, la catégorie socio-professionnelle et de nombreux autres paramètres sont pris en compte en fonction des garanties souhaitées, cela permet que l'assuré paie en fonction du seul risque qu'il représente.



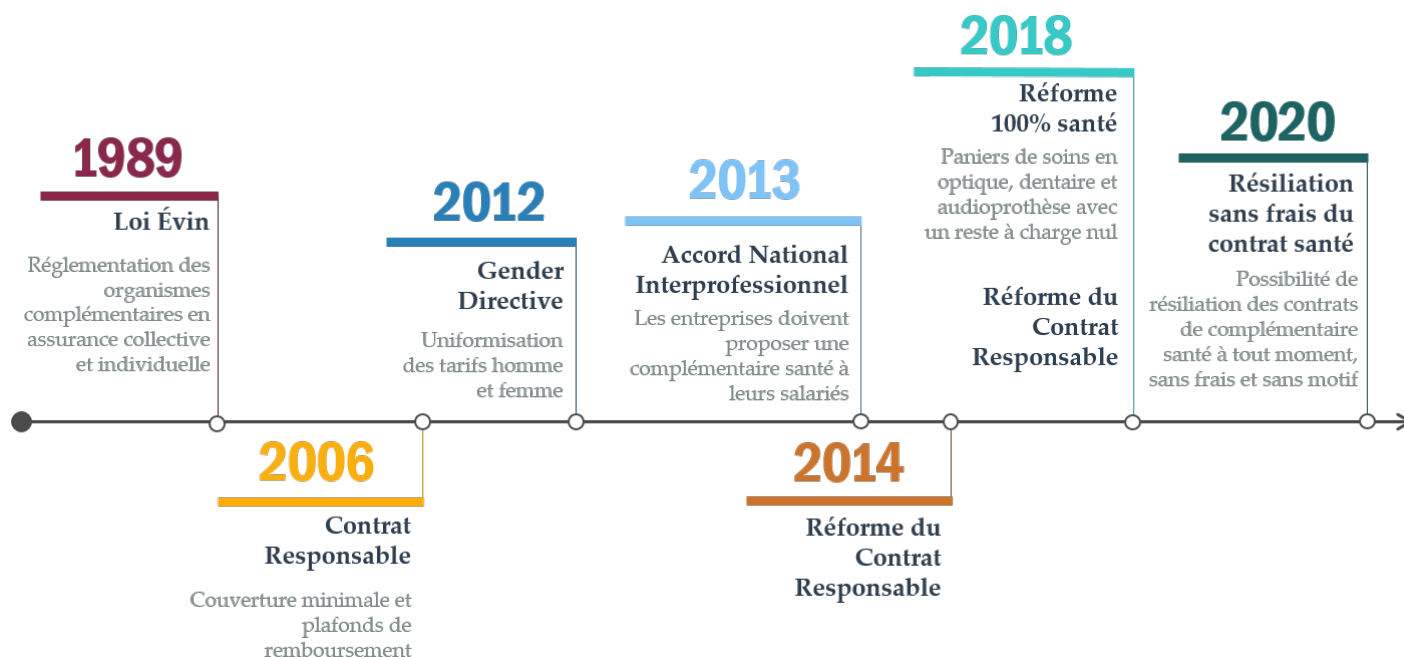
Concernant les contrats de couverture santé, l'assureur ne peut pas utiliser le seul motif de l'état de santé de l'assuré pour l'exclure, réduire ses garanties ou encore majorer le tarif de son contrat. Ces contrats permettent aux individus qui le souhaitent d'augmenter les garanties de remboursement de leurs soins, surtout pour les postes qui ne sont que très peu remboursés par la Sécurité Sociale comme l'optique ou le dentaire par exemple.

## Les contrats collectifs

Ils sont passés entre un groupe de personnes et l'assureur. Généralement, ce groupe est une entreprise ou une association. Un contrat d'assurance collective compte autant de signataires que de personnes du groupe qui y adhèrent. Les garanties peuvent malgré tout être étendues aux membres de la famille et aux ayants-droit. Elles concernent la protection sociale complémentaire des adhérents, c'est-à-dire leur retraite complémentaire, l'épargne salariale et la couverture en cas de maladie, d'incapacité, d'invalidité ou de décès.

Le montant des cotisations est uniformisé pour tous les adhérents, c'est-à-dire que les paramètres pris en compte dans les contrats individuels ne le sont plus ici (il peut tout de même y avoir une différence entre les catégories socio-professionnelles au sein d'une entreprise). Des tarifs différents s'appliquent éventuellement en fonction du nombre d'ayants-droits. Le tarif est fixé en fonction des caractéristiques de l'entreprise : le nombre de salariés, l'âge moyen des adhérents, éventuellement la répartition homme/femme, etc. Dans le cadre de la complémentaire santé, l'état de santé d'un adhérent ne peut faire augmenter sa cotisation. À la suite de la réforme de l'ANI (cf. section suivante), l'adhésion du salarié au contrat proposé par l'entreprise est obligatoire, bien qu'il existe quelques cas de dispense.

## 4 Les réformes de l'assurance santé



### 1989 - Loi Évin

La loi Évin du 31 décembre 1989 établit des règles que les organismes complémentaires doivent respecter dans le but de protéger les intérêts des assurés. Elle porte plus précisément

---

sur « la prévention et la couverture du risque décès et des risques portant atteinte à l'intégrité physique de la personne ou liés à la maternité ou des risques d'incapacité de travail ou d'invalidité ou du risque chômage » [16]. Cette loi traite en majorité des réglementations liées à la santé collective, cependant il y a aussi quelques dispositions particulières sur la santé individuelle.

Ce texte étant considéré comme le fondement de la complémentaire santé, il a été rédigé pour chacun des différents codes qui régissent l'assurance santé. Il pose un cadre à la couverture santé dans sa globalité, c'est pourquoi il serait trop long ici de le présenter de manière exhaustive : seuls quelques points clés seront donc mis en avant.

Après avoir défini le cadre précis de la prévoyance collective, il impose tout d'abord aux assureurs la prise en charge des conséquences d'une pathologie s'étant déclarée avant la souscription du contrat (ou de l'adhésion dans le cadre d'un contrat collectif). De plus, un salarié ne peut pas être discriminé lors de l'adhésion à un contrat collectif, c'est-à-dire que l'assureur ne peut pas proposer des contrats différents en fonction de l'état de santé de l'adhérent ou de son mode de vie par exemple. Dans la même optique, un assuré individuel ne peut voir le tarif de son contrat augmenter significativement à la suite d'une dégradation de santé. Une hausse de tarif est acceptée mais elle doit être la même pour tous les assurés qui ont souscrit à la même garantie.

Par ailleurs, le maintien de la couverture en frais médicaux par l'assureur est obligatoire dans certains cas comme l'incapacité, l'invalidité, le départ à la retraite ou encore le maintien de la couverture pour les ayants-droits des salariés décédés. Concernant le départ à la retraite, l'assureur a l'obligation de proposer à l'adhérent les mêmes garanties pour un tarif n'excédant pas 150% du tarif d'un salarié actif.

## **2006 - Contrat Responsable**

Le Contrat Responsable est apparu pour la première fois en 2006 afin de responsabiliser les individus sur leurs dépenses de santé. Il a ensuite été réformé en 2014 et plus récemment en 2018 pour les postes de soins impactés par la réforme « 100% santé » qui sera étudiée de manière plus approfondie par la suite. Pour qu'un contrat d'assurance santé soit considéré comme responsable, il doit respecter un cahier des charges établi dans la réglementation des assurances qui prévoit une couverture minimale de l'assuré ainsi que des plafonnements pour certains actes.

### **Couverture minimale pour l'assuré**

- Le ticket modérateur doit être intégralement remboursé par l'organisme complémentaire (hormis pour des cures thermales, de l'homéopathie et des médicaments remboursés à 15 et 35%).
- Le forfait journalier hospitalier doit également être pris en charge sans limitation de durée s'il s'agit d'établissements de santé, ce qui exclut les EHPAD par exemple.
- Les médecins ayant adhéré à un dispositif de pratiques tarifaires maîtrisées n'ont pas de limitation de dépassements d'honoraires tant qu'ils restent dans la limite définie par les garanties du contrat de l'assuré.
- Tous les deux ans, l'assuré a le droit à deux verres et une monture pris en charge de manière forfaitaire.

- 
- Suite à la réforme « 100% santé », l'organisme complémentaire se doit de prendre intégralement en charge les paniers sans reste à charge qu'elle définit.

### **Plafonnements des remboursements**

- Le remboursement d'équipements optiques fait l'objet de plafonds spécifiques. Par exemple, pour les montures, ce plafond était fixé à 150 € jusqu'à la réforme « 100% santé » et à 100 € à présent. Les verres font aussi l'objet de plafonnements en fonction de la dioptrie et de leur type.
- La participation forfaitaire de 1 € ne peut pas être prise en charge par l'organisme complémentaire.
- En cas de non respect du parcours de soin (qui pousse les assurés à aller consulter en premier lieu leur médecin traitant), les dépassements d'honoraires et la majoration du ticket modérateur ne peuvent plus être pris en charge.
- Le remboursement des dépassements d'honoraires de médecins n'ayant pas adhéré à un dispositif de pratiques tarifaires maîtrisées est plafonné à 100% de la BR.

Afin d'encourager le développement des contrats collectifs responsables, la réglementation prévoit également des avantages sociaux et fiscaux. En effet, dans la limite d'un certain plafond, l'employeur est exonéré des charges sociales ; la taxe de solidarité additionnelle (ou TSA) voit son taux passer de 20,27% à 13,27% ; et la part salariale de l'assuré est déduite du calcul de l'impôt sur le revenu. Cette démarche s'est avérée payante : près de 95% [27] des contrats proposés par les organismes complémentaires sont aujourd'hui responsables.

### **2012 - Gender Directive**

Le Conseil de l'Union Européenne avait instauré le principe de l'égalité des traitements en 2004. Cela assurait aux individus un accès aux biens ou services sans discrimination de genre. Cependant, jusqu'en 2012, les compagnies d'assurances n'avaient pas l'obligation de respecter ce principe dans la mesure où les produits qu'elles proposaient reposaient sur des risques qui pouvaient, d'un point de vue statistique, réellement différer selon le genre de l'assuré.

C'est en décembre 2012 que le Conseil de l'Union Européenne met fin à cette dérogation en obligeant les assureurs à uniformiser leurs prix entre les femmes et les hommes. Cette décision a donc fortement impacté la tarification, sans pour autant être nécessairement avantageuse pour les assurés selon le type d'assurance. En effet, les femmes ont vu leurs cotisations d'assurance auto augmenter alors que celles des hommes ont baissé ; à l'inverse, les prestations d'épargne et d'assurance vie ont augmenté pour les hommes et diminué pour les femmes. Bien qu'elle repose sur un principe noble, cette directive est assez controversée. En effet, le principe même de l'assurance est d'effectuer une segmentation de la population par groupes homogènes, et le genre est un facteur déterminant dans de nombreuses tarifications actuarielles. C'est pourquoi l'utilisation du genre comme facteur d'évaluation des risques reste autorisée tant qu'elle ne conduit pas à une différenciation des primes entre les assurés.

Cependant, en assurance santé, très peu de contrats utilisaient ce critère dans leur tarification. Cette directive n'a donc pas eu un impact trop important dans ce domaine. Les contrats

---

collectifs n'ont pas été impactés puisqu'ils ne distinguaient pas les genres avant cette directive. Ce sont finalement les domaines de l'assurance auto et de l'assurance vie qui ont été principalement impactés.

## **2013 - Accord National Interprofessionnel**

Cet accord, souvent appelé ANI, a été signé le 11 janvier 2013. Il porte sur les conditions de travail et les garanties sociales des salariés. Une grande partie de cet accord établit des règles sur la généralisation de la complémentaire santé collective dans le secteur privé. Cet accord devait être appliqué à compter du 1<sup>er</sup> janvier 2016. Ce laps de temps est dû au fait que l'accord implique une création de contrats personnalisés ou une mise en conformité de contrats déjà existants pour chaque entreprise. En effet, suite à cet accord, toutes les entreprises doivent proposer à leurs salariés une complémentaire santé incluant un minimum de garanties explicitées dans l'accord. Cette complémentaire est alors financée au minimum à 50% par l'employeur. Par ailleurs, l'accord prévoit aussi la conservation des garanties de santé et de prévoyance en cas de chômage, s'il n'est pas la conséquence d'un licenciement pour faute lourde.

La mise en place de cet accord a eu pour conséquence un transfert des contrats individuels vers des contrats collectifs. Ainsi, une déformation du portefeuille individuel des organismes complémentaires a pu être observée : la proportion de travailleurs non salariés et de retraités a augmenté. Ces derniers ayant des consommations de santé plus importantes en moyenne, leur prime a donc également augmenté.

## **2018 - Réforme « 100% santé »**

La réforme « 100% santé » a été votée par le Sénat le 4 décembre 2018. Elle impose aux assureurs et professionnels de santé de proposer aux individus des actes médicaux sans reste à charge pour les postes optique, dentaire et audio-prothèse. Cette réforme étant l'objet de l'étude, le contexte ainsi que les différentes mesures mises en œuvre seront détaillés par la suite.

Dans les grandes lignes, cette réforme permettra à tout individu possédant un contrat de complémentaire santé d'accéder à différents paniers de soins en fonction du type de produit qu'il souhaite. Un panier « 100% santé » assure des équipements de qualité sans reste à charge, l'assuré devra choisir parmi un certain nombre de produits prédéfinis dans la réforme. Pour les soins dentaires, il y a un panier reste à charge maîtrisé où les garanties proposées sont légèrement supérieures à celles du premier panier. L'assuré aura donc à payer une partie des soins, ou prothèses, mais avec un coût qui restera limité. Enfin, dans les trois postes concernés, il y a également un panier libre, où l'assuré pourra choisir les équipements et soins de son choix en assumant une plus grande partie des coûts qui leurs sont liés.

Du point de vue du prix des actes et du remboursement, cette réforme impacte à la fois la Sécurité Sociale, les organismes complémentaires et les professionnels de santé. Les bases de remboursement de la Sécurité Sociale vont en moyenne augmenter pour les paniers « 100% santé » et diminuer pour les paniers libres. Des Prix Limites de Vente (ou PLV) ainsi que des Honoraires Limites de Facturation (ou HLF) sont également mis en place, ce qui plafonne le montant de certains remboursements complémentaires et celui des honoraires de certains soins. La consommation des individus qui renonçaient aux soins pour des raisons financières

---

devrait augmenter, ce qui augmenterait dans le même temps les remboursements complémentaires pour les paniers « 100% santé ». Enfin, les professionnels de santé seront obligés de proposer deux devis, l'un se basant sur les garanties « 100% santé », l'autre non. Ils auront également l'obligation de proposer un certain panel de choix à leurs clients (un certain nombre de montures entrant dans les garanties du panier « 100% santé » par exemple).

## **2020 - Droit de résiliation sans frais du contrat de complémentaire santé**

Le 27 mars 2019, la proposition de loi permettant à un assuré de résilier un contrat de complémentaire santé à tout moment et sans frais a été adoptée. Elle sera appliquée à partir du 1<sup>er</sup> décembre 2020. Plus précisément, un assuré pourra résilier son contrat santé (individuel ou collectif) quand il veut après un an de souscription. Cela laisse plus de liberté aux assurés. En effet, jusqu'à cette proposition, l'assuré recevait un avis d'échéance une fois par an qui précisait l'unique date possible de résiliation du contrat. Cet échéancier pouvait être envoyé au plus tard 15 jours avant la date limite de résiliation, ne laissant à l'assuré que très peu de temps pour résilier son contrat.

Bien que du côté des complémentaires santé, une augmentation des frais de gestion (et donc des cotisations) soit envisagée, d'après le gouvernement cette loi pourrait augmenter considérablement la concurrence sur le marché de la complémentaire santé et donc diminuer le montant des cotisations. C'est en tout cas le constat qui a pu être fait dans le cadre des assurances auto et habitation puisque suite à la loi Hamon, qui permet aux assurés de résilier leurs contrats auto et habitation à tout moment, c'est bien une baisse des cotisations qui a été observée. Par ailleurs, un des principaux amendements du texte oblige les organismes complémentaires à indiquer le taux de redistribution de chaque contrat, c'est-à-dire la part moyenne payée par l'assuré qui lui revient sous forme de remboursements complémentaires. Cette mesure permettra aux assurés de mieux comprendre à quoi correspond leur prime de santé.

## 2 | LA RÉFORME « 100% SANTÉ »

Le renoncement aux soins en France reste toujours présent, c'est pourquoi la volonté d'y faire face en instaurant une réforme du système de santé était annoncée dans le programme d'Emmanuel Macron lors des élections présidentielles de 2017. En effet, le gouvernement annonce que 65% des personnes ayant besoin d'un appareil auditif y renoncent [19], et l'Assurance Maladie [1] a publié un dossier exposant que 19% des français renonçaient à des achats d'optique, 39% à des prothèses dentaires et 34% à des soins dentaires.

### 1 Le renoncement aux soins

Le système de santé français a pour but de faciliter l'accès aux soins aux citoyens quels que soient leur niveau de vie, leurs revenus et leur lieu de résidence. Parmi les pays d'Europe, la France fait partie de ceux présentant un faible taux de renoncement aux soins. Malgré tout, une partie non négligeable de la population (entre 1% et 3% selon les études) continue à déclarer qu'elle renonce à des soins médicaux pour différentes raisons. Dans cette optique, le gouvernement Macron a mis en place la réforme « 100% santé ». Afin de trouver des solutions pour réduire le renoncement aux soins, il est tout d'abord nécessaire d'en comprendre les différentes causes.

#### La première cause de renoncement reste la raison financière.

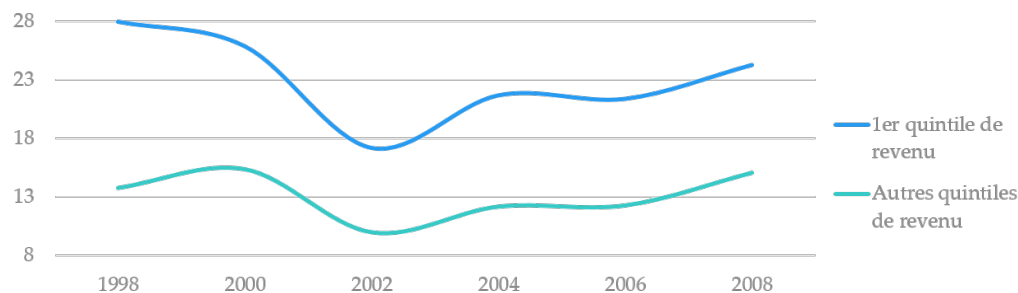


FIGURE 2.1 – Évolution du taux de renoncement aux soins selon le niveau de revenu [14]

Tout d'abord, le renoncement aux soins est fortement corrélé à l'absence de couverture complémentaire, elle-même significativement liée à la catégorie socio-professionnelle. Bien que l'ANI ait permis à de nombreux salariés d'être couverts par une complémentaire santé en obligeant leur employeur à payer 50% de la cotisation, de nombreuses personnes non salariées n'ont toujours pas les moyens de contracter une assurance santé individuelle.

D'autre part, il arrive que le remboursement complémentaire ne suffise pas et que certains assurés ne puissent pas faire face au paiement de certains actes pour lesquels le reste à charge est élevé. En 2018 en France, le reste à charge moyen était de 7,5% de la Consommation de Soins et de Bien Médicaux (CSBM) [11] ou encore 6,8% de la Dépense Courante de Santé au sens international (DCSi) [19], il est l'un des plus faibles des pays de l'OCDE.

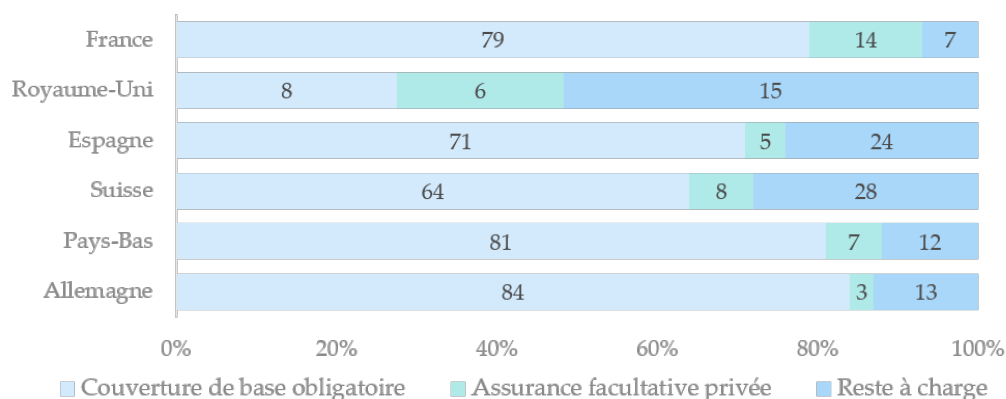


FIGURE 2.2 – Répartition de la dépense courante de santé entre financeurs en 2015 dans six pays d'Europe, en % de la DCSi [21]

Cependant, le reste à charge de l'individu a tendance à fortement augmenter pour certains postes de soins. En 2017, l'optique était le poste où les individus avaient le reste à charge le plus important avec 23% [10] des frais réels; le dentaire suivant de près avec 22,2% [1] des frais réels. Le poste des audio-prothèses est plus difficile à évaluer car il est souvent inclus dans la catégorie « autres biens médicaux ». En raisonnant en terme de coût moyen, le constat est le suivant : un appareil auditif coûte environ 1500 € et ensemble, les remboursements effectués par la Sécurité Sociale et la complémentaire sont d'environ 600 € en tout, le reste à charge pour l'assuré est donc aux alentours de 900 € par appareil, montant qui peut légèrement varier en fonction des garanties du contrat. Cela reste extrêmement élevé pour de nombreuses personnes, d'autant plus que les individus ayant besoin d'appareils auditifs sont principalement des personnes âgées et retraitées dont les revenus sont bien souvent fortement réduits.

Des mesures étaient déjà mises en place pour réduire au maximum le reste à charge. Par exemple, les assurés faisant appel à un réseau de soins pour le dentaire ou l'optique voient leur reste à charge diminuer de manière significative par rapport à ceux qui ne se rendent pas dans un réseau de soins partenaire de leur assureur. Cependant, au vu des chiffres exposés ci-dessus, ces mesures semblent insuffisantes. C'est pourquoi le gouvernement a voté une réforme qui propose un reste à charge nul sur les trois postes précédemment cités : l'optique, le dentaire ainsi que les audio-prothèses.

## D'autres causes de renoncement aux soins

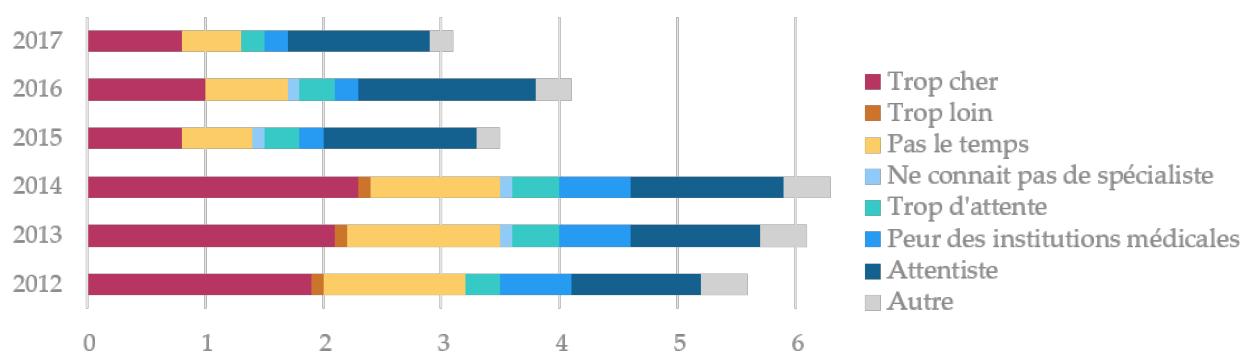


FIGURE 2.3 – Motifs du renoncement aux soins en France entre 2012 et 2017, en % de la population  
Eurostat > Data Explorer

Ce graphique met en évidence de nombreuses raisons de renoncement aux soins pour lesquelles le gouvernement n'est pas en mesure d'agir. En effet, il ne peut forcer les individus à aller consulter un médecin s'ils considèrent qu'ils n'ont pas le temps ou s'ils attendent de voir si leur état va s'améliorer de lui-même. Bien que pour le dernier point, la prévention permet de responsabiliser les individus. Par ailleurs, il apparaît que depuis 2015 la peur des institutions médicales a diminué. Ce motif de renoncement peut éventuellement être réduit par le gouvernement grâce à sa communication sur le domaine médical. Enfin, dans la catégorie « Autre » se trouvent par exemple les personnes qui n'ont pas confiance en la médecine et se méfient des spécialistes. Cette catégorie ne représentait que 0,2% de la population en 2017 mais la communication pourrait peut-être contribuer à sa réduction.

En outre, le temps d'attente est également un motif de renoncement aux soins. Pour le poste optique, afin de réduire ce type de renoncement, les opticiens vont être habilités à adapter la correction d'un patient ainsi qu'à renouveler une ordonnance effectuée par un ophtalmologiste pour les patients âgés de plus de 16 ans si ce renouvellement est associé à la prise en charge d'un équipement optique. De plus, la loi sur le financement de la Sécurité Sociale de 2019 prévoit des passerelles pour les opticiens qui souhaiteraient suivre une formation d'orthoptiste. Ces mesures devraient permettre de réduire le temps d'attente pour renouveler une paire de lunettes par exemple et ainsi, lorsque cela en était la cause, réduire le renoncement. L'optique est un cas assez particulier puisque pour la majorité des postes de soins les ordonnances ou examens doivent être effectués par un médecin. Dans ce cas, afin de limiter le temps d'attente, le numerus clausus actuellement appliqué à la fin de la première année de médecine devrait être supprimé dès la rentrée 2020. Les effets de cette mesure ne se ressentiront pas immédiatement mais elle a pour but d'anticiper la demande toujours croissante de soins médicaux sur le long terme.

Il est important de noter que la notion de renoncement aux soins médicaux est à manipuler avec une extrême prudence. En effet, elle est le reflet de ce que ressentent les personnes interrogées et surtout de la manière dont la question est posée. L'état de santé et la nécessité d'avoir recours à des soins peuvent parfois être subjectifs : certaines personnes considèrent renoncer à un soin quand d'autres, dans la même situation, ne le penseraient peut-être pas. Les chiffres montrent par exemple que les femmes déclarent plus renoncer à des soins médicaux que les hommes. Cela s'explique par le fait qu'à état de santé comparable, les femmes



---

tendent à s'estimer en plus mauvais état de santé que les hommes [14]. C'est cette subjectivité qui fait que certaines enquêtes annoncent un taux de renoncement aux soins de 50% de la population, alors que d'autres n'affichent qu'un taux de 25%.

## 2 Les principales mesures de la réforme

Comme cela a été brièvement abordé précédemment, la réforme prévoit la mise en place de plusieurs paniers de soins qui s'adaptent aux différents moyens financiers et besoins de chaque individu. De manière générale, le panier « 100% santé » permettra à l'assuré de n'avoir aucun reste à charge lors de l'achat d'un équipement optique, d'une prothèse dentaire ou auditive et les praticiens seront dans l'obligation de proposer un devis sans reste à charge à chaque patient. De plus, cette réforme modifie également le cahier des charges des contrats responsables pour les trois postes et leurs paniers. Chaque poste ayant ses particularités, il peut être intéressant de les étudier séparément.

### Le dentaire

Trois grandes branches des soins dentaires sont réformées dans le cadre de la loi de financement de la Sécurité Sociale de 2019 : les soins conservateurs, les soins chirurgicaux et les prothèses dentaires. Pour ces trois branches, une hausse des bases de remboursement est prévue. La réforme « 100% santé » ne touche que les prothèses dentaires et certains soins conservateurs, c'est pourquoi les soins chirurgicaux ne seront pas étudiés ici. Parmi les prothèses dentaires, les cinq actes concernés sont les inlay-core, les inlay-onlay, les couronnes, les bridges et les prothèses amovibles. La réforme nécessitant des efforts mutuels de la part de la Sécurité Sociale, des complémentaires santé et des praticiens, des Honoraires Limites de Facturation (HLF) sont également mis en place. Il s'agit de plafonds de prix que les praticiens sont dans l'obligation de respecter, permettant ainsi de limiter le montant que la complémentaire santé aura à rembourser.

Le 21 juin 2018, la convention prévoyant l'instauration de trois paniers de soins pour le poste dentaire a été signée par les présidents respectifs de l'Union Nationale des Organismes d'Assurance Maladie Complémentaire (UNOCAM), de la Confédération Nationale des Syndicats Dentaires (CNSD) et de l'Union Dentaire (UD). Cette convention, entrée en vigueur le 1<sup>er</sup> Janvier 2019 et mise en place progressivement jusqu'en 2023, propose aux patients le choix entre un panier « 100% santé », un panier proposant des honoraires maîtrisés ou un panier proposant des honoraires libres.

**Panier « 100% santé »** Il offre des prothèses dentaires sans aucun reste à charge pour l'assuré. Les matériaux constituant les prothèses varient selon la position de la dent : si cette dernière est visible, une prothèse dentaire céramo-métallique est prise en charge ; alors que si la dent n'est pas visible, seule une prothèse métallique entrera dans les critères de ce panier.

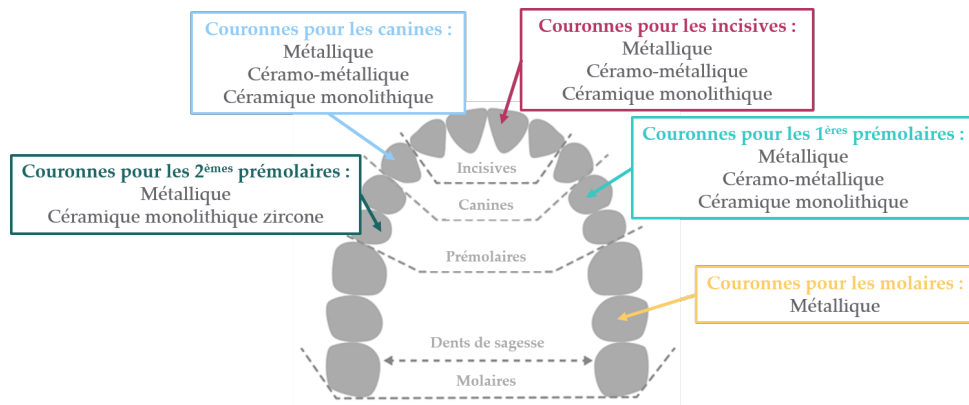


FIGURE 2.4 – Couronnes prises en charge dans le panier « 100% santé » en dentaire

**Panier maîtrisé** Il permet à l'assuré d'être plus libre dans son choix de prothèse que le panier « 100% santé » tout en fixant des Honoraires Limites de Facturation (HLF) qui assurent un reste à charge réduit. Ainsi par exemple, une couronne céramique monolithique zirconie est prise en charge pour une molaire, ce qui n'est pas le cas dans le panier « 100% santé ».

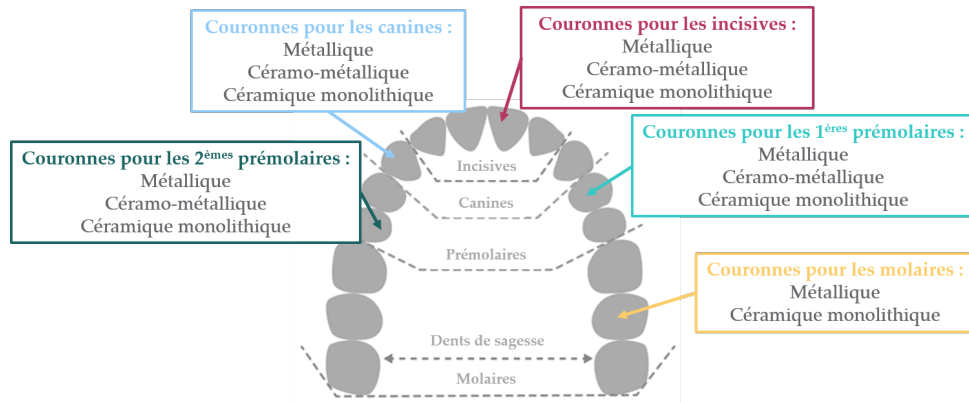


FIGURE 2.5 – Couronnes prises en charge dans le panier maîtrisé en dentaire

**Panier libre** Ce panier permet aux assurés qui peuvent faire face à un reste à charge élevé de choisir la prothèse dentaire de leur choix. C'est également le seul panier à permettre de choisir une prothèse dentaire céramo-céramique ou encore une prothèse en alliage précieux.

Pour ces trois paniers, les BR des prothèses dentaires seront augmentées au fur et à mesure de l'application de la réforme. De plus, pour certains types de soins conservateurs, la BR augmentera également entre 2019 et 2023, le but étant d'inciter les populations les plus fragiles à choisir des actes de prévention plutôt que d'attendre jusqu'à ce qu'une prothèse dentaire devienne la seule solution. Un tableau présentant les BR en dentaire à la suite de la réforme a été publié au Journal Officiel n°0195 du 25 août 2018, texte n°5.

Le tableau ci-dessous résume les différents honoraires limites de facturation par type de dent et par panier de soins.

Actes		Honoraires limites de facturation				
		Incisives	Canines	1 <sup>re</sup> Prémolaire	2 <sup>me</sup> Prémolaire	Molaires
Couronnes	Métallique	290 €				
	Céramique monolithique zircon	440 €				440 €
	Céramo-métallique	500 €		550 €		
	Céramique monolithique hors zircon	500 €		550 €		
	Céramo-céramique					
Soins prothétiques	Inlay/onlay céramique					
	Inlay/onlay composite	350 €				
	Inlay core sur couronne 100% santé	175 €				
	Inlay core sur couronne panier maîtrisé	175 €				
	Inlay core sur couronne panier libre					
	Bridge	1 465 €	1 635 €			

Panier 100% santé
Panier maîtrisé
Panier libre (garantie actuelle en €)

FIGURE 2.6 – Honoraires limites de facturation en dentaire en fonction des paniers de soins

**Conditions pour un contrat responsable [15] :** En dentaire, un contrat responsable a l'obligation de prendre en charge la totalité des dépenses liées à des prothèses entrant dans le panier « 100% santé ». Il doit également prendre en charge une couverture minimale pour les prothèses du panier maîtrisé. Il n'y a pas d'obligation de remboursement minimal ou maximal dans le cadre des prothèses du panier libre.

## L'auditif

Avant la réforme « 100% santé », le remboursement des aides auditives variait selon la classe de l'équipement (classe A à D), chaque classe étant définie en fonction du degré de perte auditive. La réforme supprime ces classes et crée deux paniers de soins : le panier « 100% santé » et le panier libre. Pour chaque panier, la réforme prévoit une augmentation progressive de la BR pour les adultes et une BR fixée à 1400€ dès 2019 pour les enfants.

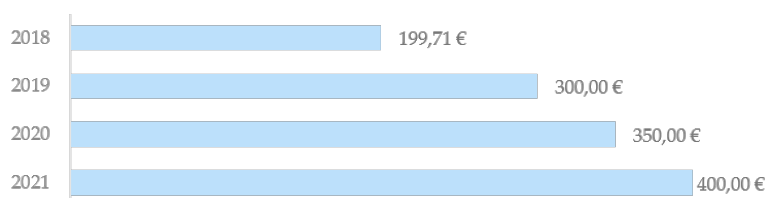


FIGURE 2.7 – Évolution de la BR pour les adultes suite à la mise en place de la réforme « 100% santé »

Par ailleurs, la réforme impose que, quelque soit le panier choisi, huit caractéristiques soient automatiquement incluses : 12 canaux de réglage minimum, au moins 2 programmes différents (un environnement calme, un environnement bruyant), réducteur du bruit, système d'amplification, anti-larsen, directivité, étanchéité, enregistrement des données<sup>1</sup>. Enfin, deux listes sont mises en place (la liste A et la liste B). Elles comportent des options

1. Notamment le nombre d'heures portées, le fonctionnement de la synchronisation binaurale le cas échéant, l'activation des programmes, le volume d'amplification utilisé

parmi lesquelles l'assuré pourra choisir en fonction du panier qu'il choisira. Ces listes sont présentées dans l'Annexe A.

**Panier « 100% santé »** Il est également appelé « Classe I ». Il propose à l'assuré de choisir trois options dans la liste A et lui garantit la prise en charge de la totalité des frais liés à un appareillage. Pour ce panier, le prix limite de vente pour un appareil pour enfant est de 1400€ et celui pour un appareil pour adulte évolue de la façon suivante :

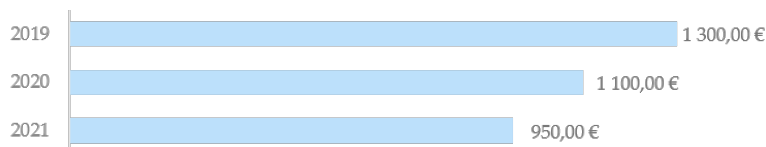


FIGURE 2.8 – Évolution des prix limites de vente pour les adultes suite à la mise en place de la réforme « 100% santé »

**Panier libre** Il est également appelé « Classe II ». Il propose à l'assuré de choisir six options dans la liste A et une dans la liste B. Il n'y a techniquement pas de prix limites de vente, mais pour que le contrat soit responsable la prise en charge ne peut pas excéder 1700€ par appareil, aussi bien pour un enfant que pour un adulte.

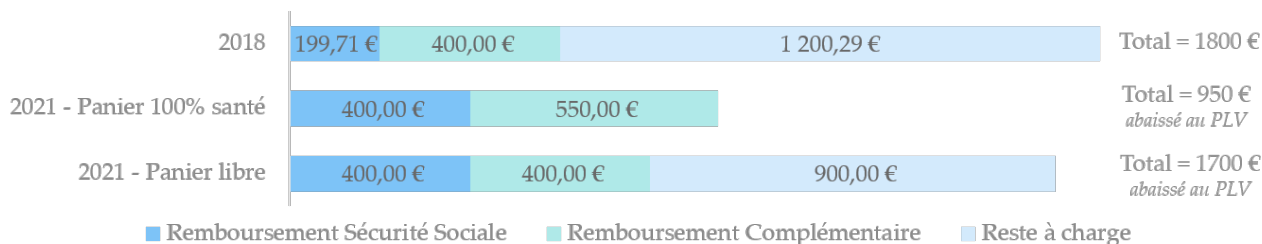


FIGURE 2.9 – Exemple de remboursement d'un appareil auditif coûtant 1 500€ avant et après la réforme « 100% santé »

## L'optique

Comme pour l'auditif, la réforme prévoit pour l'optique deux paniers de soins : le panier « 100% santé » et le panier libre. Dans chaque panier, une distinction est faite entre les verres et la monture.

**Panier « 100% santé »** Il est également appelé « Classe A » et propose une prise en charge à hauteur de 100% des frais réels.

Monture	Verres
Garantie 2 ans en cas de casse <b>Adultes :</b> La BR passe de 2,84€ à 9€ PLV fixés à 30€ 17 modèles dans deux coloris <b>Enfants :</b> La BR passe de 30,49€ à 15€ PLV fixés à 50€ pour les moins de 6ans 10 modèles dans deux coloris	Toutes les corrections sont prises en charge Plus de distinction entre adulte et enfant Augmentation des BR PLV allant de 32,50€ à 170€ Prise en charge de l'amincissement, des traitements anti-rayure, anti-reflet, des filtres et de l'appairage Garantie 3 mois en cas d'inadaptation

Pour un adulte, les prix limites de vente, incluant les 30€ de monture, du panier « 100% santé » sont les suivants :

Corrections en dioptrie	Équipements unifocaux			Équipements progressifs		
	Sphérique	Astigmat +		Sphérique	Astigmat +	
		Cylindre de 0 à 4	Cylindre > 4		Cylindre de 0 à 4	Cylindre > 4
Niveau 1 – entre 0 et 2 dioptries	95 €	105 €	135 €	180 €	210 €	260 €
Niveau 2 – entre 2 et 4 dioptries	105 €	115 €	145 €	190 €	220 €	270 €
Niveau 3 – entre 4 et 8 dioptries	125 €	135 €	165 €	210 €	240 €	290 €
Niveau 4 – entre 8 et 12 dioptries	225 €	235 €	265 €	290 €	320 €	370 €
Niveau 5 – supérieur à 12 dioptries						

FIGURE 2.10 – Prix limites de vente pour les adultes suite à la mise en place de la réforme « 100% santé »

**Panier libre** Il est également appelé « Classe B ». Ce panier permet à l'assuré de choisir la monture de son choix, quel qu'en soit le prix ou la marque et également de choisir une qualité de verres supérieure à celle du panier « 100% santé ». Cependant, la BR des montures et des verres de ce panier est drastiquement réduite à 0,05€, ce qui implique un désengagement quasi total de la Sécurité Sociale sur ces actes.

**Condition pour un contrat responsable** Pour qu'il soit responsable, la réforme impose qu'un contrat couvre entièrement les dépenses liées au panier « 100% santé » ainsi que celles liées à certains actes complémentaires. D'autre part, une couverture minimale et une maximale pour le panier libre sont mises en place. Ces plafonds sont fixés dans le Journal Officiel n°0090 du 16 avril 2019, texte n° 19. Enfin, la prise en charge d'une monture est limitée à 100€, au lieu de 150€ avant la réforme.

### 3 Les étapes de la mise en place de la réforme

La réforme « 100% santé » s'applique en trois phases. Au début de l'année 2021, le reste à charge nul sera assuré pour les trois postes et pour tous les actes concernés. La frise ci-dessous (Figure 2.11) présente de façon plus détaillée les différentes étapes de mise en place de la réforme. L'abréviation "BR" signifie Base de Remboursement, tandis que "PLV" se réfère aux Prix Limites de Ventes et "HLF" aux Honoraires Limites de Facturation.

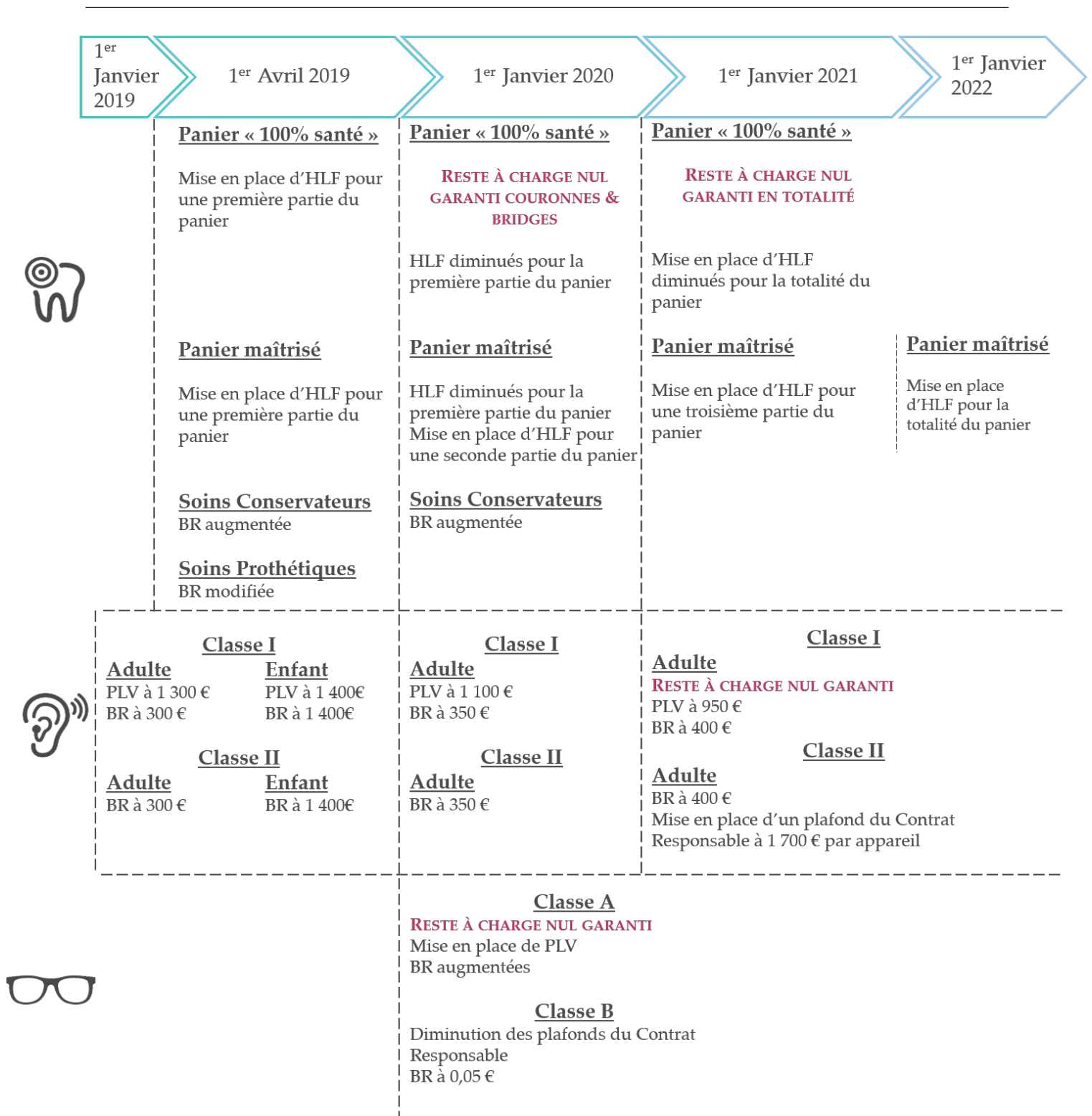


FIGURE 2.11 – Calendrier de mise en place de la réforme « 100% santé »

## 4 Les conséquences pour les contrats des complémentaires santé

Tout d'abord, la mise en place de la réforme « 100% santé » implique un changement dans les tableaux de garanties des complémentaires santé. En effet, les différents paniers de soins doivent à présent y figurer et les plafonds des contrats responsables doivent être modifiés (ou ajoutés selon les situations). Le remboursement intégral du reste à charge

---

après intervention de la Sécurité Sociale pour les paniers « 100% santé » doit également être clairement explicité. Les complémentaires santé devront donc mettre en conformité leurs contrats. C'est-à-dire, revoir la totalité des contrats existants et les mettre à jour, en prenant en compte les options qui figurent dans les contrats pour les paniers autres que le panier « 100% santé ». Les nouveaux contrats devront également être conformes aux exigences de la réforme « 100% santé ».

Par ailleurs, le gouvernement a profité de la réforme « 100% santé » pour imposer aux complémentaires santé d'améliorer la lisibilité de leurs garanties. Ainsi, depuis janvier 2019 elles doivent toutes utiliser les mêmes intitulés pour une grande partie des garanties des cinq grands postes de soins. Cette mesure vise à permettre aux assurés de comparer plus facilement les contrats de différentes complémentaires et donc de faire un choix de façon plus éclairée. De plus, pour améliorer la compréhension des garanties, une base commune d'exemples doit également être utilisée par toutes les complémentaires santé et un glossaire définissant tous les termes techniques relatifs aux soins de santé ainsi qu'à leur remboursement est aussi établi. Enfin, le gouvernement recommande également de mettre à disposition des assurés un simulateur de remboursement, qui n'est cependant pas obligatoire.

De façon plus générale, la réforme « 100% santé » aura très certainement un impact sur les prestations des complémentaires santé. En effet, même si le renoncement aux soins a plusieurs origines, il a été établi précédemment que la raison financière restait majoritaire. Ainsi, bien que le gouvernement ait demandé à ces dernières de ne pas augmenter les primes des assurés dans un premier temps, il est fort probable qu'à partir de 2021 elles n'aient pas d'autre choix. En effet, des enjeux de rentabilité et des contraintes liées au ratio sinistres sur primes (autrement appelé S/P) obligeront très probablement les complémentaires à revoir leur tarification sur les contrats santé. La réforme ayant pour but de réduire les différences de remboursements entre les niveaux de garanties, ce changement de tarif devrait donc leur être étroitement lié. Cette modification tarifaire sera à l'appréciation de chaque organisme, en fonction de sa solvabilité, de ses critères de rentabilité et de ses choix commerciaux. Il est également possible que certains organismes ne fassent évoluer leurs tarifs qu'après la mise en place totale de la réforme et les premières observations d'impacts. C'est dans ce cadre que s'inscrit l'étude de ce mémoire : mieux appréhender les différents impacts de la réforme afin d'aider les assureurs à anticiper l'évolution tarifaire.

### En résumé

La réforme « 100% santé » apporte des changements sur la prise en charge des soins des postes optique, audiologie et dentaire dans le but de réduire le renoncement aux soins. Après sa mise en place, les assurés auront la possibilité de choisir leurs soins parmi des paniers : le panier « 100% santé » pour lequel le montant qui reste à leur charge est nul, ou les paniers maîtrisés et libres qui nécessiteront qu'ils paient une partie des soins. Cette réforme se met progressivement en œuvre depuis 2019, jusqu'en 2023.

### 3 | LA BASE DE DONNÉES

Afin de respecter le Règlement Général sur la Protection des Données (RGPD), la base de données est anonymisée : les données permettant d'identifier précisément un assuré sont supprimées ou modifiées, seuls les identifiants créés par l'assureur sont conservés afin de pouvoir lier les différentes tables.

#### 1 Description de la base de données

La base de données utilisée pour l'étude est la base santé d'un client d'ADDACTIS France. Elle comporte six tables présentées dans la Figure 3.1.

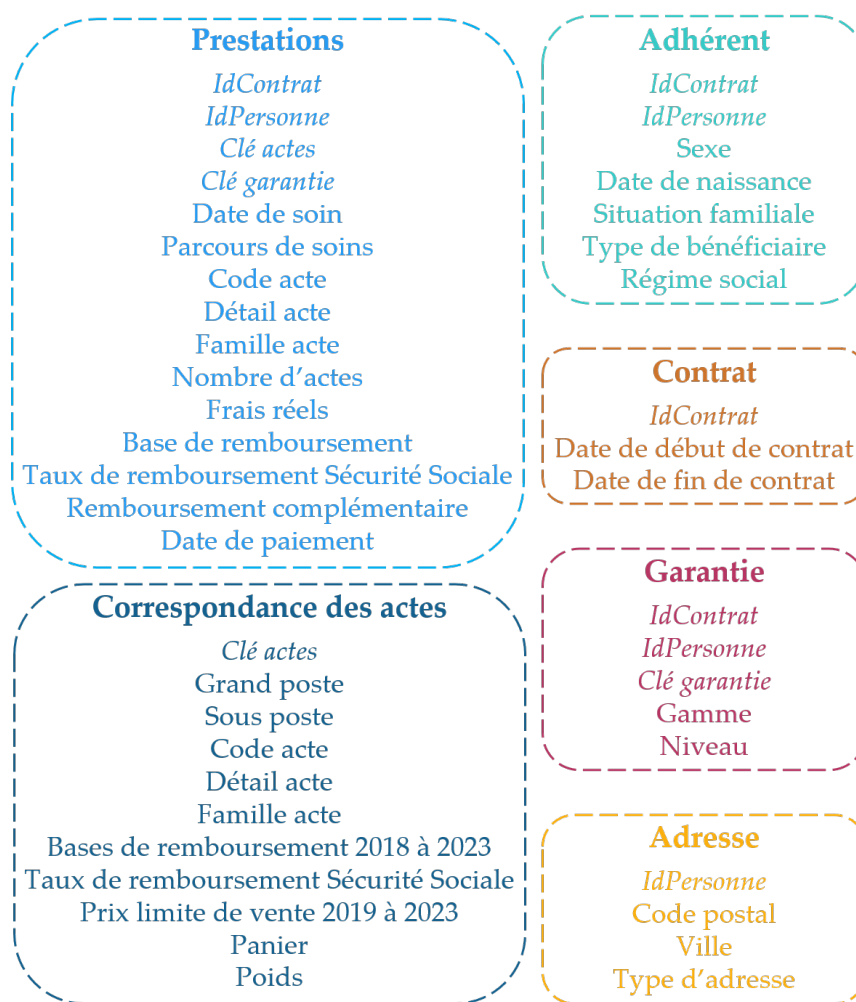


FIGURE 3.1 – Présentation des tables composant la base de données d'étude

L'Annexe B détaille la signification de toutes les variables présentées à la Figure 3.1. Cependant, certaines d'entre elles nécessitent quelques précisions dans le corps de ce mémoire quant aux différentes modalités qui les composent, que ce soit pour mieux appréhender la maille de l'étude, ou pour mieux comprendre l'information apportée par la variable. Le tableau de la Figure 3.2 détaille les modalités ambiguës.



Variable	Modalités
Régime social	Régime général ou autre régime
Situation familiale	Marié, divorcé, célibataire, concubin ou situation inconnue
Type de bénéficiaire	Assuré, conjoint, enfant ou assuré non souscripteur
Gamme de garantie	Solo, famille ou senior
Niveau de garantie	Entrée de gamme, milieu de gamme ou haut de gamme
Type d'adresse	Principale ou secondaire

FIGURE 3.2 – Modalités des variables de la base de données

Il est important de noter que l'ensemble des contrats sont des contrats individuels, les contrats collectifs ne seront pas étudiés dans ce mémoire.

## 2 Création de variables

Certaines variables nécessaires à l'étude des impacts de la réforme « 100% santé » sur le portefeuille ne sont pas directement disponibles dans la base de données. Elles doivent donc être créées en amont de l'étude. Le volume de données étant relativement important, cette étape est effectuée sous SAS.

Tout d'abord, l'âge du bénéficiaire en 2018 est calculé à partir de son année de naissance puis des classes d'âges sont créées afin d'obtenir des statistiques plus exploitables : entre 0 et 15 ans, 16 et 30 ans, 31 et 50 ans, 51 et 64 ans, et enfin 65 ans et plus. Cette dernière classe comprend un intervalle d'âges plus grand car les robustesse des données n'est pas suffisante.

De plus, il peut être pertinent aussi bien dans l'étude des données que dans l'étude de l'impact de la réforme « 100% santé » d'observer l'évolution du reste à charge de l'assuré après les remboursements de la Sécurité Sociale et de l'assurance complémentaire. Cette variable n'existant pas initialement dans la base de données, elle est calculée à partir de la formule suivante :

$$\text{Reste à charge} = \text{Frais réels} - \frac{\text{Remboursement de la Sécurité Sociale}}{\text{Remboursement complémentaire}}$$

Le remboursement de la Sécurité Sociale est donc également à calculer à partir des variables disponibles dans la base de données :

$$\frac{\text{Remboursement de la Sécurité Sociale}}{\text{Remboursement de la Sécurité Sociale}} = \frac{\text{Taux de remboursement de la Sécurité Sociale}}{\text{Base de remboursement}}$$

Par ailleurs, il est nécessaire de calculer l'exposition de chaque contrat. Cette variable mesure la durée dans l'année pendant laquelle le contrat est effectif. Ainsi par exemple, un contrat débutant au 1<sup>er</sup> juin se voit attribuer une exposition de 0,5 ; cela signifie qu'il compte en réalité pour un demi contrat dans l'année étudiée.

$$\text{Exposition} = \frac{\text{Durée de couverture sur une année}}{\text{Nombre de jours dans l'année}} \in [0 ; 1]$$

Il est finalement possible de sommer les expositions de chaque contrat afin de mesurer l'exposition totale du portefeuille, ce qui équivaut au nombre de contrats actifs dans le portefeuille au prorata du temps qu'il est observé.

De plus, la variable nombre de sinistres n'existe pas dans la base de données, seule la variable nombre d'actes est référencée. Bien que ces deux variables soient proches, il peut y avoir plusieurs actes pour un même sinistre : par exemple, pour une même ordonnance plusieurs médicaments peuvent être délivrés. Ainsi, l'utilisation de la variable nombre d'actes provoquerait une surestimation de la sinistralité réelle du portefeuille. Pour créer la variable nombre de sinistres, il est considéré que tous les remboursements pour une même référence d'actes faits le même jour pour le même assuré font partie du même sinistre.

D'autre part, la fréquence de consommation d'un assuré est une variable importante, qui permet de coupler le nombre de sinistres qu'il a déclaré avec la durée pendant laquelle il a été couvert. En effet, un assuré couvert pour les trois derniers mois de l'année déclarant trois sinistres a une fréquence de consommation de soins de santé bien plus élevée qu'un assuré couvert toute l'année ayant lui aussi déclaré trois sinistres. Ainsi, la fréquence de consommation se calcule après avoir agrégé la base de données par bénéficiaire de la façon suivante :

$$\text{Fréquence de consommation des soins de santé} = \frac{\text{Nombre de sinistres}}{\text{Exposition}}$$

Enfin, la variable géographique présente dans la base de données est le code postal (ou la ville, ces deux variables sont équivalentes), qui est bien trop précis pour la suite de l'étude. En effet, il est nécessaire de regrouper l'information géographique en un nombre restreint de modalités. L'agrégation des codes postaux peut se faire de différentes manières : selon le nombre d'habitants ou la densité de population observés par l'INSEE, ou encore selon l'exposition du portefeuille dans les différentes zones. Finalement, c'est la densité d'exposition qui est retenue. Il s'agit du rapport entre l'exposition totale d'un département dans le portefeuille assuré et la superficie de ce département. Tous les codes postaux étrangers à la France métropolitaine sont regroupés en une seule modalité « Étranger ».

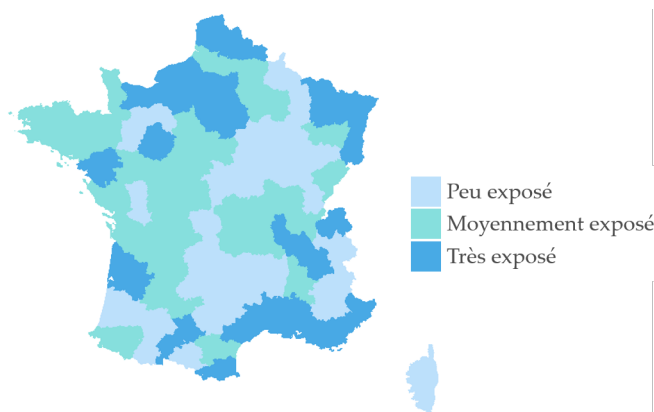


FIGURE 3.3 – Découpage des régions obtenu à partir de l'exposition du portefeuille

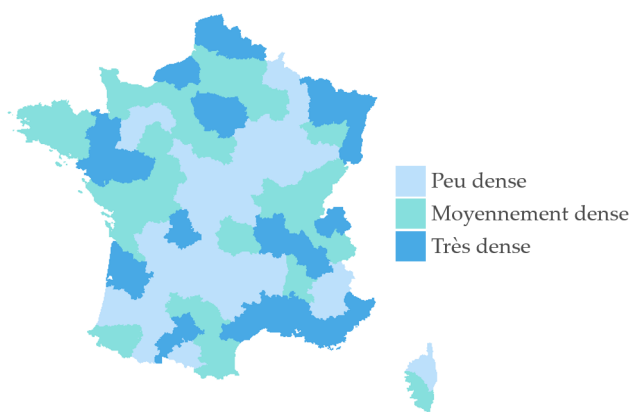


FIGURE 3.4 – Découpage des régions obtenu avec les densités de population INSEE

La carte obtenue à partir de l'exposition du portefeuille est très proche de celle de l'INSEE bien que la diagonale du vide soit légèrement sur-exposée dans le portefeuille client.

---

### 3 Qualité des données

L'appréciation de la qualité des données est définie dans la Norme de Pratique Actuarielle 1 (ou NPA 1) [13]. Ainsi, l'étude doit respecter trois points pour être considérée comme utilisable par un actuare : les données doivent être fiables et suffisantes ; elles doivent être validées par l'actuaire ; et les sources de données doivent si possible être spécifiques à l'entité.

Pour que des données soient suffisantes, elles doivent inclure les informations pertinentes nécessaires à la réalisation du travail. Les données de l'étude vérifient cette condition : elles regroupent d'une part les prestations santé d'un client du cabinet qui vont servir de base à la mise en As-if et au calcul des impacts de la réforme « 100% santé », et d'autre part les caractéristiques personnelles des assurés qui permettront l'identification de profils de risque.

La significativité des données impose qu'elles ne contiennent pas d'erreurs significatives. Il s'avère que parfois certaines valeurs aberrantes apparaissent dans la base de données. En effet, il arrive que des lignes de prestations présentent des frais réels ou des remboursements complémentaires nuls voire négatifs. Afin de conserver une cohérence dans l'étude, ces lignes sont supprimées de la base de données. D'autre part, pour chaque poste impacté par la réforme, les montants de prestations supérieurs au quantile à 99% sont considérés comme des valeurs aberrantes ou des sinistres graves (qui présentent une fréquence faible et un coût élevé) et sont donc supprimés. En effet, le but de cette étude n'est pas d'observer précisément les sinistres graves mais plutôt la masse de sinistres dits attritionnels qui représentent la grande majorité des sinistres du portefeuille.

La validation des données utilisées pour l'étude passe par la vérification de leur cohérence, de leur complétude et de leur exactitude. Dans le cas des données utilisées dans cette étude, il est impossible de les comparer avec celles d'un exercice précédent puisque l'historique des prestations avant 2018 n'a pas été fourni par le client. Cependant, la cohérence des données a été établie par l'ancienneté du portefeuille et garantie par le client. Les données sont complètes dans le sens où tous les assurés ayant souscrit à un contrat individuel sont correctement inclus dans la base.

Enfin, les données utilisées sont spécifiques au portefeuille client sur lequel l'étude est menée. Par ailleurs, les hypothèses prises tout au long de cette étude sont prises en accord avec le client et conformes au portefeuille de l'étude.

### 4 Statistiques descriptives

La base de données compte un peu plus de 200 000 individus en 2018. En comparant la pyramide des âges de la base (Figure 3.5) avec celle de la population française (Figure 3.6), il apparaît que les enfants et les 50-65 ans sont sous-représentés. Par ailleurs, la proportion de femmes dans le portefeuille client est légèrement supérieure à celle des hommes. Ces remarques seront à prendre en considération lors des différentes interprétations. En effet, le portefeuille n'étant pas une réplique fidèle de la totalité de la population française, certaines conclusions pourront être faites pour le portefeuille client mais ne pourront peut-être pas être généralisées.

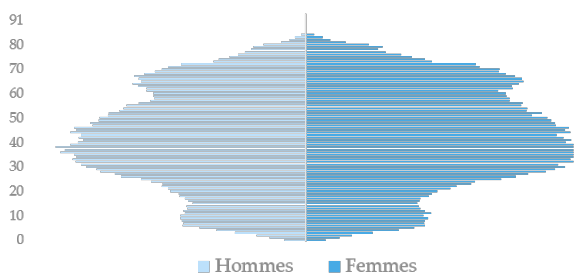


FIGURE 3.5 – Pyramide des âges du portefeuille client en 2018

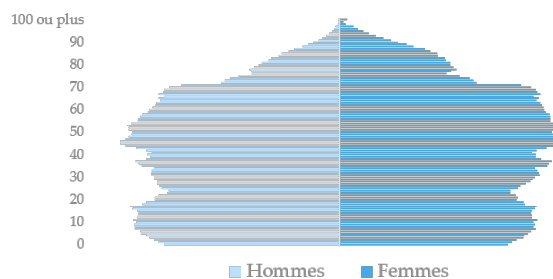


FIGURE 3.6 – Pyramide des âges de la population française en 2018

D'autre part, il peut être intéressant d'étudier la répartition des âges dans le portefeuille en fonction de l'exposition des assurés. En effet, comme cela a été décrit précédemment, l'exposition permet de considérer un individu au prorata de sa présence dans le portefeuille et ainsi d'étudier la réelle répartition du portefeuille.

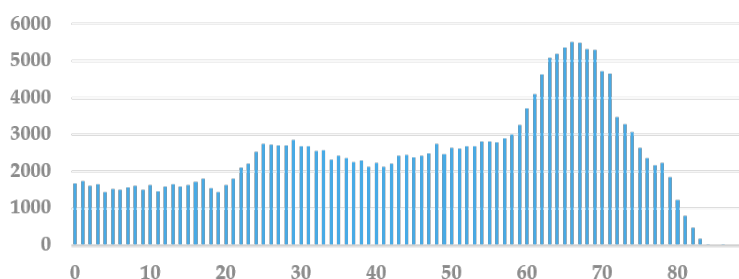


FIGURE 3.7 – Répartition du portefeuille, en terme d'exposition, selon l'âge des assurés

Ainsi, on remarque que bien que les personnes âgées ne paraissent pas plus présentes qu'une autre classe d'âge sur la figure 3.5, elle sont les plus représentées en terme d'exposition (Figure 3.7). Cela aura un impact non négligeable sur les résultats de l'étude étant donné que cette partie de la population a généralement tendance à consommer plus en terme de soins de santé. Cette considération se vérifie dans le portefeuille client lors de l'étude des frais réels, des remboursements complémentaires et des remboursements Sécurité Sociale 2018 (Figure 3.8) qui augmentent à mesure que l'âge de l'assuré augmente.

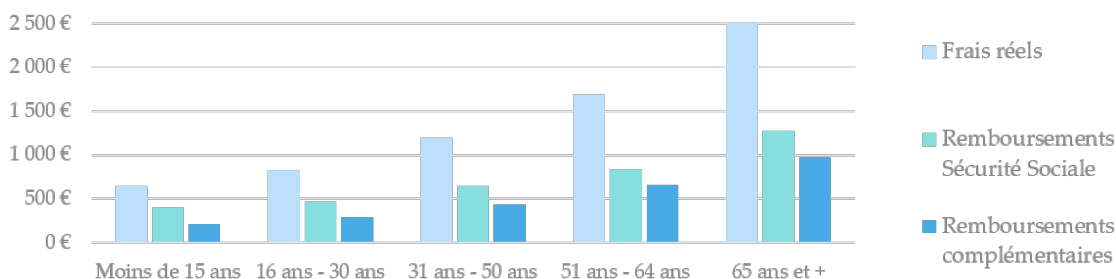


FIGURE 3.8 – Frais réels, remboursements complémentaires & remboursements Sécurité Sociale moyens pour un individu, en 2018 et selon son âge

Par ailleurs, l'assureur propose trois gammes de garanties différentes elles-mêmes déclinées en trois niveaux. Les trois garanties sont adaptées aux différentes étapes de la vie : la garantie Solo assure un unique bénéficiaire, la garantie Famille permet d'assurer une famille et la garantie Senior assure les personnes de plus de 60 ans. Les trois niveaux de garantie

(entrée, milieu ou haut de gamme) permettent à l'assuré de choisir le type de couverture qui lui convient. Le niveau entrée de gamme offre principalement à l'assuré un remboursement à 100% de la BR y compris Sécurité Sociale, alors que le niveau haut de gamme propose en général à l'assuré des prestations remboursées quasiment intégralement.

	Soins dentaires	Couronne	Bridge	Prothèses métalliques
<b>Entrée de gamme</b>	100% de la BR	275€	400€	150% de la BR
<b>Milieu de gamme</b>	100% de la BR	330€	540€	225% de la BR
<b>Haut de gamme</b>	100% de la BR	425€	780€	300% de la BR

	Prothèse auditive <i>Solo ou Famille</i>	Prothèse Auditive <i>Senior</i>	Entretien <i>Senior</i>
<b>Entrée de gamme</b>	100€	300€	20€
<b>Milieu de gamme</b>	175€	450€	30€
<b>Haut de gamme</b>	250€	600€	40€

	Monture	Verre simple	Verre complexe
<b>Entrée de gamme</b>	50€	40€	80€
<b>Milieu de gamme</b>	55€	57,5€	125€
<b>Haut de gamme</b>	80€	85€	175€

FIGURE 3.9 – Extraits des tableaux de garanties proposés par l'assureur

La Figure 3.9 présente des extraits des tableaux de garanties dans les trois postes impactés par la réforme : le dentaire, l'audiologie et l'optique. Sur les actes présentés en optique et en dentaire, les garanties sont les mêmes quelle que soit la gamme de garantie. Les garanties en euro sont additionnelles au remboursement de la Sécurité Sociale, alors que les garanties en pourcentage de la BR prennent en compte le remboursement de la Sécurité Sociale (par exemple une prothèse métallique entrée de gamme est remboursée à 150% de la BR et le taux de remboursement de la Sécurité Sociale est de 70% en dentaire ; le remboursement complémentaire sera donc de 80% de la BR).

L'étude des tableaux de garanties fait apparaître que les contrats sont responsables. Cela permettra par la suite d'appliquer dans l'étude les plafonds liés aux contrats responsables.

Par la suite, la répartition du portefeuille dans les différentes gammes ainsi que dans les niveaux de garantie devrait avoir une importance dans l'analyse des résultats de l'étude. En effet, la réforme « 100% santé » étant prévue pour réduire le reste à charge des personnes ayant renoncé à des soins pour raisons financières, il est possible d'imaginer que les assurés ayant opté pour une garantie entrée de gamme soient les plus visés par la réforme. De même, les personnes âgées sont les plus grandes consommatrices de prothèses auditives. La connaissance du portefeuille sur ce point est donc nécessaire.

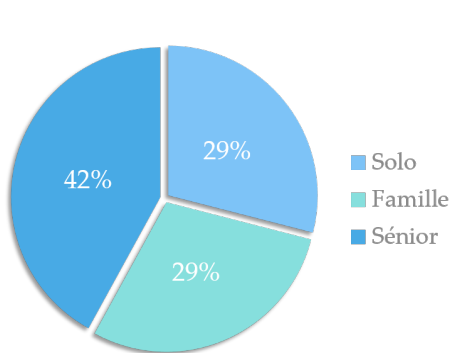


FIGURE 3.10 – Répartition des gammes de garantie en fonction de l'exposition du portefeuille

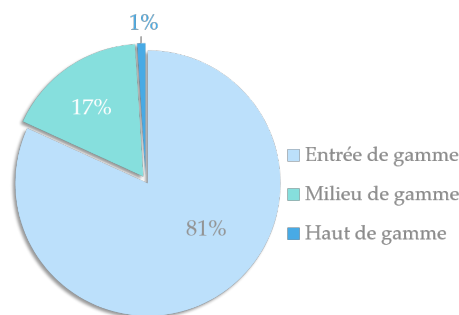


FIGURE 3.11 – Répartition des niveaux de garantie en fonction de l'exposition du portefeuille

La Figure 3.10 montre que la gamme de garantie Senior est la plus représentée dans le portefeuille en terme d'exposition. Les gammes Solo et Famille présentent la même proportion. Cette disparité sera à prendre en compte par la suite.

La Figure 3.11 met en lumière un portefeuille majoritairement axé sur des garanties entrée de gamme. Seulement 1% des assurés ont opté pour une garantie haut de gamme (toutes gammes confondues). Cela peut s'expliquer par le fait que le portefeuille soit un portefeuille individuel, ainsi très peu de salariés le composent et de nombreux TNS<sup>1</sup> ou seniors optent pour une couverture santé minimale, faute de moyens.

Suite à l'étude des répartitions en terme d'exposition dans le portefeuille, il apparaît intéressant d'étudier les caractéristiques des prestations 2018 effectuées en fonction du niveau de garantie. La Figure 3.12 détaille les frais réels, remboursements complémentaires et remboursements de la Sécurité Sociale selon le niveau de garantie et le poste auquel appartient l'acte médical. Seuls les trois postes impactés par la réforme sont étudiés ici.

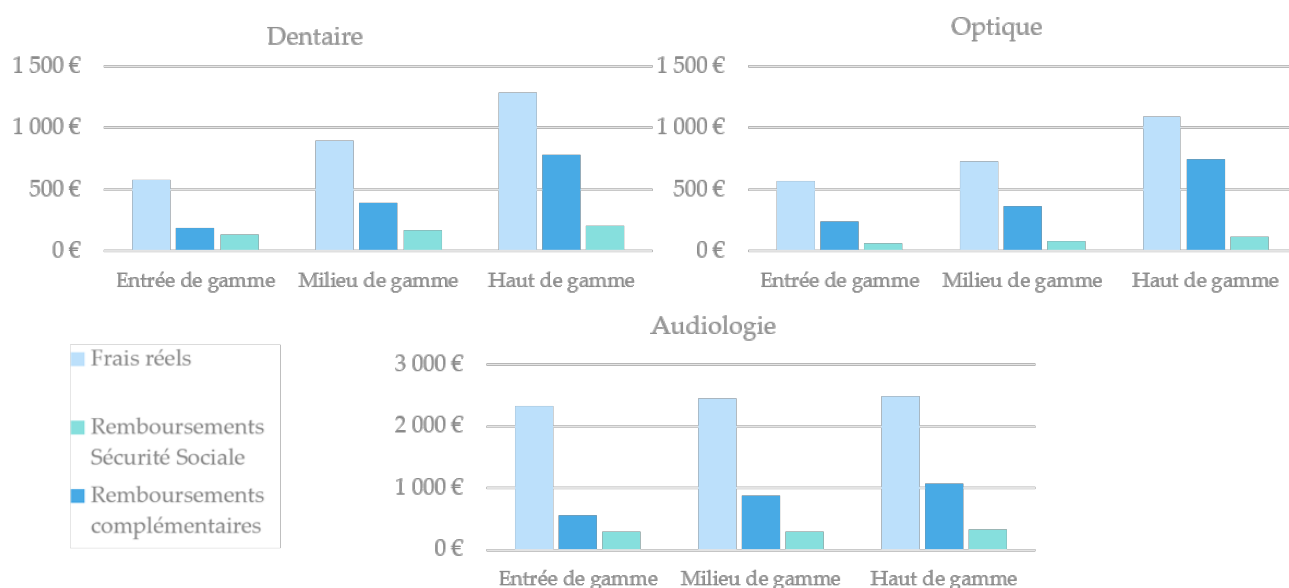


FIGURE 3.12 – Frais réels, remboursements complémentaires & remboursements Sécurité Sociale moyens pour un individu, en 2018 et selon le niveau de garantie

1. Travailleurs non salariés

Il apparaît pour les trois postes de soins que plus le niveau de garantie augmente, plus les frais réels et le remboursement complémentaire augmentent. En dentaire, les frais réels sont plus que doublés entre un assuré ayant opté pour une garantie entrée de gamme et celui ayant choisi une garantie haut de gamme. Le remboursement complémentaire est quant à lui quasiment multiplié par quatre. En optique, le constat est très similaire bien que les augmentations soient légèrement plus faibles. Cependant, en audiologie, les frais réels sont quasiment identiques pour les trois niveaux de garantie. Cela peut s'expliquer par le fait que le renoncement aux soins est très fort pour ce poste, quelle que soit la garantie de l'assuré. Le remboursement complémentaire est tout de même doublé entre une garantie entrée de gamme et une garantie haut de gamme.

D'autre part, l'étude portant sur la réforme « 100% santé », il peut être intéressant d'étudier la répartition des remboursements selon les postes de soins. Ainsi, on remarque dans la Figure 3.13 que les trois postes réformés ne représentent que 17% de la totalité des remboursements 2018. Ainsi, quelles que soient les valeurs des impacts de la réforme sur les remboursements qui seront calculés, ils seront sûrement par la suite dilués dans la masse des remboursements des postes non réformés.

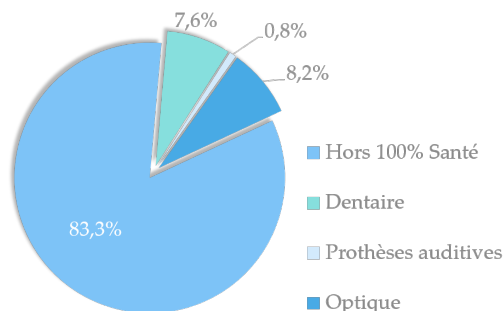


FIGURE 3.13 – Part des remboursements complémentaires par poste

Enfin, un des buts premiers de la réforme « 100% santé » étant de réduire le reste à charge pour les personnes les plus enclines à renoncer à des soins, cela peut être utile d'observer cette variable en fonction du poste et du niveau de garantie de l'assuré. Elle pourrait se déduire de la Figure 3.12 mais les afficher directement les rendra plus simples à analyser.

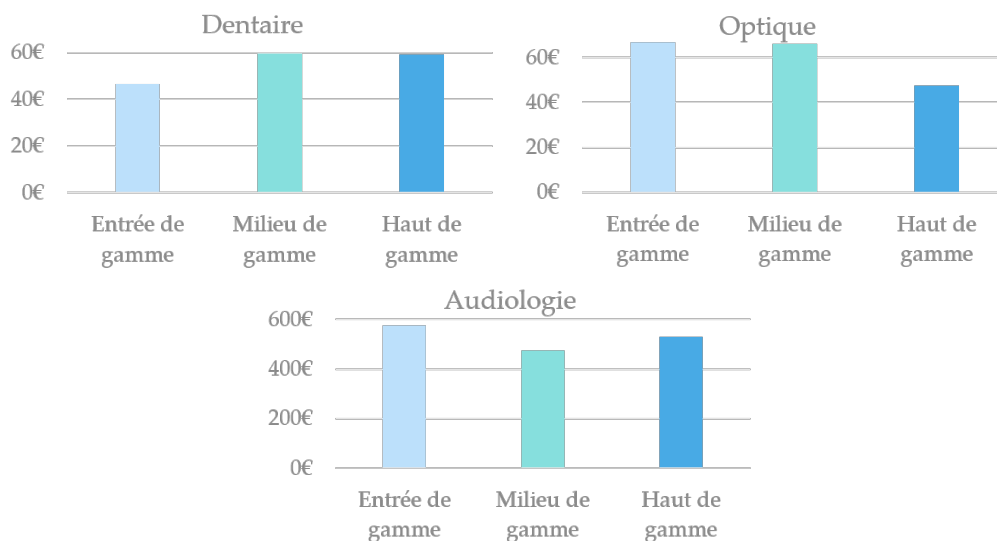


FIGURE 3.14 – Moyenne des restes à charge 2018 par poste et niveau de garantie

---

Comme attendu, le reste à charge d'un assuré pour le poste audiologie est bien plus important que celui des postes dentaire ou optique.

En dentaire, le reste à charge est légèrement inférieur pour des assurés ayant choisi une garantie entrée de gamme. À première vue, cela peut paraître surprenant puisqu'en général plus la garantie est élevée, mieux les frais sont pris en charge par la complémentaire santé et donc moins le reste à charge est élevé. Cependant, il est nécessaire de prendre également en compte le montant initial du soin. En dentaire, de nombreux actes se déclinent selon leur qualité, leur longévité et leur prix en est donc impacté. Un assuré moins bien couvert par sa complémentaire santé aura donc en général tendance à choisir un soin moins cher et ainsi, bien que le remboursement effectué par la complémentaire soit plus bas, le reste à charge le sera également (cela se vérifie à la Figure 3.12).

En optique, le reste à charge diminue à mesure que le niveau de garantie augmente. Cela s'explique par le fait que les frais réels sont quasiment identiques pour les trois niveaux de garantie (cf. Figure 3.12) et que les remboursements complémentaires augmentent avec le niveau de garantie. En effet, l'optique est un poste de soin pour lequel l'assuré est plus enclin à augmenter son reste à charge pour des raisons esthétiques (tel que le modèle de la monture par exemple).

En audiologie, le reste à charge des prestations haut de gamme est supérieur à celui des prestations milieu de gamme. Cependant, la différence entre les deux est plutôt faible (environ 50€ d'écart). Cela peut s'expliquer par le fait que les frais réels augmentent légèrement plus que le remboursement complémentaire (cf. Figure 3.12). Par ailleurs, il est important de remarquer que quel que soit le niveau de garantie, le reste à charge moyen reste très élevé, l'écart observé est négligeable et doit probablement provenir des spécificités du portefeuille étudié. Ainsi, l'impact de la réforme « 100% santé » devrait être important sur les trois niveaux de garantie.

Par la suite, il pourra être intéressant d'étudier l'évolution du reste à charge afin de juger de l'efficacité de la réforme.

### En résumé

Tout d'abord, le portefeuille sur lequel porte l'étude est vieillissant : la majorité des assurés ont plus de 60 ans. Les montants, aussi bien des frais réels, des remboursements de la Sécurité Sociale ou de ceux de la complémentaire santé augmentent avec l'âge de l'assuré.

D'autre part, trois niveaux de garantie sont déclinés pour les trois gammes de l'assureur : Solo, Famille et Senior. Le niveau Entrée de gamme est très largement plus représenté que les niveaux Milieu et Haut de gamme (81% des assurés ont opté pour de l'entrée de gamme), ce qui peut s'expliquer par le fait que l'étude porte sur un portefeuille individuel (incluant ainsi en général une grande majorité de TNS ou de seniors).

Enfin, les postes impactés par la réforme « 100% santé » représentent 16,6% des remboursements complémentaires effectués par l'assureur.



## DEUXIÈME PARTIE

### IMPACTS DE LA RÉFORME

---

La première partie de cette étude propose de calculer l'impact de la réforme « 100% santé » sur les prestations versées par le client à ses assurés. Dans un premier lieu, les impacts seront calculés sans prendre en compte la déformation de consommation de soins des assurés qui devrait être induite par la mise en place du panier « 100% santé ». Ensuite, seront ajoutées lesdites déformations de fréquences de consommation afin d'obtenir un impact plus réaliste de la réforme. Cependant, la mise en place de ces déformations repose nécessairement sur des hypothèses fortes quant au comportement du portefeuille client face à la réforme.

# 1 | TRAVAIL PRÉPARATOIRE SUR LA BASE DE DONNÉES

La méthode de calcul de l'impact de la réforme sera détaillée dans le chapitre suivant. Néanmoins, il est tout d'abord nécessaire d'agréger les données pour les rendre utilisables par la suite. Pour cela, plusieurs tâches sont à effectuer en amont.

## 1 Tables de Bases de Remboursement et de Prix Limites de Vente

En premier lieu, il est nécessaire de disposer de tables « propres » dans lesquelles se trouvent les BR et PLV poste par poste. Chaque acte médical est répertorié dans la nomenclature de la Sécurité Sociale en étant associé à un « code acte ». Ce sont ces codes qui seront par la suite utilisés afin d'identifier chaque prestation.

### Le dentaire

Les Bases de Remboursement pour les soins dentaires sont disponibles dans le Journal Officiel n°0195 du 25 août 2018, texte n°5. Cette table ne demande aucun traitement particulier, elle est composée de dix colonnes. Les colonnes importantes pour cette étude sont :

- Le code acte figurant à la Classification Commune des Actes Médicaux (CCAM).
- Le libellé, qui permet de savoir à quel acte le code CCAM correspond.
- Les nouveaux tarifs applicables entre 2019 et 2023.

Les HLF<sup>1</sup> ont été publiés dans le même texte du Journal Officiel.

### L'auditif

La nomenclature et les Bases de Remboursement pour les prothèses auditives ont été publiées au Journal Officiel n°0275 du 28 novembre 2018, texte n°126. Tout comme pour le dentaire, cette table ne nécessite pas de traitement particulier pour pouvoir l'utiliser dans l'étude. Les PLV ont été ajoutés à la main en suivant les informations publiées par le gouvernement.

---

1. principe équivalent aux PLV

---

## L'optique

La nomenclature, les Bases de Remboursement et les Prix Limites de Vente pour le poste optique ont été publiés au Journal Officiel n°0090 du 16 avril 2019, texte n°19. La nomenclature ayant été modifiée de manière importante, notamment par la suppression de la distinction des adultes et des enfants, il est tout d'abord nécessaire de créer une correspondance entre les codes LPP (Liste des Produits et Prestations) utilisés avant la réforme et ceux qui doivent être utilisés à compter de 2020. Cette correspondance servira à associer les nouveaux codes à la table de prestations 2018.

## 2 Agrégation et identification des prestations

D'autre part, la base de prestations contenant plus de huit millions de lignes, il est indispensable de regrouper les lignes de prestations identiques (sans prendre en compte les informations liées à l'assuré mais seulement celles liées à l'acte en lui-même). Deux lignes de prestations sont donc considérées comme identiques si elles ont le même code acte, la même base de remboursement et le même détail d'acte (colonne qui donne le plus de précisions possible sur l'acte remboursé). Une colonne de comptage du nombre de lignes identiques est ajoutée à la table afin de pouvoir retrouver les informations unitaires par la suite. Cette étape est effectuée sous SAS.

Une fois la table de prestations réduite, elle est importée sur Excel afin d'identifier chaque prestation. Pour les trois postes, la méthode de classification est identique : la base de remboursement permet d'identifier l'acte dans les tables de BR et PLV créées dans le paragraphe précédent puis une clé est associée à la ligne. Trois cas de figure peuvent être rencontrés :

- La BR de la prestation n'apparaît qu'une seule fois dans toute la table de BR. Dans ce cas, la clé associée est le code acte correspondant à cette unique BR.
- La BR de la prestation apparaît plusieurs fois dans la table de BR. Dans ce cas, le choix de la clé se fait manuellement en fonction de l'intitulé de l'acte et du détail.
- La BR de la prestation n'apparaît pas dans la table de BR. Le choix de la clé se fait alors également à la main sur la base de BR dont le montant est similaire (il arrive en effet que le montant de la BR apparaissant dans la base de données soit différent de quelques centimes avec un montant existant).

La table finalement créée est nommée **Table des correspondances**.

En optique, il arrive que la nouvelle nomenclature précise l'ancienne c'est-à-dire qu'un code acte avant la réforme peut correspondre à plusieurs codes actes après celle-ci. Par exemple, un verre simple avant la réforme correspondait à un code acte alors qu'après la réforme, différents codes sont associés aux différents niveaux de dioptrie du verre. Dans ce cas, la ligne de prestation initiale est dupliquée autant de fois qu'il y a de nouveaux codes et des poids sont appliqués afin de ne pas donner plus d'importance à la ligne dupliquée qu'aux autres lignes de prestations. Le choix est fait d'utiliser les poids observés par le réseau Carte Blanche [5]. Il s'agit d'une hypothèse importante, qui aura très certainement des conséquences sur les résultats de l'étude. En effet, les répartitions peuvent fortement varier d'un expert à l'autre.

		Niveau 1	Niveau 2	Niveau 3	Niveaux 4 & 5	
Unifocal	Sphérique	7%	2%	0%	0%	
	Spérocyllindrique	CYL 0-4	64%	21%	5%	1%
		CYL > 4	0%	0%	0%	0%
Progressif	Sphérique	3%	1%	0%	5%	
	Sphérique	CYL 0-4	76%	17%	2%	90%
		CYL > 4	0%	0%	0%	5%

	Verres simples
	Verres complexes
	Verres très complexes

FIGURE 1.1 – Répartition des prises en charges du réseau Carte Blanche en fonction du type de verre et du niveau de correction visuelle

Pour obtenir les répartitions de la Figure 1.1, le réseau Carte Blanche a étudié les proportions des remboursements en fonction du type de verre dans l'ensemble de ses prises en charge. Les niveaux sont ceux des corrections visuelles (mesurées en dioptrie). Plus le niveau est élevé, plus la correction est importante.

Par ailleurs, il est possible qu'un même acte se trouve dans différents paniers de soins, comme en dentaire ou en optique ; ou encore qu'il soit nécessaire de dupliquer toutes les lignes de prestations, comme pour les prothèses auditives, puisque les prestations peuvent systématiquement appartenir aux deux paniers. Dans ce cas, afin de projeter les prestations en as-if, la ligne de prestation est elle aussi dupliquée et des poids sont attribués aux lignes. Encore une fois, le choix de ces poids fait l'objet d'hypothèses importantes dont les conséquences sur les résultats de l'étude ne pourront être oubliées.

Pour les poids des différents paniers dans les postes dentaire et optique, des sources officielles ont publié des projections des répartitions attendues. Ainsi, les poids appliqués en dentaire pour les différents paniers sont ceux publiés par le gouvernement [18] et ceux appliqués en optique sont ceux parus dans une étude réalisée pour le Groupement des Industriels et Fabricants de l'Optique (ou GIFO) [2].

Le choix de la répartition des paniers en audiologie est plus compliqué : les chiffres varient énormément selon les organismes et les études. Plusieurs faits sont donc pris en compte. Tout d'abord, actuellement près de deux tiers des personnes ayant besoin d'un appareil auditif y ont renoncé. De plus, dès janvier 2020, près de 30% [24] des devis effectués pour des prothèses auditives appartiennent déjà au panier « 100% santé », alors que le remboursement à 100% du reste à charge n'est pas encore effectif. Ainsi, en tenant compte de ces deux observations, l'hypothèse que deux tiers des actes d'audiologie appartiendront au panier « 100% santé » une fois la réforme entièrement mise en place est faite.

	Panier « 100% santé »	Panier maîtrisé	Panier libre
Optique	20%	-	80%
Dentaire	46%	25%	29%
Audiologie	67%	-	33%

### 3 Création de la Table finale de prestations

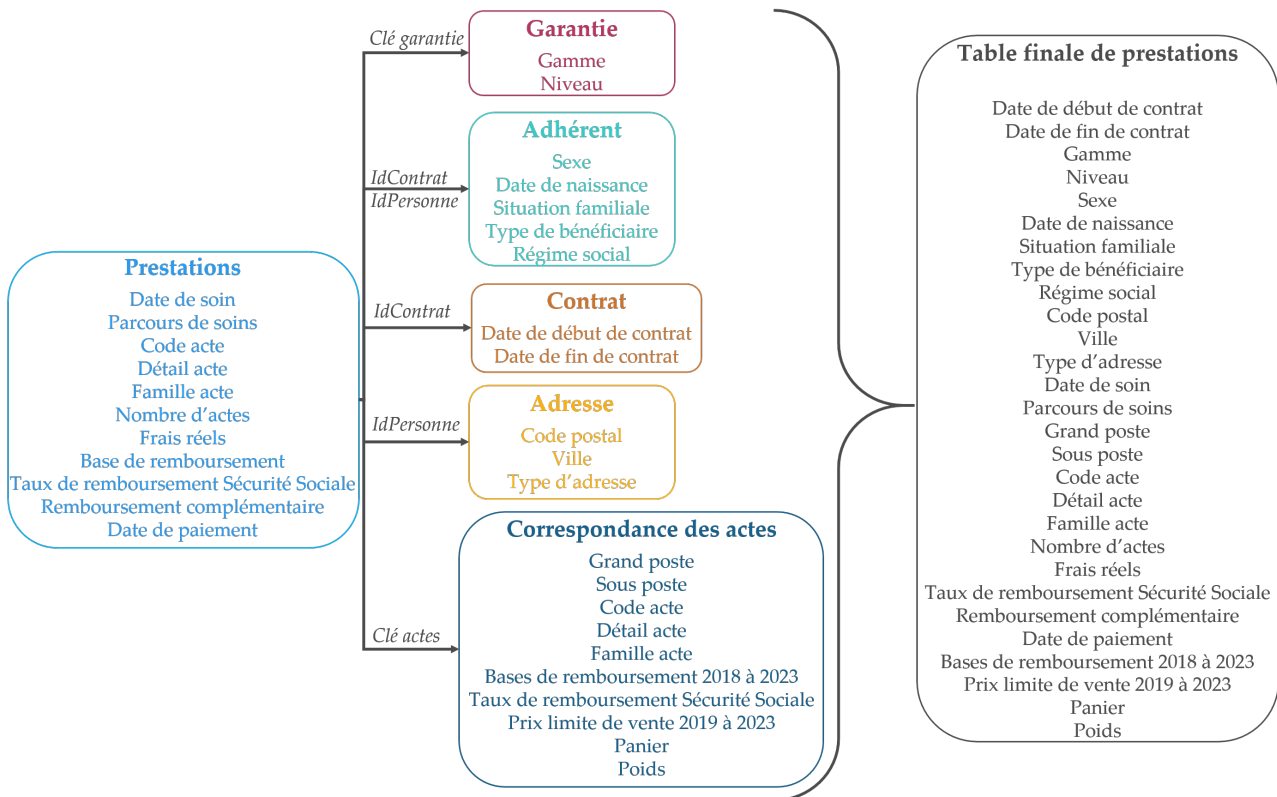


FIGURE 1.2 – Représentation de la création de la Table finale de prestations sous SAS et des clés utilisées pour lier les différentes tables

Seules les prestations liées aux trois postes de la réforme « 100% santé » seront étudiées en détails lors de l'analyse des impacts de la réforme. Ainsi, il est nécessaire de créer une table de prestations regroupant les prestations des postes optique, dentaire et audiology. À cette table doivent ensuite être ajoutées les informations des bases Contrat, Adhérent, Garantie, Adresse et de la table des correspondances (en utilisant SAS). La Figure 1.2 représente les clés utilisées pour lier les différentes tables, ainsi que les variables de chaque table.

---

L'ensemble des étapes effectuées dans les deux prochains chapitres mènent à la création d'un outil Excel. Cet outil prend en entrée une table de prestations (telle que la table finale de prestations obtenue sous SAS précédemment) ainsi qu'une table des expositions (déclinées en fonction de l'âge ainsi que de la gamme et du niveau de garantie). Les sorties et affichages de l'outil seront présentés au fur et à mesure de l'étude.

## 2 | IMPACTS SUR LES COÛTS DES TROIS POSTES CONCERNÉS

### 1 Hypothèses

La projection des prestations versées par l'assureur en 2019 s'effectue en « As-if » à partir des prestations 2018. C'est-à-dire que les prestations de 2018 sont considérées comme la base de travail sur laquelle des paramètres sont modifiés afin d'étudier les montants « comme si » ils avaient été versés en 2019, avec les nouvelles règles imposées par la réforme. En d'autres termes, pour une prestation versée en 2018, les BR sont modifiées, des PLV sont appliqués et les garanties sont ajustées afin de recalculer la prestation comme si l'acte médical et le remboursement avaient été effectués en 2019. Cela permet de mesurer le seul impact de la réforme « 100% santé » sur les prestations versées par l'assureur. Ce principe est également appliqué aux projections des prestations de 2020 à 2023, avec pour chaque année, les BR, PLV et garanties associés.

Ainsi, avant de procéder à la mise en « As-if » des prestations 2018, il est nécessaire que le tableau de garanties utilisé soit mis en conformité. En effet, comme cela a été évoqué précédemment, certains remboursements sont plafonnés par la réforme (comme par exemple le remboursement des montures optiques), d'autres sont soumis à la nécessité d'être supérieurs à un seuil dans le cadre des contrats responsables et les paniers de soins doivent apparaître dans les garanties.

Par ailleurs, il arrive régulièrement que les assureurs modifient leurs tableaux de garanties en fonction de nombreux autres paramètres (rentabilité, ajustement sur le marché, etc.), sans rapport direct avec la réforme étudiée. Étant donné que seul l'impact de la réforme « 100% santé » est l'objet de cette partie, l'hypothèse qu'aucune autre modification de garantie ne sera effectuée est faite, car prendre en compte ces changements biaiserait le calcul de l'impact pur de la réforme.

De plus, les prix de vente et honoraires de facturation seront très certainement impactés par cette réforme : les BR augmentant, il paraît cohérent de considérer que les frais réels augmenteront aussi (dans la limite des PLV appliqués). Il est cependant compliqué d'anticiper précisément cette hausse. Pour l'étude, une hausse proportionnelle à celle des bases de remboursements est choisie. Il s'agit d'une hypothèse prudente importante qu'il faudra nécessairement prendre en considération lors de l'analyse des résultats.

Enfin, l'utilisation des PLV nécessite également certaines hypothèses. Plusieurs méthodes sont possibles. Tout d'abord, il est possible de remplacer le montant des frais réels par le PLV, qu'il soit supérieur ou inférieur, ce qui revient à considérer que quelque soit le montant

---

consommé par les assurés en 2018, ils choisiront tous de payer le prix maximum autorisé une fois la réforme « 100% santé » mise en place. Cependant, cela implique de modifier considérablement les prestations. En effet, en réalité, tous les assurés ne choisiront pas de consommer le montant maximum autorisé, certains continueront de choisir des actes dont le prix est inférieur au PLV. Ainsi, une autre méthode est celle qui consiste à appliquer un minimum entre le PLV et les frais réels (augmentés comme décrit dans le paragraphe précédent). Cela permet d'appliquer le plafond qui doit l'être dans le cadre des contrats responsables ou des différents paniers de soins, sans augmenter démesurément les frais réels s'ils étaient initialement plus faibles que ledit plafond. Pour rappel, comme la Figure 1.2 le montre, pour chaque ligne de prestation, le panier est connu, la BR et le PLV aussi, cela permet d'associer à la prestation les bons paramètres.

## 2 Projections

Une fois tous les paramètres mis en place, la dernière étape de cette partie de l'étude consiste à projeter les remboursements complémentaires de 2019 à 2023. La projection est effectuée sous SAS et le résultat sera inclus dans l'outil Excel afin d'en extraire une analyse approfondie.

Le remboursement à 100% du reste à charge dans les paniers « 100% santé » doit être pris en compte. Pour l'année 2019, ce remboursement à 100% n'est pas garanti, la formule appliquée pour calculer les remboursements complémentaires (ou RC) est donc la même pour toutes les prestations et peut s'écrire comme cela :

$$RC\ 2019 = \min\left(\text{Frais réels ; } \frac{\text{Remboursement maximal autorisé}}{\text{yc Sécurité Sociale}}\right) - \frac{\text{Remboursement}}{\text{Sécurité Sociale}}$$

Avec le remboursement maximal autorisé y compris Sécurité Sociale égal à la somme du remboursement Sécurité Sociale et de la garantie assureur accordée pour l'acte étudié.

En revanche, pour les remboursements de 2020 il est nécessaire de différencier deux cas de figure :

- La prestation n'est pas remboursée à 100%, quelque soit le panier. Dans ce cas, la formule utilisée est la même que celle utilisée pour 2019.
- La prestation est une prestation d'optique, une couronne dentaire ou un bridge, auquel cas le reste à charge pour l'assuré doit être nul. Le RC vaut alors la différence entre la totalité des frais réels et le remboursement de la Sécurité Sociale.

Pour les années suivantes, les formules de 2020 restent valables, en faisant bien attention que dès 2021, tous les actes des paniers « 100% santé » des trois postes sont entièrement remboursés.

## 3 Résultats

Les projections obtenues sont synthétisées dans le tableau de la Figure 2.1 issu de l'outil Excel mis en place.

	2018	2019	2020	2021	2022	2023
<b>DENTAIRE</b>	5 092 k€	5 225 k€	5 719 k€	6 676 k€	6 709 k€	6 710 k€
<b>AUDIOLOGIE</b>	558 k€	571 k€	574 k€	1 005 k€	1 005 k€	1 005 k€
<b>OPTIQUE</b>	5 477 k€	5 477 k€	5 173 k€	5 173 k€	5 173 k€	5 173 k€
<b>AUTRES POSTES</b>	55 604 k€	55 604 k€	55 604 k€	55 604 k€	55 604 k€	55 604 k€
<b>TOTAL</b>	<b>66 730 k€</b>	<b>66 876 k€</b>	<b>67 070 k€</b>	<b>68 458 k€</b>	<b>68 491 k€</b>	<b>68 491 k€</b>

FIGURE 2.1 – Projection des remboursements complémentaires

Afin d’avoir une meilleure lisibilité des résultats, il est possible de calculer les impacts de la réforme sur les prestations à partir de ces projections. L’impact est considéré comme étant la hausse (ou la baisse) des montants observés entre deux années. La formule utilisée est donc la suivante :

$$\text{Impact entre l'année } i \text{ et l'année } j = \frac{\text{Prestations de l'année } j}{\text{Prestations de l'année } i} - 1, \quad \text{avec } i < j$$

Tout d’abord, les impacts année après année sont calculés, afin de comprendre chaque année comment les prestations ont évolué suite à la réforme. Puis l’impact total (entre 2018 et 2023) est calculé afin de visualiser, à terme, ce que la réforme « 100% santé » change du point de vue des prestations.

Les impacts ainsi calculés sont présentés dans le tableau de la Figure 2.2, également issu de l’outil Excel.

	Entre 2018 & 2019	Entre 2019 & 2020	Entre 2020 & 2021	Entre 2021 & 2022	Entre 2022 & 2023	Entre 2018 & 2023
<b>DENTAIRE</b>	2,60%	9,47%	16,73%	0,49%	0,01%	<b>31,77%</b>
<b>AUDIOLOGIE</b>	2,34%	0,61%	74,98%	0,00%	0,00%	<b>80,16%</b>
<b>OPTIQUE</b>	0,00%	-5,55%	0,00%	0,00%	0,00%	<b>-5,55%</b>
<b>AUTRES POSTES</b>	0,00%	0,00%	0,00%	0,00%	0,00%	<b>0,00%</b>
<b>TOTAL</b>	<b>0,22%</b>	<b>0,29%</b>	<b>2,07%</b>	<b>0,05%</b>	<b>0,00%</b>	<b>2,64%</b>

FIGURE 2.2 – Impacts de la réforme sur les remboursements complémentaires

Il reste important de rappeler une hypothèse prudente prise en début d’étude : le prix d’un acte médical augmente dans les mêmes proportions que la BR qui lui est associée, à moins qu’un PLV n’existe et soit inférieur à ce prix. Ainsi, pour tous les actes du panier « 100% santé », puisque l’assureur se doit de rembourser la totalité du reste à charge et que ce dernier augmente suite à l’hypothèse émise, alors la prestation de l’assureur augmente également. D’autre part, comme cela a été expliqué dans l’étude préliminaire de la base de données, 81% des assurés ont fait le choix d’une garantie entrée de gamme, pour laquelle les montants remboursés sont très faibles, ce qui laissait jusqu’alors un reste à charge important à l’assuré. Cela explique en partie l’impact très élevé obtenu dans l’étude.

L’impact global de la réforme tous postes confondus est de +2,64% sur les prestations complémentaires. Néanmoins, les résultats obtenus sont très différents selon le poste observé, il est donc important de les étudier indépendamment les uns des autres :



— Les postes n'étant pas réformés présentent un impact nul de 2019 à 2023. Cela est cohérent puisque la mise en As-if ne s'effectue qu'autour des modifications induites par la réforme, et non autour de la déviance intrinsèque du portefeuille. Les prestations 2018 pour ces postes n'ont donc pas été modifiées dans l'étude.

— Les prestations dentaires augmentent continuellement de 2019 à 2023.

En 2019, l'impact est dû à la modification des BR. Pour certains actes, la BR augmente, ce qui augmente les frais réels et donc le remboursement complémentaire. C'est notamment le cas pour les soins dentaires, dont la BR augmente continuellement entre 2018 et 2022 (cf. Figure 2.3). Pour d'autres actes, la BR diminue, ce qui engendre un désengagement de la Sécurité Sociale et donc une hausse du remboursement complémentaire. C'est le cas des actes Inlay-core (cf. Figure 2.4).

Panier de soins	Niveau de garantie	Année de remboursement	Base de remboursement	Remboursement Sécurité Sociale	Prix limite de vente	Frais réels	Remboursement complémentaire
Libre	Haut de gamme	2018	40,97 €	28,68 €	-	40,97 €	12,29 €
		2019	53,00 €	37,10 €	-	53,00 €	15,90 €
		2020	60,95 €	42,67 €	-	60,95 €	18,29 €
		2021	63,60 €	44,52 €	-	63,60 €	19,08 €
		2022 & 2023	65,50 €	45,85 €	-	65,50 €	19,65 €

FIGURE 2.3 – Exemple de prestations pour des soins dentaires

Panier de soins	Niveau de garantie	Année de remboursement	Base de remboursement	Remboursement Sécurité Sociale	Prix limite de vente	Frais réels	Remboursement complémentaire
100% santé	Milieu de gamme	2018	122,56 €	85,79 €	-	200,00 €	114,21 €
		2019	90,00 €	63,00 €	230,00 €	200,00 €	130,00 €
		2020 à 2023	90,00 €	63,00 €	175,00 €	175,00 €	112,00 €

FIGURE 2.4 – Exemple de prestations pour un Inlay-core

La hausse entre 2019 et 2020 est plus importante : la prise en charge de la totalité du reste à charge pour les couronnes et les bridges est effective, ce qui augmente nécessairement les prestations pour le panier « 100% santé ». Pour les paniers Modéré et Libre, la BR 2020 est supérieure à celle de 2019 ce qui engendre une hausse des frais réels et donc du remboursement complémentaire. La Figure 2.5 illustre ces phénomènes.

Panier de soins	Niveau de garantie	Année de remboursement	Base de remboursement	Remboursement Sécurité Sociale	Prix limite de vente	Frais réels	Remboursement complémentaire
100% santé	Entrée de gamme	2018 & 2019	107,50 €	71,75 €	-	467,50 €	32,25 €
		2020 à 2023	120,00 €	84,00 €	500,00 €	500,00 €	416,00 €
Modéré	Entrée de gamme	2018 & 2019	107,50 €	71,75 €	-	467,50 €	32,25 €
		2020 à 2023	120,00 €	84,00 €	-	521,86 €	36,00 €
Libre	Entrée de gamme	2018 à 2021	107,50 €	71,75 €	-	467,50 €	32,25 €
		2022 & 2023	120,00 €	84,00 €	-	521,86 €	36,00 €

FIGURE 2.5 – Exemple de prestations pour une couronne en céramique

Par la suite, en 2021, le reste à charge nul est garanti pour la totalité des actes prévus par la réforme « 100% santé », ce qui explique la très forte hausse. L'exemple de la

Figure 2.6 permet de visualiser ce cas de figure.

Panier de soins	Niveau de garantie	Année de remboursement	Base de remboursement	Remboursement Sécurité Sociale	Prix limite de vente	Frais réels	Remboursement complémentaire
100% santé	Milieu de gamme	2018 à 2020	182,75 €	127,93 €	-	1 600,00 €	328,95 €
		2021 à 2023	182,75 €	127,93 €	1100,00 €	1 100,00 €	972,08 €

FIGURE 2.6 – Exemple de prestations pour une prothèse dentaire amovible définitive en résine

Enfin, en 2022 et 2023, quelques BR continuent d'augmenter, ce qui explique le faible impact positif observé.

Au global, les prestations dentaires augmentent de 32% entre 2018 et 2023, cette hausse n'est pas surprenante au vu des PLV qui restent relativement élevés et de la hausse des BR.

- Les prestations audiolologiques subissent une augmentation importante dès 2019. Cela est dû à l'augmentation conséquente des BR chez les enfants comme chez les adultes. En revanche, en 2020, seuls les actes pour les adultes voient leur BR augmenter très légèrement. La hausse de prestations est donc beaucoup plus faible. Puis, en 2021, le remboursement en totalité du reste à charge doit être assuré pour les paniers « 100% santé », ce qui implique une très forte augmentation. Enfin, après 2021, les paramètres restent inchangés, les impacts de 2022 et de 2023 sont donc nuls. Les Figures 2.7 et 2.8 présentent un exemple de remboursement d'une prothèse auditive.

Panier de soins	Niveau de garantie	Année de remboursement	Base de remboursement	Remboursement Sécurité Sociale	Prix limite de vente	Frais réels	Remboursement complémentaire
100% santé	Entrée de gamme	2018	199,71 €	119,83 €	-	490,00 €	370,17 €
		2019	300,00 €	180,00 €	1300,00 €	736,07 €	500,00 €
		2020	350,00 €	210,00 €	1100,00 €	858,75 €	500,00 €
		2021 à 2023	400,00 €	240,00 €	950,00 €	950,00 €	710,00 €
Libre	Entrée de gamme	2018	199,71 €	119,83 €	-	490,00 €	370,17 €
		2019	300,00 €	180,00 €	-	736,07 €	500,00 €
		2020	350,00 €	210,00 €	-	858,75 €	500,00 €
		2021 à 2023	400,00 €	240,00 €	1700,00 €	981,42 €	500,00 €

FIGURE 2.7 – Exemple de prestations pour une prothèse auditive

Panier de soins	Niveau de garantie	Année de remboursement	Base de remboursement	Remboursement Sécurité Sociale	Prix limite de vente	Frais réels	Remboursement complémentaire
100% santé	Milieu de gamme	2018	199,71 €	119,83 €	-	1 250,00 €	400,00 €
		2019	300,00 €	180,00 €	1300,00 €	1 300,00 €	400,00 €
		2020	350,00 €	210,00 €	1100,00 €	1 100,00 €	400,00 €
		2021 à 2023	400,00 €	240,00 €	950,00 €	950,00 €	710,00 €
Libre	Milieu de gamme	2018	199,71 €	119,83 €	-	1 250,00 €	400,00 €
		2019	300,00 €	180,00 €	-	1 877,72 €	400,00 €
		2020	350,00 €	210,00 €	-	2 190,68 €	400,00 €
		2021 à 2023	400,00 €	240,00 €	1700,00 €	1 700,00 €	400,00 €

FIGURE 2.8 – Exemple de prestations pour une prothèse auditive

L'audiologie est le poste qui enregistre la plus forte hausse globale : +80% entre 2018 et 2023. Cela est cohérent, tout d'abord parce que le gouvernement a fortement augmenté les BR au travers de la réforme « 100% santé ». Mais aussi parce que les PLV ne sont pas particulièrement restrictifs, du moins vis-à-vis du portefeuille étudié : la majorité des appareils auditifs présentaient un prix de vente inférieur à ces PLV en 2018. Les résultats d'une étude sur un autre portefeuille auraient donc pu être différents.

- Les prestations optiques font exception parmi les trois postes observés : l'impact de la réforme est au global négatif (-5,6%). Le choc s'applique en une seule fois, en 2020. En effet, la réforme ne prévoit pas de mise en place progressive pour ce poste, tous les changements (ainsi que le reste à charge nul pour le panier « 100% santé ») sont effectifs en 2020. Alors que le but de la réforme reste de permettre aux assurés d'avoir une meilleure couverture complémentaire, l'étude prédit une réduction des remboursements complémentaires.

Cependant en optique, bien que le but final reste le même, afin d'alléger la contribution de la Sécurité Sociale et des complémentaires santé, la réforme prévoit un désengagement presque total des actes du panier libre. En effet, pour ce panier, la BR plonge à 0,05€ pour une monture et pour les verres, avec des PLV qui réduisent fortement les frais réels de l'assuré. Ainsi, les assurés ayant choisi le panier libre devront prendre en charge en quasi totalité leur dépense, alors que ceux ayant opté pour le panier « 100% santé » se verront remboursés à 100%. La Figure 2.9 propose plusieurs exemples de prestations qui illustrent les différents cas de figure.

	Panier de soins	Niveau de garantie	Année de remboursement	Base de remboursement	Remboursement Sécurité Sociale	Prix limite de vente	Frais réels	Remboursement complémentaire
Verre	100% santé	Entrée de gamme	2018	10,37 €	6,22 €	-	189,00 €	50,00 €
			2020	39,00 €	23,4 €	130 €	130,00 €	106,60 €
Verre	Libre	Entrée de gamme	2018	10,37 €	6,22 €	-	189,00 €	50,00 €
			2020	0,05 €	0,03 €	130 €	130,00 €	50,00 €
Verre	Libre	Entrée de gamme	2018	10,37 €	6,22 €	-	304,00 €	4,15 €
			2020	0,05 €	0,03 €	130 €	130,00 €	0,02 €
Monture	Libre	Entrée de gamme	2018	2,84 €	1,70 €	-	415,14 €	60,00 €
			2020	0,05 €	0,03 €	100 €	100,00 €	60,00 €

FIGURE 2.9 – Exemples de prestations optique

## 4 Conséquences pour l'assureur

Outre le calcul de la valeur de l'impact brut de la réforme « 100% santé », l'intérêt de cette étude est de mieux appréhender les retombées de la réforme sur le portefeuille, notamment en fonction du niveau de garantie de l'assuré. En effet, la réforme a pour but d'améliorer les remboursements pour les assurés ayant une couverture entrée de gamme, tout en réduisant certaines prestations pour les assurés avec un niveau de garantie élevé (comme les montures optiques).

Tout d'abord, il est intéressant d'étudier des statistiques sur le portefeuille après l'application de la réforme « 100% santé », et de les comparer avec celles de la Partie I, section 4.

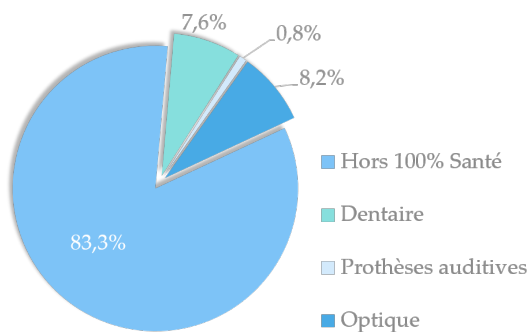


FIGURE 2.10 – Part des remboursements complémentaires par poste avant la réforme « 100% santé »

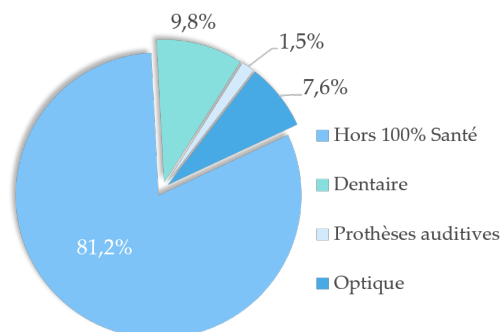


FIGURE 2.11 – Part des remboursements complémentaires par poste après la réforme « 100% santé »

Comme le montre la Figure 2.11, la répartition des remboursements complémentaires est légèrement modifiée par la mise en place de la réforme. Le poids des postes dentaire et prothèse auditive augmente, ce qui s'explique par l'impact positif obtenu lors de la projection de prestations de l'assureur. D'un autre côté, celui du poste optique diminue, étant donné que l'impact pour ce poste est négatif. Au total, le poids des postes réformés dans la totalité des prestations du portefeuille augmente de 2,1%.

Par ailleurs, l'étude d'impacts selon le niveau de garantie permet de mieux appréhender la réforme « 100% santé » et ses objectifs. Ainsi, la Figure 2.12 permet de visualiser l'évolution des montants de prestations selon le poste de l'acte et le niveau de garantie de l'assuré. Plus précisément, il s'agit d'un montant de prestations moyen calculé sur la base du nombre d'assurés ayant consommé au moins un acte médical appartenant au poste étudié :

$$\text{Montant moyen de prestations pour un poste et un niveau de garantie} = \frac{\text{Montant total de prestations pour ce poste et ce niveau de garantie}}{\text{Nombre d'assurés ayant souscrit à ce niveau de garantie et ayant consommé au moins un acte dans ce poste}}$$

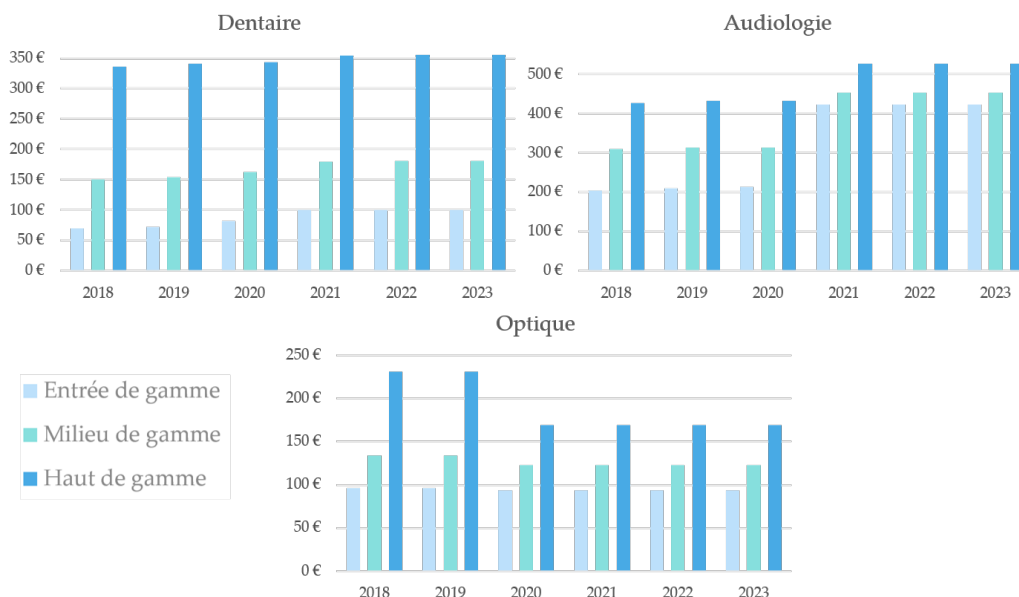


FIGURE 2.12 – Évolution des montants de prestations moyens par niveau de garantie et par poste

---

Les tableaux des Figures 2.1 et 2.2 ont déjà permis d’analyser le calendrier de la mise en place progressive de la réforme « 100% santé ». Ce calendrier apparaît aussi clairement sur les graphiques de la Figure 2.12 : l’application en deux étapes en dentaire avec un reste à charge nul assuré pour les couronnes et le bridges en 2020 et pour les autres actes en 2021 s’observe tout particulièrement sur l’entrée de gamme ; l’application de la totalité des mesures de la réforme en optique en 2020 est spécialement visible sur le haut de gamme ; et le reste à charge nul assuré en 2021 pour les prothèses auditives impacte visiblement tous les niveaux de garantie.

En plus de permettre de visualiser les conclusions émises suite aux résultats de l’étude, la Figure 2.12 permet également d’examiner les montants de prestations selon le niveau de garantie. Il apparaît donc que les conséquences de la réforme sont très différentes selon le niveau de garantie de l’assuré en audiologie et en optique, alors qu’en dentaire la tendance des trois courbes est relativement identique malgré une augmentation légèrement plus forte à mesure que la garantie diminue. L’augmentation des prestations en dentaire a déjà été expliquée dans l’analyse de la Figure 2.2.

En audiologie, la légère augmentation observée entre 2018 et 2020, due à l’augmentation progressive des BR, est très similaire pour les trois niveaux de garantie, ce qui est cohérent puisque la BR ne dépend pas de la garantie de l’assuré. Puis, la très forte augmentation en 2021, qui s’explique par le remboursement de la totalité du reste à charge pour le panier « 100% santé », est bien plus prononcée pour les assurés ayant opté pour une garantie entrée de gamme que pour ceux ayant opté pour des garanties supérieures. En effet, à terme, le remboursement complémentaire moyen pour l’entrée de gamme (qui double entre 2018 et 2023) est presque égal à celui du milieu de gamme. Les montants de prestations des garanties milieu et haut de gamme augmentent également, ce qui s’explique par le fait qu’avant la réforme « 100% santé », le reste à charge restait très élevé pour des prothèses auditives quelque soit le niveau de garantie (cf. Figure 3.14). En 2023, les montants de prestations sont beaucoup plus élevés qu’en 2018, et surtout plus proches les uns des autres : les inégalités en matière de remboursement complémentaire sont très visiblement réduites pour le poste audiologie.

En optique, la tendance des montants de prestations au cours des années diffère également selon les niveaux de garantie. En effet, les prestations pour les assurés ayant opté pour une garantie entrée de gamme ne diminuent que très peu : la très grande majorité des actes appartenant au panier « 100% santé », ils sont remboursés en totalité à partir de 2020. Par ailleurs, les assurés ayant opté pour une garantie milieu de gamme voient leurs remboursements complémentaires diminuer en 2020. Ce phénomène s’explique principalement par la mise en place de PLV qui réduisent les frais réels ou de plafonds imposés par le contrat responsable. Pour la même raison, les assurés ayant opté pour une garantie haut de gamme voient leur remboursement complémentaire diminuer de façon plus importante : les frais réels étaient initialement plus élevés, le plafonnement aux PLV a donc un impact plus fort sur ces prestations. Néanmoins, une diminution des remboursements complémentaires n’implique pas nécessairement une hausse du reste à charge des assurés, c’est d’ailleurs l’objet de l’étude de la section suivante.

## 5 Conséquences pour l'assuré

Le reste à charge suite à un soin médical est nécessairement lié au remboursement complémentaire, mais aussi au remboursement de la Sécurité Sociale. Il est donc pertinent à ce stade de l'étude d'observer les conséquences de la réforme « 100% santé » sur le montant des restes à charge.

	2018	2019	2020	2021	2022	2023
<b>DENTAIRE</b>	50,10 €	46,69 €	41,59 €	31,05 €	31,10 €	31,10 €
<b>AUDITIF</b>	544,68 €	622,48 €	598,58 €	238,44 €	238,44 €	238,44 €
<b>OPTIQUE</b>	66,11 €	66,06 €	24,65 €	24,65 €	24,65 €	24,65 €

FIGURE 2.13 – Projection des restes à charge moyens

	Entre 2018 & 2019	Entre 2019 & 2020	Entre 2020 & 2021	Entre 2021 & 2022	Entre 2022 & 2023	Entre 2018 & 2023
<b>DENTAIRE</b>	- 7%	- 11%	- 25%	0%	0%	- 38%
<b>AUDITIF</b>	14%	- 4%	- 60%	0%	0%	- 56%
<b>OPTIQUE</b>	0%	- 63%	0%	0%	0%	- 63%

FIGURE 2.14 – Impacts de la réforme sur le reste à charge

Les tableaux des Figures 2.13 et 2.14 présentent l'évolution de la moyenne des restes à charge entre 2018 et 2023, pour les trois postes touchés par la réforme « 100% santé » ainsi que les impacts de la réforme sur ces restes à charge. Les impacts sont calculés de la même manière que pour le tableau de la Figure 2.2. Il apparaît qu'à terme, pour les trois postes réformés, le reste à charge diminue.

L'observation de l'évolution du reste à charge en fonction du niveau de garantie pourra permettre de mieux comprendre les impacts de la Figure 2.14.



FIGURE 2.15 – Évolution de la moyenne des restes à charge par poste et niveau de garantie

## Dentaire

D'après la Figure 2.15, en moyenne le reste à charge diminue continuellement entre 2018 et 2021, et ce pour les trois niveaux de garantie. Tout comme pour l'analyse de l'évolution des remboursements complémentaires, il apparaît que la tendance de décroissance est très similaire pour les différents niveaux de garantie. En 2018, le reste à charge était quasiment identique pour les garanties milieu et haut de gamme. Un écart se crée à mesure que la réforme « 100% santé » se met en place et en 2023, le reste à charge moyen d'une prestation milieu de gamme est inférieur à celui d'une prestation haut de gamme.

Panier de soins	Niveau de garantie	Année de remboursement	Frais réels	Remboursement complémentaire	Remboursement Sécurité Sociale	Reste à charge
100% santé	Entrée de gamme	2018 & 2019	467,50 €	32,25 €	75,25 €	360,00 €
		2020 à 2023	500,00 €	416,00 €	84,00 €	0,00 €
Modéré	Entrée de gamme	2018 & 2019	467,50 €	32,25 €	75,25 €	360,00 €
		2020 à 2023	521,86 €	36,00 €	84,00 €	401,86 €
Libre	Entrée de gamme	2018 à 2021	467,50 €	32,25 €	75,25 €	360,00 €
		2022 & 2023	521,86 €	36,00 €	84,00 €	401,86 €

FIGURE 2.16 – Exemples de prestations pour une couronne en céramique

Panier de soins	Niveau de garantie	Année de remboursement	Frais réels	Remboursement complémentaire	Remboursement Sécurité Sociale	Reste à charge
100% santé	Milieu de gamme	2018 à 2020	1 600,00 €	328,95 €	127,93 €	1 143,12 €
		2021 à 2023	1 100,00 €	972,07 €	127,93 €	0,00 €

FIGURE 2.17 – Exemple de prestations pour une prothèse dentaire amovible définitive en résine

Les Figures 2.16 et 2.17 sont la poursuite des exemples présentés aux Figures 2.5 et 2.6. Ces exemples permettent de visualiser la différence de reste à charge entre les prestations de panier « 100% santé » et les autres : là où les actes du panier « 100% santé » présentent un reste à charge nul, le reste à charge des autres paniers augmente. Le même constat s'applique pour les niveaux de garantie milieu et haut de gamme, avec des différences de reste à charge au fil de l'application de la réforme plus faibles que pour l'entrée de gamme. La réforme « 100% santé » a donc l'effet prévu sur les prestations dentaires. Au global et pour tous les niveaux de garantie confondus, elle réduit le reste à charge moyen des assurés de 38%.

## Audiologie

Le fait que le reste à charge d'une prestation haut de gamme soit supérieur à celle d'une prestation milieu de gamme a été expliqué lors de l'analyse de la Figure 3.14. Une augmentation importante de la moyenne du reste à charge s'observe entre 2018 et 2019 (+14%). Cela s'explique par le fait que les frais réels augmentent en 2019 à cause de l'augmentation des BR et de l'hypothèse impliquant que les frais réels augmentent proportionnellement à la BR. Ce phénomène est observable sur la Figure 2.18, qui reprend la suite de l'exemple de la Figure 2.7.

Panier de soins	Niveau de garantie	Année de remboursement	Frais réels	Remboursement complémentaire	Remboursement Sécurité Sociale	Reste à charge
100% santé	Entrée de gamme	2018	490,00 €	370,17 €	119,83 €	0,00 €
		2019	736,07 €	500,00 €	180,00 €	56,07 €
		2020	858,75 €	500,00 €	210,00 €	148,75 €
		2021 à 2023	950,00 €	710,00 €	240,00 €	0,00 €
Libre	Entrée de gamme	2018	490,00 €	370,17 €	119,83 €	0,00 €
		2019	736,07 €	500,00 €	180,00 €	56,07 €
		2020	858,75 €	500,00 €	210,00 €	148,75 €
		2021 à 2023	981,42 €	500,00 €	240,00 €	241,42 €

FIGURE 2.18 – Exemples de prestations pour une prothèse auditive

À partir de 2020, bien que la BR continue d'augmenter, les PLV deviennent plus restrictifs ce qui entraîne une baisse des frais réels. La Figure 2.19 illustre ce propos (il s'agit de la suite de l'exemple présenté à la Figure 2.8).

Panier de soins	Niveau de garantie	Année de remboursement	Frais réels	Remboursement complémentaire	Remboursement Sécurité Sociale	Reste à charge
100% santé	Milieu de gamme	2018	1 250,00 €	400,00 €	119,83 €	730,17 €
		2019	1 300,00 €	400,00 €	180,00 €	720,00 €
		2020	1 100,00 €	400,00 €	210,00 €	490,00 €
		2021 à 2023	950,00 €	710,00 €	240,00 €	0,00 €
Libre	Milieu de gamme	2018	1 250,00 €	400,00 €	119,83 €	730,17 €
		2019	1 877,72 €	400,00 €	180,00 €	1 297,72 €
		2020	2 190,68 €	400,00 €	210,00 €	1 580,68 €
		2021 à 2023	1 700,00 €	400,00 €	240,00 €	1 060,00 €

FIGURE 2.19 – Exemples de prestations pour une prothèse auditive

Enfin, dès 2021, l'obligation pour l'assureur de rembourser la totalité du reste à charge pour les actes du panier « 100% santé » implique une diminution très importante du reste à charge moyen, et ce pour les trois niveaux de garantie (cf. Figures 2.18 et 2.19). Ceci est cohérent puisque comme cela a été évoqué précédemment, en audiologie, le renoncement aux soins est plus général que dans les autres postes de soins. De plus, le choix de la répartition des ventes dans les paniers de soins joue un rôle important dans l'analyse du reste à charge. En effet, le choix a été fait de considérer que suite à la réforme, deux tiers des prothèses auditives appartiendront au panier « 100% santé » et seront donc intégralement remboursés par la Sécurité Sociale et la complémentaire santé. Le reste à charge moyen s'en trouve donc fortement réduit.

## Optique

Le résultat pourrait sembler au premier abord contradictoire. En effet, puisque la Figure 2.12 montre une diminution des remboursements complémentaires, une augmentation du reste à charge serait attendue en 2020. De plus, l'hypothèse que seuls 20% des actes



font partie du panier « 100% santé » implique que seuls 20% des actes présentent un reste à charge nul en 2020, ce qui ne permet pas d'expliquer une diminution aussi importante de la moyenne globale des restes à charge. Il est donc nécessaire d'étudier des lignes de prestations afin de mieux comprendre la diminution générale du reste à charge. Les prestations étudiées appartiennent au panier libre puisque la diminution du reste à charge des prestations du panier « 100% santé » s'explique immédiatement. Différents cas de figure se présentent alors.

Le plus souvent, les frais réels diminuent du fait de la mise en place de PLV (que ce soit pour les verres ou les montures), cependant les remboursements complémentaires peuvent fortement diminuer selon la prestation.

Pour les montures, bien que la baisse de la BR et la présence du PLV devrait impliquer une baisse des remboursements complémentaires, il s'avère que les garanties de l'assureur étaient déjà majoritairement inférieures à ces PLV. Le remboursement de la Sécurité Sociale diminue, mais il était déjà très faible avant la réforme en comparaison avec les frais réels. Ainsi, dans le cas des montures, les frais réels diminuent, le remboursement complémentaire reste quasiment le même et la diminution du remboursement de la Sécurité Sociale est négligeable, ce qui implique une baisse du reste à charge. Ce phénomène est présenté dans la Figure 2.20

Panier de soins	Niveau de garantie	Année de remboursement	Frais réels	Remboursement complémentaire	Remboursement Sécurité Sociale	Reste à charge
Libre	Entrée de gamme	2018	415,14 €	60,00 €	1,99 €	353,44 €
		2020	100,00 €	60,00 €	0,03 €	39,97 €

FIGURE 2.20 – Exemple d'évolution de prestations pour une monture

D'autre part, pour les verres, le remboursement complémentaire peut diminuer suite à la mise en place de PLV et à la baisse des BR, ou rester identique si la garantie est inférieure au PLV. Malgré tout, les PLV impliquent également une baisse importante des frais réels de l'assuré qui sont parfois extrêmement élevés (et ne respectent pas le plafond du contrat responsable valable en 2018). Cette baisse est plus importante que celles des remboursements complémentaires et de la Sécurité Sociale, ce qui engendre une baisse du reste à charge (cf. exemples des Figures 2.21 et 2.22).

Panier de soins	Niveau de garantie	Année de remboursement	Frais réels	Remboursement complémentaire	Remboursement Sécurité Sociale	Reste à charge
Libre	Entrée de gamme	2018	304,00 €	4,15 €	7,26 €	293,63 €
		2020	130,00 €	0,02 €	0,03 €	129,95 €

FIGURE 2.21 – Exemple d'évolution de prestations pour un verre optique

Panier de soins	Niveau de garantie	Année de remboursement	Frais réels	Remboursement complémentaire	Remboursement Sécurité Sociale	Reste à charge
100% santé	Entrée de gamme	2018	189,00 €	50,00 €	7,26 €	132,78 €
		2020	130,00 €	106,60 €	27,30 €	0,00 €
Libre	Entrée de gamme	2018	189,00 €	50,00 €	7,26 €	132,78 €
		2020	130,00 €	50,00 €	0,03 €	79,97 €

FIGURE 2.22 – Exemples d'évolution de prestations pour un verre optique

Vraisemblablement, pour le poste optique, ce sont les praticiens qui encaissent les pertes dues à la réforme « 100% santé » liées à la limitation des prix de vente. Cependant, l'objectif de la réforme « 100% santé » de réduction de reste à charge est atteint.

Il va de soi que l'hypothèse sur le fait que tous les contrats soient responsables a un impact important sur les résultats obtenus : en 2018, certains frais réels sont très supérieurs aux plafonds responsables, ce qui engendre un reste à charge très important, qui réduit donc considérablement en 2020.

### En résumé

L'impact global de la réforme « 100% santé » une fois sa mise en place terminée en 2023 résulte en une hausse des prestations de 2,64%. C'est le poste audiologie qui enregistre la plus forte augmentation : 80,16%, alors que la projection pour le poste optique prévoit une réduction des remboursements complémentaires de 5,55%. Quant aux remboursements complémentaires pour le poste dentaire, ils augmentent de 31,77%.

De plus, pour les postes audiologie et dentaire, les augmentations de remboursements complémentaires sont bien plus importantes pour les assurés ayant une garantie entrée de gamme, bien qu'en audiologie l'augmentation est significative pour les trois niveaux de garantie. En optique à l'inverse, les prestations entrée et milieu de gamme sont peu modifiées, alors que celles du haut de gamme sont fortement réduites.

D'autre part, l'évolution des restes à charge n'est pas exactement l'inverse de celle des remboursements complémentaires. En effet, pour les postes dentaire et audiologie et quel que soit le niveau de garantie, le reste à charge diminue (très fortement en audiologie), ce qui s'explique principalement par l'augmentation des remboursements complémentaires et parfois également par le plafonnement des frais réels. En revanche en optique, bien que les remboursements complémentaires diminuent, les restes à charge diminuent également du fait du plafonnement des frais réels.

---

## 3 | IMPACT DE LA RÉFORME SUR LE COMPORTEMENT DES ASSURÉS

Après avoir mesuré l'impact pur de la réforme « 100% santé », un autre impact est à prendre en compte : celui du changement de comportement des assurés face à cette réforme.

En effet, la réforme a pour but de réduire le reste à charge des assurés afin de leur permettre un accès plus large aux actes dentaires, optiques et audiolinguistiques, mais aussi afin de réduire le renoncement aux soins dans ces postes. Ainsi, certains assurés qui n'ont pas consommé en 2018 mais qui en auraient eu besoin, devraient le faire au fur et à mesure de la mise en place de la réforme.

Ainsi, la réduction du renoncement aux soins impactera la fréquence de consommation des assurés, ce qui aura nécessairement des conséquences sur les prestations des assureurs qui viendront s'ajouter à celles précédemment décrites.

### 1 Hypothèses

Tout d'abord, seule l'augmentation de fréquence de consommation intrinsèque au portefeuille est étudiée : l'apparition potentielle de nouveaux assurés n'est pas prise en compte ici car le portefeuille étudié est en « run-off » (c'est-à-dire que les produits ne sont plus commercialisés). De plus, il est considéré que les assurés ne feront pas le choix de changer de gamme ni de niveau de couverture suite à l'application de la réforme « 100% santé ». En effet, bien que cette situation soit probable, elle est délicate à anticiper.

D'autre part, la majoration des fréquences de consommation n'est effectuée que sur les actes entrant dans le cadre de la réforme « 100% santé ». Cette hypothèse s'explique par le fondement même de l'étude : l'impact de la réforme sur la consommation des assurés. Le changement de comportement sur la consommation d'actes n'étant pas réformés ne sera donc pas pris en compte dans cette étude. Les actes du panier Modéré en dentaire ne sont pas non plus concernés par la déformation de fréquence qui sera appliquée.

Par ailleurs, les soins dentaires font partie du panier Libre. Cependant, la pose de prothèse dentaire est presque systématiquement accompagnée d'actes de soins dentaires. Ainsi, l'hypothèse que la fréquence de consommation de ces soins soit augmentée de la même façon que les actes dentaires du panier « 100% santé » est faite.

De plus, la communication sur la réforme « 100% santé » se fait progressivement et les premières années, tous les assurés ne seront pas au courant de son existence ou de son fonctionnement. Ainsi, le choc de la déformation de fréquence de consommation est pondéré par l'estimation de la quantité d'assurés effectivement au courant de la réforme. Pour l'année 2020, cette estimation est exprimée dans plusieurs études de marché et dépend évidemment de l'organisme qui les publie. Santéclair a mis en avant en juillet 2020 [25] que seules 36% des personnes interrogées disent connaître précisément la réforme « 100% santé ». C'est ce chiffre qui sera donc utilisé. Pour les années suivantes, les chiffres suivants sont choisis :

2020	2021	2022	2023
36%	75%	100%	100%

FIGURE 3.1 – Proportion d’assurés correctement informés sur la réforme « 100% santé »

Enfin, les taux de déformation de la consommation des assurés diffèrent selon le poste de soins ainsi que selon la gamme et le niveau de garantie de l’assuré. Trois scénarios sont alors envisagés afin d’étudier la sensibilité de l’impact de la réforme « 100% santé » suite à la déformation de fréquence de consommation. Les taux de déformation sont établis pour un scénario central et deux autres scénarios (l’un minorant et l’autre majorant) sont ensuite créés à partir du scénario central. Le choix de ces taux de déformation constitue l’hypothèse majeure de cette partie.

Ainsi, étant donné que la réforme « 100% santé » a pour but de réduire le renoncement aux soins pour les trois postes concernés, les données sur ce renoncement semblent être une base de départ solide. Pour rappel, les taux de renoncement aux soins ont été détaillés dans la Partie I, ils sont de 19% pour l’optique, 39% pour les prothèses dentaires, 34% pour les soins dentaires et 65% pour les prothèses auditives. Aussi, ces chiffres sont déclinés selon le niveau et la gamme de garantie. Par exemple, les personnes âgées étant les plus touchées par la nécessité de porter une prothèse auditive, leur taux de déformation de fréquence peut être considéré comme plus élevé que le taux général de renoncement aux soins pour les niveaux entrée et milieu de gamme. De plus, quelque soit le poste et la gamme de garantie, le taux de déformation appliqué pour l’entrée de gamme est supérieur à celui du milieu de gamme, lui-même supérieur à celui du haut de gamme. Cela s’explique par l’hypothèse que le renoncement aux soins est plus important lorsque la garantie est plus faible mais également par le fait que les remboursements complémentaires pour l’entrée de gamme augmentent fortement en dentaire et audiologie et restent stables en optique (cf. Figure 2.12). Les taux de déformation choisis pour le scénario central sont donc présentés à la Figure 3.2.

NIVEAU		OPTIQUE	AUDIO	PROTHESES DENTAIRES	SOINS DENTAIRES
SOLO	Entrée de gamme	25%	65%	54%	54%
	Milieu de gamme	19%	50%	27%	27%
	Haut de gamme	8%	10%	5%	5%
FAMILLE	Entrée de gamme	25%	70%	65%	65%
	Milieu de gamme	19%	65%	39%	39%
	Haut de gamme	8%	30%	18%	18%
SENIOR	Entrée de gamme	19%	73%	59%	59%
	Milieu de gamme	14%	68%	36%	36%
	Haut de gamme	6%	35%	9%	9%

FIGURE 3.2 – Taux de déformation pour le scénario central

Les taux de déformation des scénarios minorants et majorants sont calculés à partir de ceux du scénario central auxquels des chocs (positifs pour le scénario majorant, négatifs pour le scénario minorant) sont appliqués selon la gamme et le niveau de garantie. Les chocs choisis sont les suivants (par exemple, pour le scénario minorant, la déformation de fréquence entrée de gamme Solo pour l’optique devient  $25\% - 10\% = 15\%$ ) :

	Scénario Minorant	Scénario Majorant
<b>Entrée de gamme</b>	- 10%	30%
<b>Milieu de gamme</b>	- 20%	20%
<b>Haut de gamme</b>	- 30%	10%

FIGURE 3.3 – Chocs appliqués aux taux de déformation du scénario central pour obtenir les scénarios minorant et majorant

Il arrive par ailleurs que le taux de déformation du scénario minorant devienne négatif (comme par exemple pour l'entrée de gamme en optique, pour les seniors qui auraient un taux de déformation de  $-6\%$ ). Dans ce cas, le taux de déformation est fixé à  $0\%$ .

## 2 Mise en place des taux de déformation de la fréquence de consommation

La déformation de la fréquence de consommation est directement appliquée aux montants de prestations. C'est-à-dire que lorsque la fréquence de consommation d'un acte doit être modifiée, c'est en réalité le montant de prestations pour cet acte qui l'est au travers de la formule suivante :  $RC_{déformé} = RC_{initial} \times (1 + \text{taux de déformation totale})$ , où le taux de déformation totale est le produit entre le taux de déformation de la fréquence de consommation (Figures 3.2 et 3.3) et la proportion d'assurés correctement informés sur la réforme « 100% santé » (Figure 3.1). Cette formule se justifie par le fait que le RC initial est calculé par :  $RC_{initial} = \text{fréquence}_{initiale} \times \text{coût moyen}$ .

Ainsi, la proportion de remboursements de la complémentaire santé augmente comme si c'était la consommation des assurés qui augmentait.

## 3 Résultats

Les impacts de la réforme « 100% santé » sont alors plus importants que ceux observés précédemment (Figure 2.2) :

	Sans déformation	Scénario minorant	Scénario central	Scénario majorant
<b>DENTAIRE</b>	31,8%	60,2%	67%	90,5%
<b>AUDIOLOGIE</b>	80,2%	158,8%	177,5%	214,5%
<b>OPTIQUE</b>	- 5,6%	-4,1%	-2%	2,8%
<b>TOTAL</b>	2,6%	5,6%	6,7%	8,9%

FIGURE 3.4 – Impacts de la réforme sur les remboursements complémentaires incluant la déformation de fréquence de consommation des assurés

D'après les hypothèses émises, l'impact global de la réforme « 100% santé » que l'assureur aurait à supporter serait compris entre  $5,6\%$  et  $8,9\%$ . Les taux étant plus élevés en audiologie, l'impact de la déformation est d'autant plus fort pour ce poste, quel que soit le scénario.

Par ailleurs, le scénario majorant affiche un impact positif pour le poste optique, ce qui signifie que dans le cas extrême d'un changement de comportement très important chez les assurés, les prestations d'optique pourraient se voir augmenter de 2,8%.

La Figure 3.5 présente les différentes évolutions des remboursements complémentaires selon le scénario de déformation appliqué. Comme attendu, plus les taux de déformation sont élevés, plus les remboursements complémentaires le sont. Par ailleurs, ces trois graphiques mettent en avant le calendrier de la réforme « 100% santé ». En effet, les différents chocs précédemment observés pour les remboursements sans déformation de fréquence de consommation (Figure 2.1) sont exacerbés par l'application des dites déformations de fréquence. Par ailleurs, lorsque l'impact brut de la réforme est nul ou presque (entre 2021 et 2023 pour le dentaire et l'audiologie et entre 2020 et 2023 pour l'optique), l'application de déformations de fréquences implique une hausse annuelle des remboursements complémentaires non négligeable. Pour le scénario central, cette hausse est de l'ordre de 6% en dentaire, 5% en audiologie et 1% en optique. Pour le scénario majorant en optique, cette hausse est de plus de 2%, ce qui compense la baisse de 2020 due à l'application pure de la réforme et implique donc un impact global positif pour ce poste.

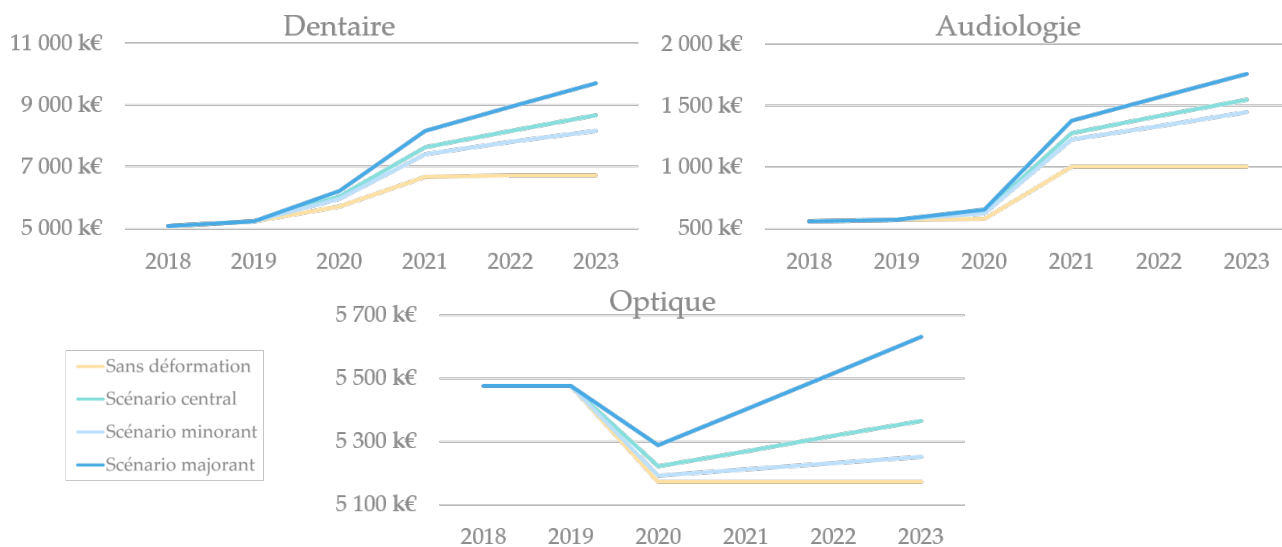


FIGURE 3.5 – Comparaison de l'évolution des remboursements complémentaires selon le scénario de déformation

## 4 Approche rentabilité assureur

Plutôt que de tenter de prédire les taux de déformation de la fréquence de consommation dans le but d'obtenir un impact pour la réforme « 100% santé », il est possible de faire le raisonnement inverse en considérant un choc maximal que l'assureur est prêt à encaisser suite à cette réforme et d'étudier les taux de déformation qui mènent à cet impact. Cette méthode permet d'obtenir des taux de déformation extrêmes au delà desquels les produits vendus par l'assureur devront obligatoirement être modifiés ou supprimés. La rentabilité d'un assureur est calculée à partir du ratio S/P (c'est-à-dire le ratio entre le montant des sinistres payés par l'assureur et le montant des primes payées par les assurés). L'impact maximal fixé par le client d'Addactis est une perte de 15 points sur le S/P en considérant qu'il n'y a pas d'augmentation des cotisations payées par les assurés.

Ainsi, les taux de déformation sont répartis selon les postes, les gammes et le niveau de garantie. L'ordre des taux reste le même que celui du renoncement aux soins, à savoir que les taux les plus importants sont appliqués à l'audiologie, puis aux prothèses dentaires suivies des soins dentaires et que l'optique présente les taux les plus faibles.

NIVEAU		OPTIQUE	AUDIO	PROTHESES DENTAIRES	SOINS DENTAIRES
SOLO	Entrée de gamme	105%	180%	170%	155%
	Milieu de gamme	70%	135%	105%	95%
	Haut de gamme	40%	60%	55%	50%
FAMILLE	Entrée de gamme	105%	185%	180%	175%
	Milieu de gamme	70%	150%	125%	115%
	Haut de gamme	40%	80%	60%	60%
SENIOR	Entrée de gamme	90%	190%	175%	165%
	Milieu de gamme	65%	155%	120%	110%
	Haut de gamme	40%	85%	60%	50%

FIGURE 3.6 – Taux de déformation de la fréquence de consommation des assurés pour obtenir un impact global de la réforme de 15%

Pour rappel, si une déformation de +100% est appliquée à un acte, cela signifie que les assurés consommeront deux fois plus cet acte après la mise en place de la réforme « 100% santé ». Les taux observés dans le tableau de la Figure 3.6 sont pour la plupart supérieurs à 100% pour l'entrée de gamme. Le taux de déformation pour un senior entrée de gamme en audiologie atteint même 190%, ce qui signifie qu'un assuré consommerait quasiment 3 fois plus qu'en 2018.

Avec ces taux de déformation, les impacts suivants sont obtenus :

	<i>Sans déformation</i>	<b>Impacts extrêmes</b>
<b>DENTAIRE</b>	31,8%	150 %
<b>AUDIOLOGIE</b>	80,2%	326 %
<b>OPTIQUE</b>	- 5,6%	10 %
<b>TOTAL</b>	<b>2,6%</b>	<b>15 %</b>

FIGURE 3.7 – Impacts de la réforme sur les remboursements complémentaires incluant la déformation de fréquence de consommation des assurés

Afin de mieux visualiser l'impact de ces déformations de fréquences, il est possible de représenter les fréquences de consommation selon les niveaux de garantie et les postes avant et après application des déformations.

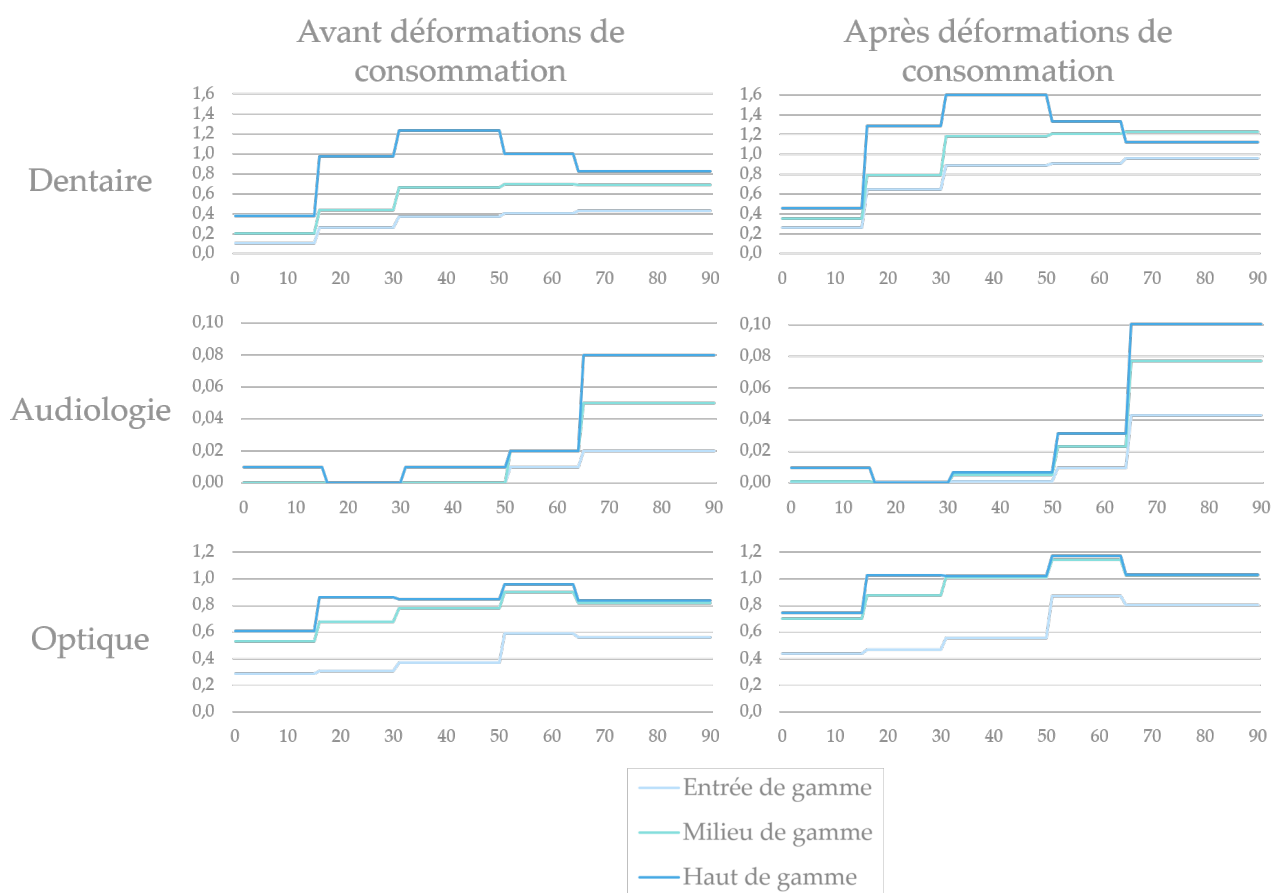


FIGURE 3.8 – Fréquences de consommation des assurés en fonction de leur âge, avant & après déformation de fréquence

Il apparaît que quelque soit le niveau de garantie, l'augmentation de fréquence est significative. En dentaire, les disparités observées entre les niveaux de garantie avant application des déformations sont fortement réduites une fois cette application effective, bien que la fréquence de consommation dans le haut de gamme reste supérieure. En audiologie, la fréquence de consommation en entrée de gamme reste inférieure à celle des deux autres niveaux de garantie. Par ailleurs, en optique, les niveaux milieu et haut de gamme en viennent presque à se confondre et la fréquence de l'entrée de gamme pour la tranche d'âge 50 ans et plus augmente fortement. Ainsi, dans le cas extrême d'un impact global de 15%, l'objectif du gouvernement de réduire le renoncement aux soins des plus démunis serait effectivement atteint.

Cette partie de l'étude permet donc d'observer que les déformations de fréquences de consommation nécessaires pour atteindre un impact global de 15% sont très supérieures aux taux de renoncement aux soins. Cette constatation permet à l'assureur de mieux appréhender les conséquences de la réforme « 100% santé » : dans le cadre de cette étude, et en suivant des hypothèses sélectionnées avec sa participation, il devrait être capable d'absorber la hausse des prestations.



---

## En résumé

Suite à la réforme « 100% santé », la façon de consommer des assurés devrait être modifiée puisque son but est de réduire le renoncement aux soins. Ainsi, pour étudier l'impact de ce changement de comportement, des taux de déformation sont appliqués à la consommation de 2018. Ces taux de déformation sont d'abord choisis en fonction des taux de renoncement aux soins pour chaque poste, puis déclinés selon la gamme et le niveau de garantie. Puisque ce choix reste une hypothèse lourde, des scénarios minorants et majorants sont effectués afin d'obtenir un intervalle d'impacts. L'impact global du scénario central (basé sur les taux de renoncement aux soins) est de 6,7% et cet impact devrait être compris entre 5,6% et 8,9%.

D'autre part, il est possible de considérer un choc maximal que l'assureur est prêt à encaisser (ici ce choc est fixé à +15% par le client d'Addactis). L'intérêt de cette approche est d'observer les taux de déformation nécessaires à l'obtention de ce choc maximal. Ici, les taux obtenus sont très largement supérieurs à ceux du renoncement aux soins.

## TROISIÈME PARTIE

### SEGMENTATION DU PORTEFEUILLE

La consommation de soins par les assurés ne dépend pas seulement de la gamme et du niveau de garantie. La base de données créée dans la section 1.3 de la Partie II contient de nombreuses variables qui pourraient permettre de mieux appréhender le comportement actuel des assurés et donc leur potentiel changement de consommation suite à la réforme « 100% santé ». Cette partie a donc pour but de développer une segmentation du portefeuille du client afin d'identifier des profils d'assurés ayant une consommation similaire en terme de remboursements complémentaires. Il est important de noter que la segmentation qui sera obtenue ne pourra pas reposer sur un quelconque historique de données liées aux changements consécutifs à la réforme « 100% santé ». Il s'agit ici d'une segmentation uniquement basée sur les données 2018 qui a pour but de donner à l'assureur un état de son portefeuille à cette date, ce qui pourra lui permettre de le piloter et d'émettre des hypothèses quant à la prévision du comportement des assurés après 2018.

Il existe deux types d'analyse d'une base de données : l'analyse non supervisée et l'analyse supervisée. L'analyse non supervisée ne tient pas compte d'une variable cible, c'est-à-dire que l'étude s'effectue sur la totalité des variables sans tenter d'en expliquer une par rapport aux autres. À l'inverse, l'analyse supervisée permet d'étudier le lien entre des variables explicatives et une variable cible. Ce type d'analyse a par exemple pour but de prédire la variable cible ou bien de créer des groupes d'individus qui sont voisins pour cette variable.

# 1 | ÉTUDE NON SUPERVISÉE

Avant d'étudier les profils d'assurés par rapport aux remboursements complémentaires (qui sont donc la variable cible, ou variable à expliquer), il peut être intéressant d'étudier dans un premier temps les variables explicatives uniquement. En effet, si certaines variables explicatives sont fortement liées, cela pourrait biaiser l'identification de profils d'assurés par rapport à la variable cible. Ainsi, le premier chapitre présente une analyse non supervisée de la base de données.

## 1 Présentation théorique des modèles

### Définitions préliminaires

Il est nécessaire de rappeler certaines notions afin de poser le cadre théorique des modèles.

**L'inertie** Elle est définie par : *Inertie totale = Inertie interclasse + Inertie intraclasse*.

Plus précisément, d'après le théorème fondamental de Huygens, soit  $(C = x_i)_{i \in \llbracket 1; n \rrbracket}$  un groupe d'individus de centre de gravité  $G$  qui doit être divisé en  $K$  classes  $(C_1, \dots, C_K)$  de centre de gravité  $(G_1, \dots, G_K)$  et de taille  $(n_1, \dots, n_K)$  et soit  $d$  une mesure de distance entre deux individus, alors

$$\sum_{i=1}^n d^2(x_i, G) = \sum_{k=1}^K n_k \cdot d^2(G_k, G) + \sum_{k=1}^K \sum_{i=1}^{n_k} d^2(x_i, G_k)$$

L'inertie interclasse indique à quel point les différents groupes sont différents alors que l'inertie intraclasse mesure la concentration et la similarité au sein d'un même groupe.

**Un médoïde** Il s'agit d'un point appartenant à une classe. Il est le point pour lequel la distance avec l'ensemble des autres points du groupe est minimisée. En d'autres termes, avec  $M$  le médoïde,  $d$  la distance et  $(x_i)_{i \in \llbracket 1;n \rrbracket}$  :

$$M = \operatorname{argmin}_{m \in \llbracket 1;n \rrbracket} \sum_{i=1}^n d(x_i, m)$$

**La distance de Gower** Il s'agit de la distance qui sera utilisée dans cette étude et qui est capable de mesurer la distance entre deux individus, quelle que soit la nature des variables qui composent la base de données (aussi bien qualitative que quantitative). Elle se base sur la notion de dissimilarité partielle et permet d'obtenir une matrice de dissimilarité qui pourra par la suite être utilisée dans des algorithmes tels que PAM.

Ainsi, si la base de données compte  $p$  variables, alors pour deux individus  $x_i$  et  $x_j$ ,  $p$  dissimilarités partielles  $d_{ij}^v \in [0;1]$  seront calculées. Chacune d'entre elles dépend de la nature de la variable  $v$ .

Si elle est qualitative,  $d_{ij}^v = 1_{\{x_i^v \neq x_j^v\}}$  ;

Et si elle est quantitative,  $d_{ij}^v = \frac{|x_i^v - x_j^v|}{\max(x^v) - \min(x^v)}$  .

Avec  $x_i^v$  la valeur de  $v$  pour l'individu  $i$  et  $\max(x^v)$  le maximum de  $v$ .

Enfin, la distance de Gower entre les individus  $x_i$  et  $x_j$  est caractérisée par

$$Gower(x_i, x_j) = \frac{1}{p} \cdot \sum_{v=1}^p d_{ij}^v$$

Par ailleurs, cette mesure permet également de pondérer les différentes variables. Ainsi, si le poids de la variable  $v$  est  $\omega_v$ , la distance de Gower devient :

$$Gower(x_i, x_j) = \frac{1}{\sum_{v=1}^p \omega_v} \cdot \sum_{v=1}^p \omega_v \cdot d_{ij}^v$$

## Test de Cramer

Le test de Cramer permet de mesurer l'intensité des relations observées entre les variables. Les relations sont observées à l'aide du test du  $\chi^2$ . Ce test  $\chi^2$  détermine s'il y a ou non une relation entre deux variables en utilisant les hypothèses suivantes :

- hypothèse nulle  $H_0$  : les deux variables sont indépendantes ;
- hypothèse alternative  $H_1$  : un lien existe entre les deux variables, la connaissance d'une variable apporte une information pour la prédiction de l'autre variable.

L'inconvénient de ce test  $\chi^2$  est qu'il n'est pas stable lorsque la taille de l'échantillon augmente. Le test de Cramer permet de remédier à ce problème puisqu'il reste stable si la taille de l'échantillon augmente dans les mêmes proportions inter-modalités. En effet, en plus de la valeur du test du  $\chi^2$ , le test de Cramer prend en compte la taille de l'échantillon ( $N$ ) et le degré de liberté de la base de données ( $DDL$ ). Ainsi,  $V_{\text{Cramer}} = \sqrt{\frac{\chi^2}{N \cdot DDL}}$ , et plus  $V_{\text{Cramer}}$  est proche de 1, plus l'intensité de la relation entre les deux variables est grande.

---

## L'algorithme PAM

L'algorithme PAM (Partitioning Around Medoids) est également appelé k-médoïdes. Il s'agit d'une version plus robuste du k-means qui permet de prendre en compte des variables qualitatives. Il prend en paramètre le nombre de classes voulu par l'utilisateur et associe à chaque itération un groupe à chaque individu.

**L'algorithme** En entrée de l'algorithme PAM doivent être portés une matrice de distance (ou de dissimilarité) et le nombre de clusters souhaités ( $K$  ici). Les  $K$  médoïdes sont le plus généralement choisis aléatoirement mais ils peuvent également être imposés par l'utilisateur. L'algorithme s'effectue de la façon suivante :

- |   |   |
|---|---|
| 1 | parmi les individus, $K$ médoïdes sont choisis aléatoirement;   |
| 2 | chaque individu est affecté au médoïde le plus proche (selon la distance choisie pour calculer la matrice de distance);   |
| 3 | recherche pour chaque groupe si un individu diminue la moyenne des dissimilarités. Si cet individu existe, il est choisi en tant que médoïde et si plusieurs individus vérifient cette condition, celui qui assure la plus grande réduction de dissimilarité est sélectionné; |
| 4 | si à l'étape 3 au moins un médoïde est modifié, recommencer à l'étape 2; sinon le clustering est terminé.   |

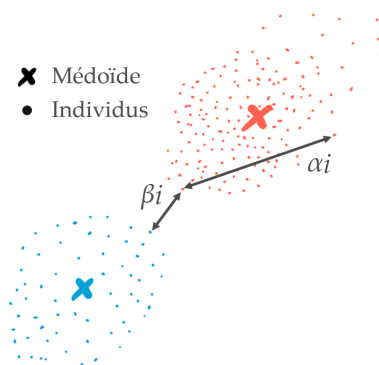
**La validation des résultats** Elle est effectuée au travers du coefficient de silhouette qui permet de mesurer la qualité de classement des individus en cluster. Ce coefficient indique si un point appartenant à un cluster  $C_k$  est correctement éloigné des autres clusters obtenus. Pour ce faire, la silhouette prend en compte la distance moyenne de l'individu  $x_i$  à son groupe  $\alpha_i$  et celle de l'individu au groupe voisin le plus proche  $\beta_i$  :

$$\alpha_i = \frac{1}{\text{Card}(C_k)} \cdot \sum_{\substack{j \in C_k \\ j \neq i}} d(x_i, x_j) \quad \& \quad \beta_i = \min_{k' \neq k} \frac{1}{\text{Card}(C_{k'})} \cdot \sum_{j \in C_{k'}} d(x_i, x_j)$$

La silhouette de l'individu  $x_i$  est alors donnée par la formule :

$$s(i) = \frac{\beta_i - \alpha_i}{\max(\alpha_i, \beta_i)}$$

Il est en effet possible que l'appartenance d'un individu  $i$  à un cluster  $C_k$  soit liée à sa proximité avec le médoïde de ce cluster. Néanmoins, si les individus appartenant à  $C_k$  sont très dispersés dans l'espace, il arrive que l'individu  $i$  soit plus proche d'individus appartenant à un autre cluster  $C_{k'}$ . C'est ce qu'illustre la figure ci-dessous :



---

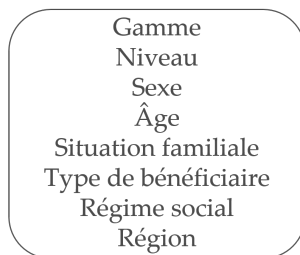
Il est également possible de vérifier la qualité du clustering dans sa totalité en faisant la moyenne des silhouettes de chaque individu et de chaque groupe :

$$S = \frac{1}{K} \cdot \sum_{k=1}^K \left( \frac{1}{\text{Card}(C_k)} \cdot \sum_{i \in C_k} s(i) \right)$$

## 2 Variables sélectionnées pour l'étude

Pour cette étude, seules les variables qui peuvent être connues avant la survenance d'un sinistre sont conservées. En effet, l'intérêt de cette étude est d'abord de connaître le profil d'un assuré pour ensuite pouvoir prédire le montant moyen de ses futures prestations. Cela pourra finalement permettre de mieux comprendre sa façon de consommer dans les trois postes concernés par la réforme « 100% santé », donc la manière dont cette consommation va pouvoir éventuellement changer suite à cette réforme. Pour ce faire, il n'est pas possible de prendre en compte des variables liées au sinistre qui ne s'est pas encore produit.

Ainsi, les variables sélectionnées dans la base de données sont les suivantes :



- L'âge est découpé en tranches : de 0 à 15 ans, de 16 à 30 ans, de 31 à 50 ans, de 51 à 64 ans et enfin plus de 65 ans.
- La situation familiale est composée de trois modalités : Célibataire, En couple et Inconnue.
- Le type de bénéficiaire (pouvant aussi être appelé Lien familial dans cette étude) se décline en trois modalités : Assuré, Conjoint et Enfant.
- Le régime social présente deux modalités : le régime général et les autres régimes.
- Les régions ont déjà été traitées dans la Partie I. La variable sélectionnée est celle qui lie la région à sa densité de population INSEE. Il y a finalement quatre modalités : Peu dense, Moyennement dense, Très dense et Étranger.

Ainsi, les lignes de prestations initiales sont regroupées de la façon suivante : lorsque les modalités sont identiques pour les huit variables explicatives conservées, les montants de prestations 2018 sont sommés, ainsi que les expositions. Seules les dix variables (les explicatives, la somme des remboursements 2018 et la somme des expositions) sont finalement conservées.

Chaque poste présentant des montants de prestations très différents les uns des autres, chaque étude sera déclinée pour chacun des trois postes de soins. Ainsi, une fois tous les traitements effectués, les bases pour les postes dentaire et optique comportent environ 1500 lignes, ce qui est raisonnable dans le cadre des modèles utilisés. Cependant, pour le poste audiologie, la base ne contient que 250 lignes, ce qui risque de poser certains problèmes pour l'algorithme PAM.

### 3 Le test de Cramer

La Figure 1.1 illustre les résultats des tests de Cramer pour les trois postes réformés.

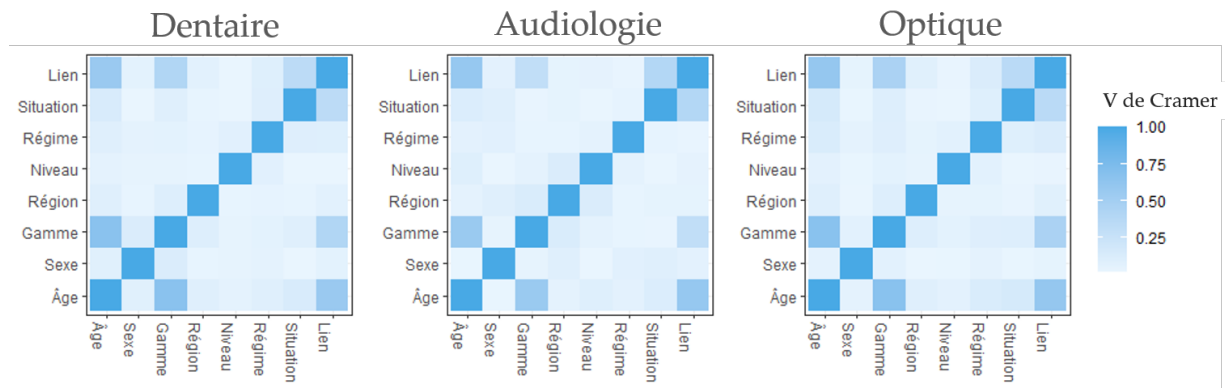


FIGURE 1.1 – Tests de Cramer pour les trois postes de l'étude

Les résultats des trois tests sont très proches. La majorité des variables sont décorrélées. Cependant, l'âge présente une corrélation plus élevée avec le lien familial et la gamme de garantie. Ces corrélations s'expliquent par le fait que les garanties sont découpées en fonction de l'âge de l'assuré et que dans le lien familial se trouve la modalité 'Enfant' nécessairement liée à l'âge. La gamme de garantie et le lien familial sont également légèrement corrélés, encore une fois car la présence d'enfant(s) dans le foyer familial conditionne la gamme de garantie choisie. Enfin, la situation familiale et le lien familial sont également corrélés puisque le fait d'être en couple induit souvent la modalité 'Conjoint' et la présence d'enfant(s).

Le fait d'avoir des variables peu corrélées assure une meilleure cohérence dans les résultats d'une approche supervisée. Cependant, l'étude non supervisée se base justement sur les ressemblances entre les variables et leurs modalités. Les résultats de ce chapitre seront donc à prendre avec le recul nécessaire.

### 4 L'algorithme PAM

L'algorithme PAM utilise une distance à définir au préalable. Pour cette étude, c'est la distance de Gower précédemment décrite qui est choisie. Elle est de plus pondérée par l'exposition de chaque ligne afin de prendre en compte la répartition des individus.

La première étape de l'algorithme PAM est de choisir le nombre de clusters à obtenir. Pour ce faire, il est possible de représenter graphiquement la valeur de la moyenne des silhouettes pour différents  $K$ . Dans le cadre de l'étude, il serait intéressant de créer plus de deux groupes.  $K$  est donc choisi entre 3 et 20 pour les trois postes.

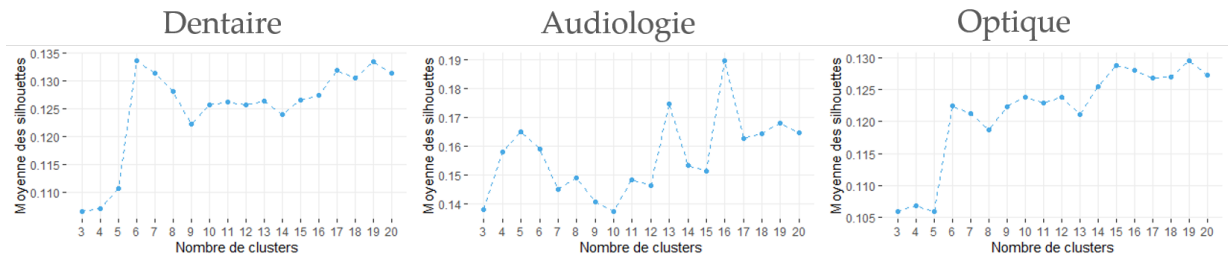


FIGURE 1.2 – Valeur des moyennes des silhouettes en fonction du nombre de clusters

En dentaire, le  $K$  optimal est  $K = 6$ . Pour les postes audiologie et optique, les  $K$  optimaux sont élevés (respectivement  $K = 16$  et  $K = 19$ ). Cependant, il est nécessaire de prendre en compte l’objectif de cette partie : regrouper les individus similaires et visualiser des groupes les plus homogènes possible. Pour le poste audiologie, étant donné qu’il y a beaucoup moins de lignes, il faudrait réduire le nombre de clusters. Pour le poste optique, le nombre de 19 clusters est excessif. Aussi, il faudrait trouver un équilibre entre un nombre trop grand de clusters et une silhouette élevée.  $K = 5$  pour l’audiologie et  $K = 6$  pour l’optique sont des choix raisonnables. Afin d’étudier l’influence du nombre de clusters sur les résultats de l’algorithme PAM, pour les postes audiologie et optique, deux  $K$  différents sont choisis : en plus de ceux choisis précédemment,  $K = 16$  pour l’audiologie et  $K = 15$  pour l’optique seront testés.

Afin de vérifier la qualité de l’algorithme, il est possible d’étudier la valeur de la silhouette pour chaque individu (et non plus la moyenne de la silhouette sur les individus) au travers du silhouette plot. Plus il y a d’individus ayant une silhouette supérieure à la moyenne, plus le clustering est efficace. Par ailleurs, il est possible que la silhouette d’un individu soit négative, ce qui signifie que l’individu est probablement excentré par rapport aux autres individus de son groupe, et que donc il paraît plus proche d’un autre groupe.

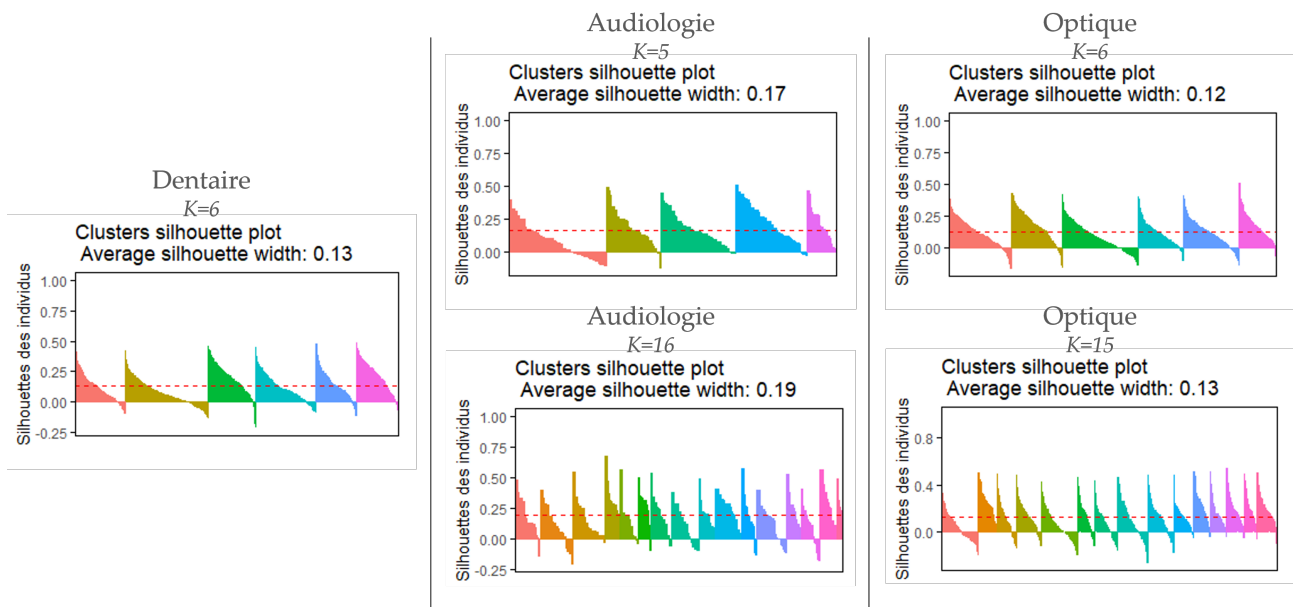


FIGURE 1.3 – Valeur des moyennes des silhouettes pour l’algorithme PAM en fonction du nombre de clusters



Le silhouette plot du poste dentaire montre que des individus de chaque cluster ont une silhouette supérieure à la moyenne, preuve d'un bon clustering, mais quelques individus ont des silhouettes négatives dans chaque cluster. Au vu des corrélations très faibles obtenues suite au test de Cramer, le résultat de ce clustering est satisfaisant.

En audiologie, le silhouette plot pour  $K = 5$  est meilleur que celui pour  $K = 16$ , ce qui s'explique par le faible nombre de lignes : séparer peu de lignes en un trop grand nombre de clusters donne de nombreux individus mal classés (ie. avec une silhouette négative). C'est donc  $K = 5$  qui est retenu pour la suite de l'étude.

En optique, bien que la base de données présente un plus grand nombre de lignes, le constat est le même que pour l'audiologie : un nombre de clusters plus petit donne de meilleurs résultats. C'est donc  $K = 6$  qui est choisi pour la suite de l'étude.

Enfin, pour visualiser les groupes créés par l'algorithme PAM pour les  $K$  retenus pour chaque poste, il est possible d'utiliser plusieurs méthodes. Tout d'abord, il est possible de faire apparaître les clusters sur les axes obtenus suite à une ACM. Il est également possible de projeter les individus sur le résultat de l'algorithme t-SNE. La théorie de l'ACM ainsi que celle de l'algorithme t-SNE sont développées dans les Annexes C et D. Les deux résultats sont fortement différents :

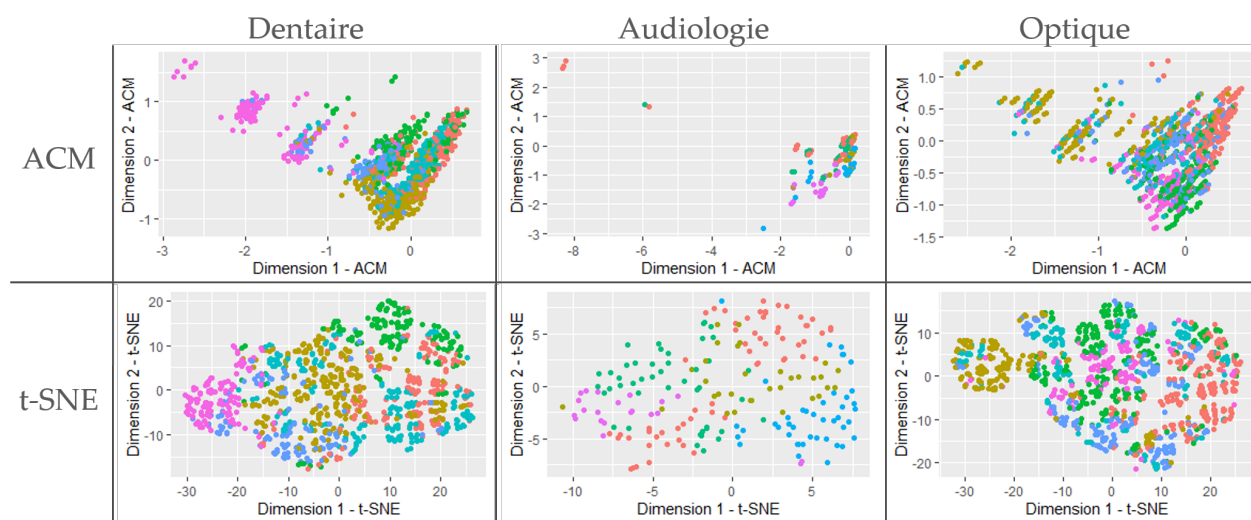


FIGURE 1.4 – Visualisations des résultats des algorithmes PAM

Les résultats issus de la projection des clusters sur les dimensions obtenues suite à l'algorithme t-SNE sont bien plus lisibles que ceux issus de l'ACM. Malgré tout, les résultats ne sont pas concluants : il est impossible d'identifier clairement des groupes distincts. Ce résultat permet de mettre en évidence que, sans variable à expliquer, il n'est pas possible de créer des profils d'assurés. C'est pourquoi la suite de l'étude reposera sur des méthodes supervisées avec pour but d'expliquer le montant de remboursement effectué en 2018 en fonction des huit variables explicatives.

---

## En résumé

Le but de cette troisième partie est d'identifier des profils d'assurés ayant une consommation similaire en terme de remboursements complémentaires. Cependant, il est important dans un premier temps d'étudier les variables dites explicatives (celles qui serviront à créer lesdits profils). C'est à cet effet que l'étude non supervisée a été menée.

Cette étude non supervisée montre que sans variable cible, le regroupement d'individus ne mène pas à des résultats concluants. En effet, quelle que soit la méthode de visualisation des clusters utilisée, aucun groupe ne se distingue clairement.

---

## 2 | ÉTUDE SUPERVISÉE

L'objectif de ce chapitre est de proposer à l'assureur une segmentation de son portefeuille en fonction de la consommation de soins médicaux des assurés, plus précisément celle de soins audiolinguistiques, dentaires et optiques. Cette segmentation est effectuée à l'aide de méthodes dites supervisées. La méthode la plus utilisée pour segmenter un portefeuille est celle des arbres de décisions car elle offre un bon compromis entre la volonté de segmentation et l'interprétabilité des résultats. Néanmoins, comme l'étude théorique pourra le montrer, cette méthode n'est pas la plus robuste et ne mène pas le plus souvent à la meilleure prédiction. Ainsi, afin d'estimer la qualité des profils d'individus identifiés à l'aide des arbres, d'autres méthodes seront utilisées pour prédire la variable cible.

### 1 Présentation théorique des modèles

Tout d'abord, il est nécessaire d'étudier les différents modèles d'un point de vue théorique. Cela permet notamment d'identifier les points forts ainsi que les faiblesses de chacun, mais aussi d'expliquer les choix qui seront faits lors de la segmentation du portefeuille de l'assureur.

#### Définition préliminaire : le GridSearch

La plupart des modèles repose sur des hyperparamètres que l'utilisateur doit donner en entrée de l'algorithme. Le choix de ces hyperparamètres est ardu car il arrive souvent qu'en les calibrant un par un, les résultats ne soient pas ceux escomptés. Ainsi, la méthode du Grid Search est une méthode d'optimisation permettant de tester une série de paramètres pour un modèle, d'en comparer les performances et donc de trouver la meilleure combinaison.

La méthode du Grid Search est la suivante : pour chaque paramètre, l'utilisateur choisit un ensemble de valeurs à tester. Ensuite, un modèle est créé pour chacune des combinaisons de paramètres (qui constituent une « grille »). Cette méthode nécessite donc d'implémenter autant de modèles qu'il existe de combinaisons parmi les choix de l'utilisateur, elle est de ce fait très rapidement coûteuse en terme de temps d'exécution.

#### Modèles Linéaires Généralisés (GLM)

Pour commencer, les modèles linéaires généralisés (ou GLM) sont les plus fréquemment utilisés en assurance. Ils permettent en effet d'obtenir la prédiction d'une variable cible à partir d'une combinaison d'un lot de variables explicatives déterminées. Il est également possible de lire dans le résultat d'un modèle GLM un début de piste pour une potentielle segmentation. En effet, pour chaque modalité des variables explicatives, un coefficient est associé. Lorsque ce coefficient est élevé, cela signifie qu'une variation de la modalité aura un impact important sur le résultat de la prédiction.

Néanmoins, ce type de modèle n'offre pas de segmentation claire et définitive. C'est pourquoi cette méthode sera simplement utilisée dans un but comparatif. Ainsi, seuls les aspects théoriques intéressants pour l'étude seront détaillés ici. Pour une étude plus approfondie de

---

la théorie des Modèles linéaires généralisés, le lecteur pourra se référer au livre de CHARPENTIER A. [6].

Soient  $Y$  la variable cible devant appartenir à une famille exponentielle, et  $(X_1, \dots, X_n)$  l'ensemble des variables explicatives, le modèle GLM étudie alors la relation entre  $Y$  et  $(X_1, \dots, X_n)$  au travers d'une fonction de lien  $g$  dérivable et inversible telle que :

$$g(\mathbb{E}[Y|X = x]) = X \cdot \beta$$

Tout d'abord, un modèle statistique  $(\Omega, F, (\mathbb{P}_{\theta, \phi})_{\theta \in \Theta, \phi > 0})$  est une famille exponentielle si la densité  $f$  des probabilités  $\mathbb{P}_{\theta, \phi}$  est telle que  $\mathbb{P}_{\theta, \phi}[Y = y] = f_{\theta, \phi}(y) = c_{\theta}(y) \cdot \exp\left(\frac{y\theta - a(\theta)}{\phi}\right)$  avec,  $\theta$  le paramètre canonique,  $\phi$  le paramètre de dispersion et  $a$  une fonction convexe de classe  $\mathcal{C}^2$ . Par exemple, les distributions de Bernoulli, Binomiale, Géométrique, de Poisson, Normale, Exponentielle et Gamma appartiennent à une famille exponentielle.

D'autre part, les coordonnées du paramètre  $\beta$  sont calculées par la méthode du maximum de vraisemblance. Elles permettent d'obtenir la prédiction de la variable cible par la formule suivante :

$$\mathbb{E}[Y|X = x] = g^{-1}(X \cdot \beta)$$

Ce sont ces coordonnées du paramètre  $\beta$  qui peuvent mener à une première interprétation de segmentation du portefeuille, comme cela pourra être étudié lors de la partie pratique de ce chapitre. Il reste néanmoins important de considérer qu'un modèle GLM n'est pas initialement fait pour segmenter un portefeuille et donc que cette méthode ne peut être retenue que pour permettre de challenger les résultats de segmentation obtenus avec d'autres méthodes ainsi que les résultats de prédictions de la variable cible <sup>1</sup>.

Enfin, un GLM est un modèle paramétrique, c'est-à-dire qu'il nécessite un a priori sur la distribution de la variable cible ou plus précisément celle de  $Y|X = x$ . Le choix de la distribution ainsi que de la fonction de lien détermine grandement les résultats obtenus et fera l'objet d'une étude plus approfondie lors de l'application du modèle GLM au portefeuille de l'étude. Chaque distribution est liée à une fonction de lien canonique, mais il est également possible de forcer le choix de la fonction de lien.

## Arbres CART

Les arbres de décision sont très largement utilisés du fait de leur interprétation simple et intuitive. Il existe de nombreux types d'arbres de décision, dans cette étude seuls les arbres binaires de régression, qui sont également appelés arbres CART (*Classification And Regression Trees*) seront retenus. Ces arbres permettent de diviser un groupe d'individus selon des critères rendant les deux groupes induits plus homogènes. Un des intérêts de ces arbres de décision est que contrairement aux modèles paramétriques comme les GLM, ils ne nécessitent pas qu'une hypothèse soit faite sur la distribution des données. Enfin, la méthode des arbres CART permet de prendre en compte plus facilement les non linéarités entre les variables explicatives et la variable cible.

---

1. Les métriques ( $R^2$  et Indice de Gini) utilisées pour comparer les résultats sont présentées à la suite de la théorie

Afin de comprendre le mécanisme des arbres CART, il est nécessaire d'introduire les notions suivantes :

- le nœud est le point où un groupe est divisé en deux à partir d'une condition qui sera définie dans la suite de l'étude théorique. Il existe le nœud parent (qui possède des successeurs) et le nœud enfant (qui possède un prédécesseur);
- la racine est le premier nœud de l'arbre, et n'a donc pas de nœud parent;
- les nœuds internes sont les nœuds intermédiaires qui rendent la sélection plus fiable;
- les feuilles sont le résultat final de l'arbre de décision, elles n'ont pas de nœud enfant;
- la profondeur de l'arbre est le nombre maximal de nœuds présents entre la racine et les feuilles.

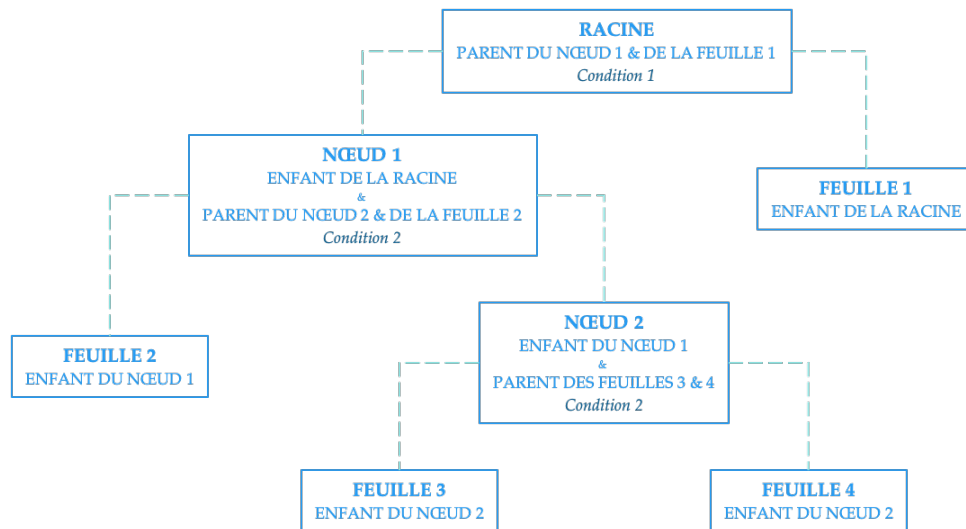


FIGURE 2.1 – Illustration des notions définies, la profondeur de cet arbre vaut 3

La méthode des arbres CART repose sur la division d'un nœud parent  $N$  en deux nœuds enfants  $N_G$  (*nœud gauche*) et  $N_D$  (*nœud droit*). Le choix des nœuds enfants repose sur un couple (caractéristique, seuil) noté  $(j, t)$  de telle sorte que :

$$N_G(j, t) = \{x \in N : x_j < t\} \quad \text{et} \quad N_D(j, t) = \{x \in N : x_j \geq t\},$$

où  $x_j$  représente la  $j$ -ième variable explicative de l'individu  $x$

Le choix du couple  $(j, t)$  se fait par l'utilisation de deux notions : l'impureté du nœud, qui n'est autre que la variance du nœud dans le cas d'une régression; et le gain d'information, donné par la formule :

$$IG(j, t) = Var(N) - \frac{Card(N_G(j, t))}{Card(N)} \cdot Var(N_G(j, t)) - \frac{Card(N_D(j, t))}{Card(N)} \cdot Var(N_D(j, t))$$

Ainsi, le couple  $(j, t)$  permettant d'obtenir l'impureté minimale et le gain d'information maximal est retenu. Deux nouveaux nœuds sont alors créés. La valeur de la prédiction de la variable cible pour chaque nœud est la moyenne de cette dernière pour le nœud considéré.

---

Cette procédure de division de nœuds peut techniquement être répétée jusqu'à ce que chaque nœud ne contienne plus qu'un seul élément. Cependant, il s'agirait d'une situation de sur-apprentissage : le modèle connaîtrait parfaitement les données à partir desquelles il a été établi mais ne serait pas performant dans la prédiction de nouvelles données, ce qui reste le but final d'un tel modèle. Aussi, il est nécessaire d'ajouter en paramètre du modèle un critère d'arrêt. Il peut s'agir par exemple d'une profondeur maximale de l'arbre, d'un nombre minimum d'individus dans les feuilles ou encore d'un seuil (appelé complexité) à partir duquel l'impureté est considérée comme acceptable et donc pour lequel le nœud ne sera pas divisé. L'élagage d'un arbre repose sur une théorie développée dans l'ouvrage de BREIMAN L. et al [4] auquel le lecteur pourra se référer pour plus de détails.

En pratique, pour chacun des postes la méthodologie suivante est appliquée :

- un GridSearch est effectué pour une complexité maximale comprise entre  $10^{-5}$  et  $10^{-4}$  par pas de  $10^{-5}$ , et pour une profondeur d'arbre maximale allant de 3 à 6.
  - Pour chaque couple de paramètres du GridSearch, un arbre CART est calculé.
  - Cet arbre est ensuite élagué en utilisant le paramètre de complexité. En effet, il existe une complexité optimale (inférieure ou égale à la complexité maximale entrée en paramètre de l'arbre CART) qui permet de réduire le plus possible l'erreur de prédiction de l'arbre CART.
  - Le coefficient  $R^2$  de cet arbre est calculé sur la base de test créée précédemment et gardé en mémoire.
- l'arbre CART final est celui dont les paramètres maximisent la valeur du  $R^2$  et qui est ensuite élagué à l'aide de la complexité optimale.

Une fois l'arbre CART obtenu, il est important de prendre en compte que toutes les variables utilisées n'ont pas la même importance dans le processus de segmentation des données. En effet, chaque critère de division des nœuds ne se vaut pas : certains permettent un gain d'information plus important que d'autres. De plus, la construction de la règle de classification repose sur le choix, à chacune des étapes, de la division qui maximise la variation d'hétérogénéité. Il se peut qu'une autre division, impliquant la même ou une autre variable explicative, puisse avoir une variation d'hétérogénéité très proche de l'optimale, tout en étant inférieure. Il est donc possible de mesurer l'importance de chaque variable explicative, afin de déterminer lesquelles sont les plus utiles à la prédiction. En complément de l'ouvrage de BREIMAN L. et al [4], le lecteur intéressé par une analyse plus approfondie de l'obtention de ces importances pourra se référer au Chapitre 7 du cours de MALOT-TULEAU C. [17].

La méthode des arbres CART reste néanmoins généralement peu robuste : un faible changement dans la base d'apprentissage par exemple peut entraîner des résultats très différents.

## Forêts aléatoires (ou Random Forest)

Les forêts aléatoires sont une amélioration des arbres de décision. En effet, avec un seul arbre, il n'est pas rare que le sur-apprentissage soit important, ou que certaines spécificités des données ne soient pas prises en compte. Ainsi, les forêts aléatoires produisent un certain nombre d'arbres (fixé par l'utilisateur) et permettent d'obtenir en prédiction finale la moyenne des prédictions faites par chacun des arbres. L'utilisation de forêts aléatoires dans la prédiction d'une variable donne souvent des résultats satisfaisants. Néanmoins, son analyse est plus complexe que celle d'un simple arbre CART et ne permet pas d'obtenir

---

une segmentation du portefeuille (et donc dans le cadre de l'étude, des profils d'assurés). Les résultats obtenus par Random Forest pourront tout de même être comparés avec ceux des arbres CART, cela permettra de mesurer la qualité des profils obtenus par le modèle CART et ainsi de prendre du recul sur la segmentation obtenue. La construction d'une forêt aléatoire repose sur deux piliers.

D'une part, un bagging est appliqué aux arbres CART. Cela consiste à entraîner chaque arbre CART sur un sous ensemble aléatoire de la base d'apprentissage, plutôt que sur la base entière. Le résultat de prédiction est la moyenne des prédictions de chaque arbre CART. Le fait que le sous ensemble soit choisi aléatoirement pour chaque arbre permet de réduire la corrélation entre les arbres et donc d'améliorer les prédictions faites sur une base indépendante de la base d'apprentissage. La taille du sous ensemble est un paramètre choisi par l'utilisateur.

D'autre part, le sous échantillonnage des variables (ou features bagging) consiste à introduire une part d'aléatoire supplémentaire. Pour chaque arbre implémenté, seulement un sous échantillon de variables explicatives (dont la taille est fixée par l'utilisateur) est utilisé dans la division des nœuds alors qu'habituellement la division s'effectue en prenant en compte toutes les variables.

## Extrem Gradient Boosting (ou XG Boost)

L'algorithme XG Boost est une optimisation de l'algorithme d'arbres de Gradient boosting. Il a été développé dans l'ouvrage de CHEN T. et GUESTRIN C. [7] et est bien résumé dans le mémoire de OTTOU P. [22], textes auxquels pourra se référer le lecteur. De plus, comme les forêts aléatoires, il repose sur des arbres de décision mais aussi sur le procédé de Boosting, ce qui en fait un bon candidat pour vérifier la qualité des résultats du modèle CART.

Le boosting est un algorithme itératif qui, à chaque étape, donne davantage de poids aux données difficilement prédictibles afin d'améliorer le modèle en prédiction. L'algorithme XG Boost utilise également une validation croisée afin d'améliorer la robustesse du modèle final. C'est ce mélange de différentes méthodes qui donne à cet algorithme la réputation d'être en général plus performant que les autres algorithmes basés sur les arbres de décision. Néanmoins, sa force fait aussi sa faiblesse : l'application de telles méthodes fait de l'XG Boost un modèle assez coûteux en terme de temps de calcul, et il arrive que l'amélioration qu'il est capable d'apporter sur les prédictions de la variable cible ne contrebalance pas cette perte de temps.

Il existe de nombreuses fonctions sous  $R$  qui permettent d'obtenir un XG Boost. Dans cette étude, c'est le package *xgboost* qui est utilisé. Il permet notamment d'entrer une série de paramètres possibles qui mènent à la création de plusieurs modèles. Le modèle final est celui qui minimise le  $MSE$ , et donc celui qui maximise le  $R^2$ .

Dans l'étude, le paramétrage du modèle repose sur les quantités suivantes :

- le nombre  $k$  d'échantillons pour la validation croisée (la base initiale d'apprentissage est divisée en  $k$  échantillons, puis à chaque étape de la validation croisée,  $k - 1$  échantillons sont utilisés pour l'apprentissage du modèle, et le dernier échantillon est utilisé pour la validation du modèle);

- le nombre d'itérations du boosting (plus il est grand, plus la précision du modèle est importante mais plus le risque de sur-apprentissage l'est également);
- la profondeur maximale de chacun des arbres créés (comme pour l'arbre CART, une profondeur trop grande peut engendrer un sur-apprentissage, mais une profondeur trop petite ne permet pas un apprentissage correct);
- le nombre de variables choisies aléatoirement pour chaque arbre implémenté dans le modèle.

## Métriques de validation de modèle

L'idéal pour comparer des modèles serait d'utiliser une métrique globale, indépendante de tous les paramètres utilisés pour optimiser chacun des modèles. Une telle métrique n'est actuellement pas disponible, c'est pourquoi deux mesures sont choisies dans cette étude pour comparer et valider les modèles. Tout d'abord, le coefficient de détermination permet d'identifier si le modèle est capable de prendre correctement en compte la volatilité des données. Puis la courbe de Lorenz et l'indice de Gini permettent de mesurer la qualité de segmentation du modèle (c'est-à-dire sa capacité à conserver l'hétérogénéité des données dans la prédiction). Ainsi, un modèle qui optimise à la fois le coefficient de détermination et l'indice de Gini peut être considéré comme un bon modèle.

### Coefficient de détermination

Le coefficient de détermination, noté  $R^2(Y, \hat{Y})$ , calculé pour un modèle dont les prédictions du vecteur  $Y$  sont notées  $\hat{Y}$ , est défini par la formule suivante :

$$R^2(Y, \hat{Y}) = 1 - \frac{MSE(Y, \hat{Y})}{Var(Y)}, \quad \text{avec } MSE(Y, \hat{Y}) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad \& \quad Var(Y) = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$$

où  $n$  est le nombre d'observations et  $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$  la moyenne empirique des observations.

### Courbe de Lorenz et indice de Gini

L'indice de Gini mesure la qualité de segmentation et d'hétérogénéité d'un modèle. Il permet de connaître à quel point le modèle est capable de prendre en compte la volatilité des observations. Il se calcule à partir de la courbe de Lorenz où sont placés en abscisse le nombre de sinistres cumulés et en ordonnées les montants de prestations cumulés. Un exemple de courbe de Lorenz est présentée à la Figure 2.2. L'indice de Gini est donc calculé par la formule :  $\frac{A}{A+B}$ .

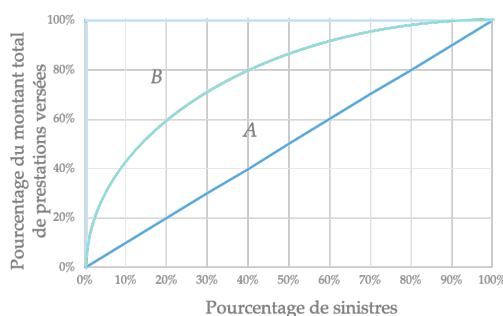


FIGURE 2.2 – Exemple d'une courbe de Lorenz



---

La lecture de cette courbe se fait de la façon suivante : 20% de sinistres représentent 60% des prestations versées par l'assureur. Les sinistres ont été ordonnés de façon décroissante en fonction de leur coût. Ainsi, les premiers 20% des montants de prestation représentent les sinistres les plus onéreux, et les derniers pourcentages représentent les sinistres les moins onéreux.

Dans le cadre d'une utilisation à des fins de comparaison de modèles, la courbe de Lorenz et l'indice de Gini sont observés pour les variables initiales du portefeuille, ainsi que pour les prédictions de chaque modèle. Plus l'indice de Gini d'un modèle est proche de celui du portefeuille, plus la segmentation de ce dernier est correctement représentée par le modèle.

### En résumé

L'étude supervisée est utilisée ici dans le but d'obtenir une segmentation intéressante du portefeuille de l'assureur afin de lui permettre de mieux le piloter lors de la mise en place de la réforme « 100% santé ». Pour ce faire, différents modèles vont être implémentés.

Tout d'abord, le modèle GLM étant le modèle le plus utilisé actuellement sur le marché de l'assurance, il est important de l'implémenter dans cette étude afin qu'elle puisse s'intégrer dans une offre d'aide au pilotage pour les assureurs. Ce modèle permet d'obtenir une prédiction relativement sûre et stable des remboursements complémentaires effectués par l'assureur. Il offre également la possibilité d'entrevoir une première segmentation « naïve » du portefeuille.

Néanmoins, la segmentation obtenue par GLM ne peut être une segmentation définitive. De plus, les potentielles non-linéarités des données ne sont pas correctement représentées par ce modèle. C'est pourquoi un modèle d'arbres CART est alors mis en place. Il offre une segmentation du portefeuille claire et sans équivoque et permet d'obtenir une prédiction de la variable cible pour chacun des groupes créés. Il est sans doute le meilleur compromis actuellement disponible entre la nécessité d'obtenir une segmentation lisible et utilisable du portefeuille et une prédiction correcte de la variable cible.

Le modèle CART est cependant connu pour être un modèle assez peu robuste. C'est pourquoi il a été perfectionné au travers des forêts aléatoires et du modèle XG Boost. Ces deux modèles offrent généralement une prédiction plus précise de la variable cible et prennent mieux en compte la volatilité de cette dernière. Étant donné que ces deux modèles ne permettent pas d'obtenir une segmentation lisible du portefeuille (en effet, la présence de plusieurs arbres, du processus de validation croisée, etc. empêche de connaître précisément les caractéristiques des individus présents dans un groupe), ils seront utilisés dans le but de vérifier que les résultats du modèle CART restent raisonnablement corrects et donc que la segmentation finale peut-être utilisée par l'assureur pour piloter son portefeuille.

## 2 Variables sélectionnées pour l'étude

Pour cette étude, la variable cible est le montant du remboursement complémentaire versé par l'assureur. Le choix des variables explicatives est détaillé dans la suite de cette section. De plus, la consommation variant fortement selon le poste étudié, une segmentation différente sera proposée pour chacun des postes impactés par la réforme « 100% santé ».

La gamme de garantie (Solo, Famille ou Senior) est trop fortement corrélée à l'âge du bénéficiaire (cf. Figure 1.1 du chapitre précédent). En théorie, il est nécessaire de ne garder qu'une des deux variables afin de ne pas sous estimer leur importance dans le modèle. Cependant, du point de vue métier, il est fortement recommandé que la gamme de garantie soit intégrée au profil de risque, et l'âge fait également partie des variables indispensables à la création de profils. Ainsi, une variable Âge / Gamme est créée, elle permet en une seule variable de connaître la tranche d'âge dans laquelle se situe l'assuré ainsi que la gamme de garantie à laquelle il a souscrit.

De plus, la variable sous poste est ajoutée pour l'étude du poste dentaire, car les montants de prestation de soins dentaires ou de prothèses dentaires sont trop différents pour être traités de la même façon. En optique, le sous poste (verre ou monture) ne sera pas conservé car, au sein du portefeuille étudié, les moyennes des prestations des deux catégories sont très proches. En effet, les garanties pour des verres optique proposées par le client n'étant pas très élevées par rapport à la moyenne du marché, il n'y a que très peu de remboursements très importants pour les verres optiques. Les variables sexe, région, niveau de garantie, régime social, situation familiale ainsi que lien familial sont prises en compte quelque soit le poste étudié.

D'autre part, contrairement à l'étude non supervisée, la notion d'individu est nécessaire ici et l'agrégation des lignes n'est donc pas exactement la même. Ainsi, à partir de la base de données créée dans la section 1.3 de la Partie II, pour chaque poste, les lignes sont regroupées par individus, le nombre de sinistres est compté, et le montant de remboursement moyen d'un sinistre est calculé à partir de ce nombre de sinistres et du montant total versé à l'assuré.

Agrégation par individu

Identifiant bénéficiaire	Âge Gamme	Sexe	Niveau de garantie	Régime social	Région	Situation familiale	Lien familial	Montant de remboursement 2018	
xxx	32 ans Solo	Masculin	Milieu de gamme	Général	Moyennement dense	Célibataire	Assuré	50 €	
xxx	32 ans Solo	Masculin	Milieu de gamme	Général	Moyennement dense	Célibataire	Assuré	60 €	
Identifiant bénéficiaire	Âge Gamme	Sexe	Niveau de garantie	Régime social	Région	Situation familiale	Lien familial	Montant de remboursement 2018 moyen	Nombre de sinistres
xxx	32 ans Solo	Masculin	Milieu de gamme	Général	Moyennement dense	Célibataire	Assuré	55 €	2

FIGURE 2.3 – Exemple d'agrégation des lignes par individu

Enfin, pour vérifier l'efficacité de prédiction de chaque méthode (et donc de segmentation pour l'arbre CART), il est nécessaire d'extraire de la base de données une partie des lignes avant d'appliquer les algorithmes. Cela permet de tester les modèles sur de nouvelles données (qui n'ont pas été utilisées pour entraîner le modèle) et ainsi de prévenir le phénomène de sur-apprentissage. La base d'apprentissage de l'algorithme représente 80% de la base de donnée originale, laissant 20% de cette dernière en base de validation. Les lignes sélection-

---

nées pour créer la base de validation sont choisies aléatoirement parmi la totalité de la base de données. Cette notion est détaillée dans la section suivante.

### 3 Méthode de l'étude

La segmentation est obtenue à partir d'un modèle d'arbre CART. Comme l'étude théorique a pu le mettre en avant, ce type de modèle est très sensible à la base d'apprentissage utilisée lors de l'implémentation. Ainsi, afin de s'assurer que les résultats de métriques obtenus ne sont pas une exception liée au choix de la base d'apprentissage, les modèles seront implémentés sur 1 000 bases d'apprentissage différentes (c'est-à-dire que les 80% des lignes de la base initiale sélectionnées pour créer la base d'apprentissage seront choisies aléatoirement 1 000 fois).

Le fait d'utiliser plusieurs bases d'apprentissage lors de l'implémentation des données permet d'obtenir une distribution des valeurs des métriques utilisées ( $R^2$  et indice de Gini) et ainsi de mieux comparer les performances des différents modèles.

Néanmoins, la lisibilité d'un arbre CART perd toute sa valeur lorsque 1 000 arbres sont implémentés pour connaître la performance du modèle. Ainsi, la segmentation finalement retenue sera celle du modèle CART présentant la valeur moyenne pour la métrique de validation  $R^2$ . Le fait de faire 1 000 modèles CART différents permettra simplement de vérifier que cette valeur n'est pas une valeur aberrante liée à une certaine base d'apprentissage.

Enfin, la méthode d'obtention de la segmentation du portefeuille ainsi que la présentation des résultats et leur analyse étant quasiment identiques pour les trois postes étudiés, le poste dentaire sera très détaillé, puis afin d'éviter une quelconque redondance, les postes audiologie et optique seront présentés plus rapidement. Une analyse complète est néanmoins disponible aux Annexes E et F pour ces deux postes de soins.

## 4 Segmentation et analyse pour le poste dentaire

Pour commencer, c'est la segmentation du portefeuille par rapport à la consommation des assurés dans le poste dentaire qui est étudiée. Ce poste présente une particularité : les montants remboursés pour les prothèses dentaires sont bien plus élevés que ceux des autres actes du poste. C'est pourquoi la variable informant du sous poste est conservée lors de la création des profils.

### Modèle linéaire généralisé (GLM)

Le modèle GLM est un modèle paramétrique et nécessite donc d'intuiter la distribution de la variable cible sachant les variables explicatives. Pour ce faire, il est possible de comparer graphiquement la distribution des données par rapport aux distributions usuellement utilisées dans des GLM. Les montants de prestations étant nécessairement positifs et continus, la loi testée sera la loi Gamma. Les paramètres de cette loi sont calculés en fonction de l'espérance et de la variance des observations.

De plus, la distribution des montants est à analyser en deux temps : la distribution des prestations pour les prothèses dentaires d'une part, et celle des autres actes dentaires d'autre part. Et les prédictions des prestations dentaires seront également faites en deux temps au travers de deux modèles GLM différents.

Les Figures 2.4 et 2.5 montrent que les distributions de Gamma ne sont pas parfaites pour modéliser les prestations : pour les prothèses dentaires un pic à 200€ est mal représenté par la loi Gamma, alors que pour les autres actes dentaires la densité de la loi Gamma n'est pas aussi importante que celle du portefeuille. Néanmoins, il s'agit de la loi la plus proche de la distribution observée du portefeuille et elle sera donc choisie pour les deux modèles GLM. En effet, il est rare que les données d'observation suivent parfaitement une loi connue, il est donc nécessaire de poser l'hypothèse lourde de la distribution puis de ne pas l'oublier lors de l'analyse finale des résultats.

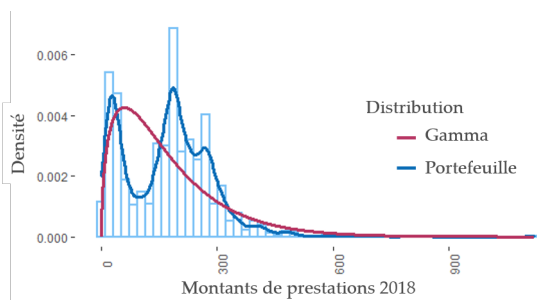


FIGURE 2.4 – Prothèses dentaires

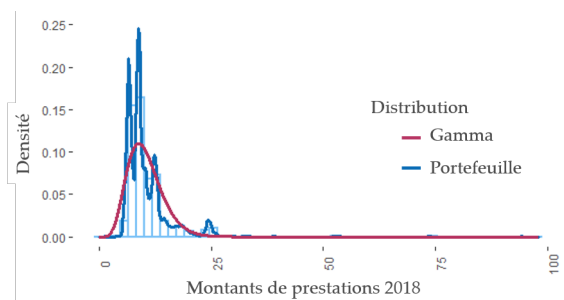


FIGURE 2.5 – Autres actes dentaires

Suite à la modélisation GLM, comme cela a été évoqué dans la partie théorique, une première segmentation « naïve » du portefeuille peut se dessiner en observant les coefficients du modèle pour les différentes modalités des variables. Les Figures 2.6 et 2.7 représentent ces coefficients pour les deux modèles GLM.

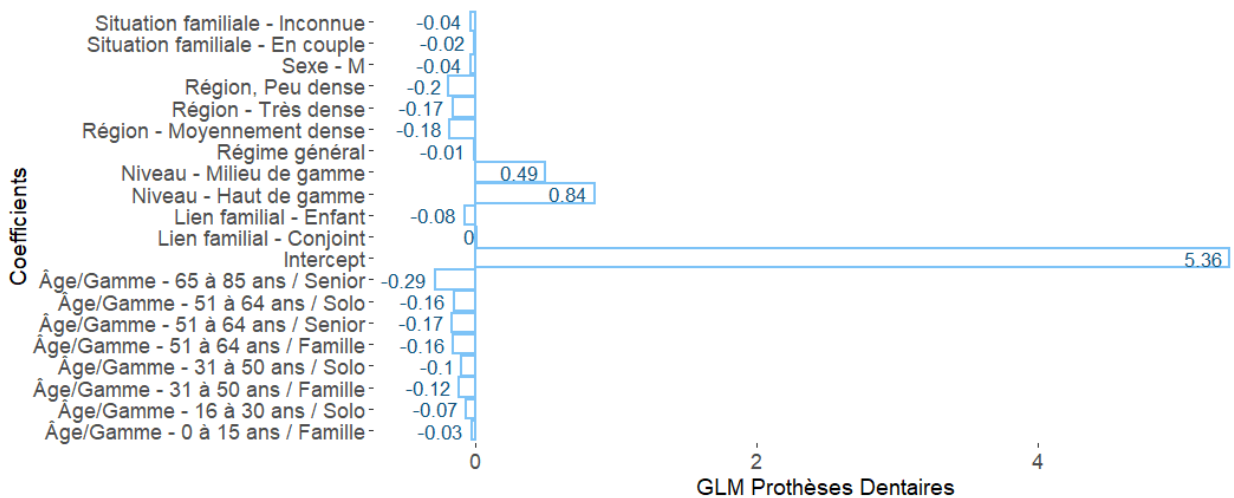


FIGURE 2.6 – Coefficients pour le modèle GLM implémenté sur les prothèses dentaires

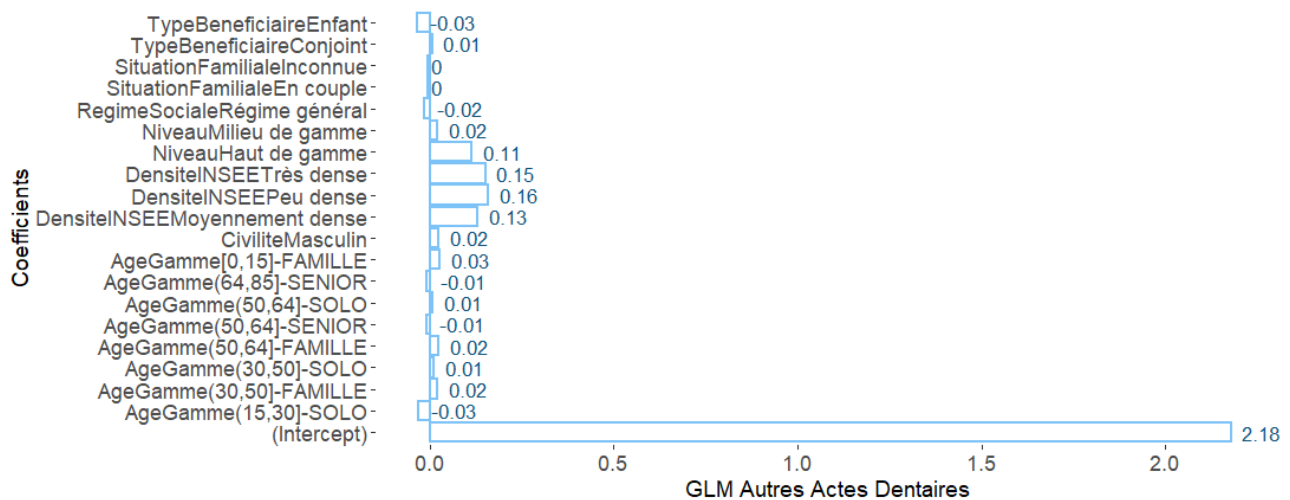


FIGURE 2.7 – Coefficients pour le modèle GLM implémenté sur les autres actes dentaires

La valeur du coefficient pour l'Intercept est la plus élevée pour les deux modèles. Les valeurs des autres coefficients décrivent si le montant de prestation sera supérieur ou inférieur lorsque la modalité de la variable prise pour calculer l'Intercept est remplacée par une autre. Ainsi par exemple pour les autres soins dentaires, les prestations sont plus faibles pour les assurés vivant à l'étranger par rapport aux autres zones géographiques (car pour les trois autres zones : peu, moyennement ou très dense, le coefficient est positif). Et c'est l'inverse pour les prothèses dentaires puisque les coefficients sont négatifs. Il est alors possible d'identifier que :

- les prestations entrée de gamme (présentes dans l'Intercept des deux modèles) sont inférieures aux prestations milieu et haut de gamme, ce qui est cohérent ;
- les remboursements faits pour les personnes de plus de 65 ans sont plus faibles que les autres ;
- pour les prothèses dentaires les prestations faites pour les hommes sont inférieures à celles faites pour les femmes alors qu'elles sont supérieures pour les autres soins dentaires.

Ces deux graphiques permettent donc d'avoir une première approche de segmentation selon le montant des remboursements prédit.

D'autre part, il est nécessaire d'observer la table des déviations en faisant un test ANOVA. Ce test renvoie les p-valeurs de chaque variable et permet donc de savoir quelles variables sont significatives (*p-valeur inférieure à 0.05*) et lesquelles ne le sont pas (*p-valeur supérieure à 0.05*). Ainsi d'après ce test, pour les prothèses dentaires les variables sexe et lien familial ne sont pas significatives pour un seuil de 5%, alors que pour les autres actes dentaires, la situation familiale et le lien familial ne sont pas significatifs pour ce seuil. Cela signifie que pour les prothèses dentaires, la modélisation par un GLM n'utilise pas les variables sexe et lien familial pour prédire les montants de prestations.

Par ailleurs, une fois les prédictions pour les prothèses dentaires et celles pour les autres actes dentaires assemblées en une même prédiction  $\hat{Y}$ , le  $R^2$  et l'indice de Gini peuvent être calculés, ce qui permettra de comparer l'efficacité des différents modèles par la suite. Pour cette première modélisation, le  $R^2$  vaut 69,64% et l'indice de Gini est de 0,815.

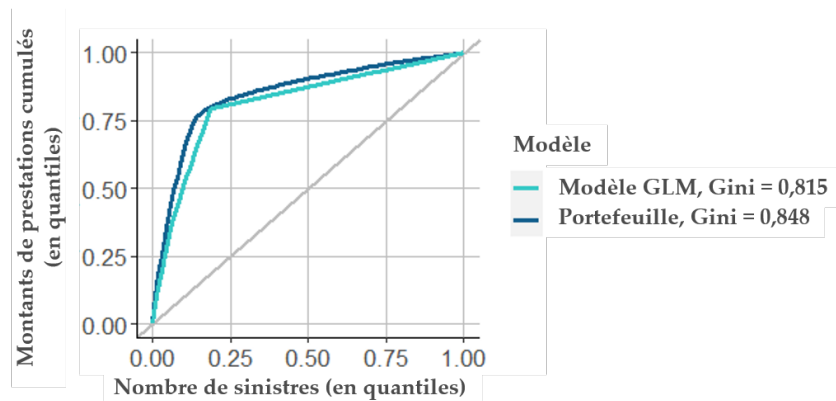


FIGURE 2.8 – Courbe de Lorenz associée au modèle GLM

La courbe de Lorenz du modèle GLM étudié est très proche de celle des observations du portefeuille, ce qui permet de considérer que les montants de remboursement sont correctement répartis par le modèle. Cependant, les deux courbes s'éloignent entre le quantile à 5% et le quantile à 20% des sinistres, ce qui signifie que pour cette tranche l'ordre des prédictions ne respecte pas exactement celui des observations. En effet, 15% des sinistres représentent dans le portefeuille 75% des remboursements alors que le modèle n'en prévoit que 55%.

Par ailleurs, l'indice de Gini ne permet pas de savoir si les prédictions du modèle sont proches des valeurs observées mais seulement de savoir si l'ordre entre les montants des sinistres prédits se rapproche de celui des montants observés. Ainsi, pour s'assurer que les prédictions sont cohérentes en ordre de grandeur, les moyennes doivent être comparées, et ce pour les prothèses dentaires d'une part puis pour les autres actes d'autre part.

PROTHÈSES DENTAIRES	Moyenne	AUTRES ACTES DENTAIRES	Moyenne
Base de test	170€	Base de test	10€
Valeurs prédites	168€	Valeurs prédites	10€

FIGURE 2.9 – Moyennes de la base de test et des prédictions obtenues avec le modèle GLM

Les moyennes des valeurs prédites sont très proches des valeurs observées, le modèle conserve donc bien l'ordre de grandeur des montants de prestations.

Enfin, comme cela a été décrit précédemment, le modèle GLM qui a été étudié ici est celui dont la valeur de  $R^2$  est proche de la moyenne des valeurs des 1 000 itérations. Par la suite, il est intéressant de connaître la distribution des valeurs de  $R^2$  et du coefficient de Gini. Les box-plots des Figures 2.10 et 2.11 présentent ces distributions. Cela permettra par la suite de comparer les performances des différents modèles.

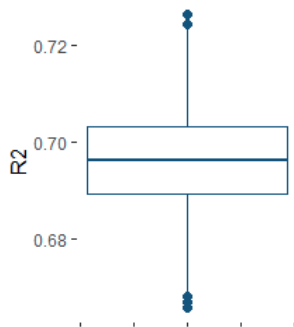


FIGURE 2.10 –  $R^2$  obtenus pour les modélisations GLM

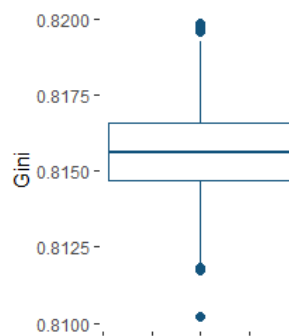


FIGURE 2.11 – Indices de Gini obtenus pour les modélisations GLM

## Arbre CART

La capacité de segmentation des modèles GLM étant très limitée, la méthode des arbres CART va être utilisée pour créer des profils d'assurés ayant la même consommation de soins, en terme de remboursement complémentaire.

L'arbre CART obtenu est le suivant :

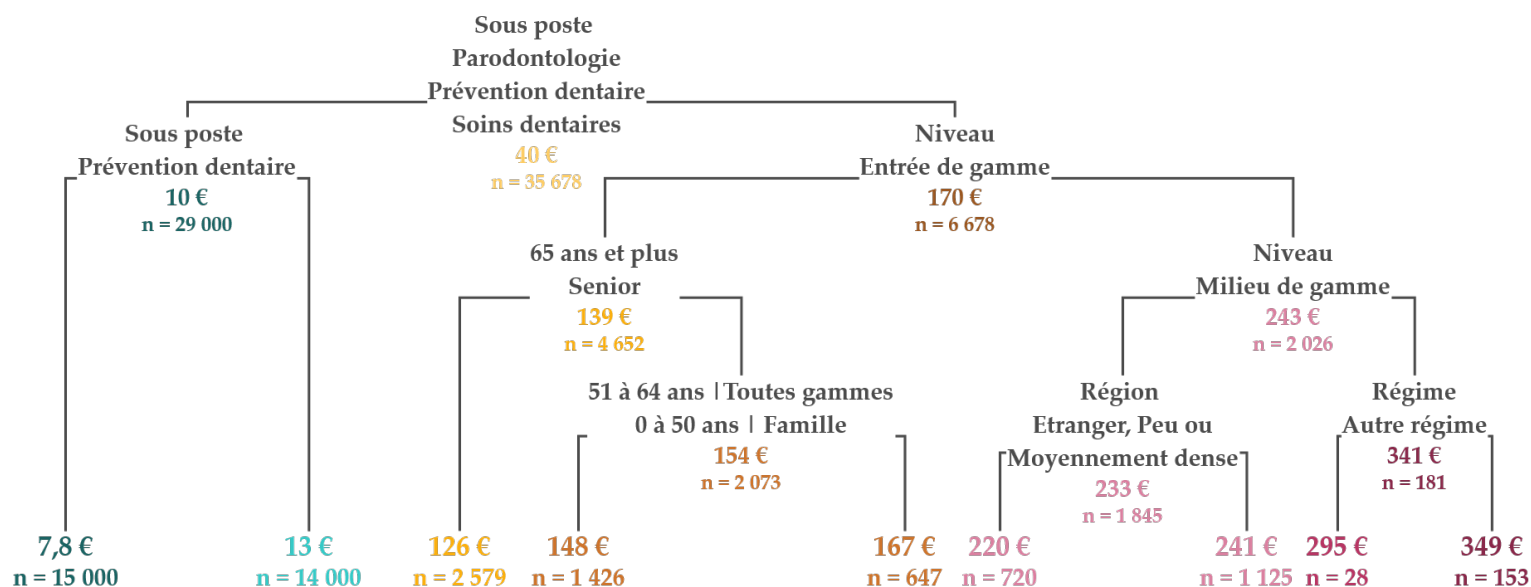


FIGURE 2.12 – Arbre CART obtenu

Comme attendu, c'est le sous poste qui est utilisé dans les premiers nœuds de l'arbre : les prothèses dentaires (couronnes, bridges, ...) affichent un montant de remboursement nettement supérieur à ceux des autres sous postes. De même, les actes de prévention dentaire sont les moins onéreux. D'autre part, la variable combinant l'âge et la gamme de garantie apparaît très souvent lors de la segmentation des individus : pour les prothèses dentaires, les personnes âgées (plus de 65 ans) présentent des remboursements inférieurs à ceux des autres assurés. Enfin, le niveau de garantie est pris en compte avec en moyenne les soins haut de gamme plus chers que les autres et les soins entrée de gamme moins chers. Globalement, la segmentation obtenue à l'aide de cet arbre CART vient confirmer les hypothèses prises dans le Chapitre 3 de la Partie II.

Cette modélisation permet d’offrir à l’assureur une segmentation assez précise de son portefeuille, basée sur les différentes variables dont il dispose.

Par ailleurs et pour rappel, toutes les variables utilisées lors de la segmentation des individus n’ont pas la même importance. Cette importance permet de déterminer quelles sont les variables qui jouent un rôle principal dans les profils obtenus et lesquelles sont plus accessoires. Les importances des variables pour l’arbre CART présenté à la Figure 2.12 sont les suivantes :

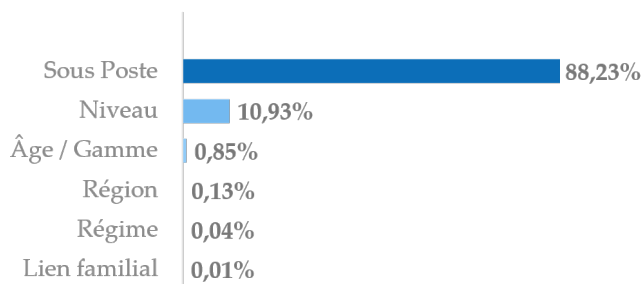


FIGURE 2.13 – Importance des variables dans la modélisation CART

La Figure 2.13 fait très clairement apparaître que le sous poste est la variable la plus importante lors de la segmentation des individus pour le poste dentaire, ce qui s’explique par la grande hétérogénéité du portefeuille en termes de montants de prestations. L’importance du niveau de garantie est également significative. Néanmoins, les importances respectives des autres variables sont inférieures à 1%. Cela signifie que dans les profils finalement obtenus, bien que la totalité des variables permettent d’obtenir des informations plus complètes, seuls le sous poste et le niveau permettent de créer des profils d’individus dont la consommation est réellement éloignée.

D’autre part, l’arbre CART n’est effectivement pas très profond et ne prend pas en compte un grand nombre de variables. Cela peut provenir de la qualité des données, ainsi que du nombre d’observations et de variables relativement faible pour des modèles dits de « Machine Learning ». Aussi, avec des données de meilleure qualité les profils obtenus auraient pu être plus précis, mais il est préférable d’obtenir des profils moins précis et plus fiables que des profils liés à un sur-apprentissage.

De plus, en complément de cette segmentation il est possible de proposer à l’assureur une analyse plus « grossière », issue de l’étude de ce modèle. En effet, le seul montant de remboursement ne suffit pas toujours à identifier si un assuré serait peu ou très enclin à modifier sa consommation de soins de santé suite à la réforme. Il peut arriver que le montant du remboursement semble bas sans pour autant qu’il corresponde à un renoncement aux soins de la part de l’assuré : par exemple, les remboursements de prévention dentaire sont faibles car l’acte en lui-même est peu onéreux ; ou encore les soins sont généralement moins chers dans les zones géographiques qui ne sont pas considérées comme très denses, ce qui impacte le montant du remboursement. Ainsi, la segmentation proposée à la Figure 2.12 permet d’identifier différents profils d’assurés<sup>2</sup> :

2. Les profils sont étudiés pour les prothèses dentaires car pour les autres actes le coût est suffisamment faible pour considérer que le changement de comportement des assurés sera négligeable.



### Très enclin à modifier sa consommation

Niveau de garantie : Entrée de gamme  
Âge : particulièrement les plus de 65 ans  
RC moyen pour une prothèse dentaire : 126 €

### Plutôt enclin à modifier sa consommation

Niveau de garantie : Milieu de gamme  
RC moyen pour une prothèse dentaire : 233 €

### Peu enclin à modifier sa consommation

Niveau de garantie : Haut de gamme  
RC moyen pour une prothèse dentaire : 341 €

Finalement, la segmentation obtenue par modélisation CART ne permet que de confirmer les considérations faites dans le Chapitre 3 de la Partie II : le niveau de garantie ainsi que l'âge de l'assuré sont les principaux facteurs d'un potentiel changement de consommation face à cette réforme « 100% santé ». Néanmoins, elle permet également à l'assureur d'avoir un meilleur aperçu de son portefeuille pour les postes réformés et ainsi de mieux piloter ses offres dans ce domaine. Encore une fois, un nombre de variables plus important initialement aurait peut-être permis à la segmentation d'apporter plus d'informations sur d'éventuels changements de comportement.

Par la méthode utilisée pour créer le modèle CART, la valeur du  $R^2$  est maximale pour la base d'apprentissage choisie, elle vaut ici 69,75%, ce qui est supérieur à la valeur obtenue pour le modèle GLM. Une comparaison plus complète sera faite à la fin de cette partie, une fois que les résultats seront obtenus pour tous les modèles, mais cette constatation permet d'ores et déjà de supposer que la segmentation obtenue avec le modèle CART est satisfaisante en terme de capacité prédictive. La courbe de Lorenz ainsi que l'indice de Gini associé sont présentés à la Figure 2.14.

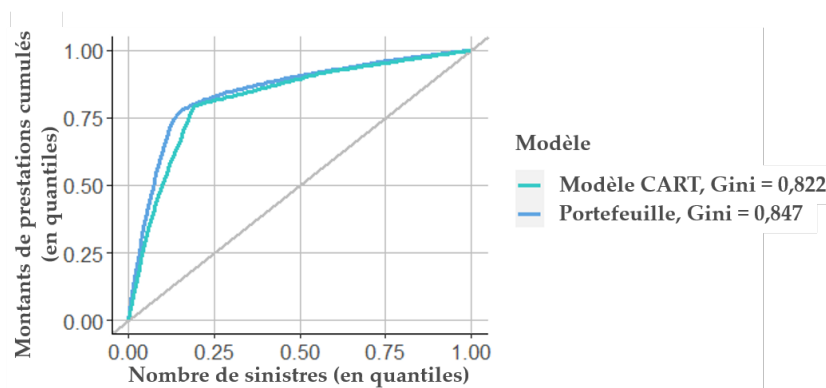


FIGURE 2.14 – Courbe de Lorenz associée au modèle CART

Tout comme pour le modèle GLM, la courbe de Lorenz du modèle CART est très proche de celle des observations du portefeuille, ce qui permet de considérer que les montants de remboursement sont correctement ordonnés par le modèle. Les mêmes conclusions que pour

le modèle GLM peuvent être tirées car les deux modèles fournissent des courbes de Lorenz très proches.

Les moyennes des prédictions du modèle CART sont comparées à celles des observations à la Figure 2.15. Les valeurs sont très proches, le modèle conserve donc bien l'ordre de grandeur des montants de prestations.

PROTHÈSES DENTAIRES	Moyenne	AUTRES ACTES DENTAIRES	Moyenne
Base de test	170€	Base de test	10€
Valeurs prédites	171€	Valeurs prédites	10€

FIGURE 2.15 – Moyennes de la base de test et des prédictions obtenues avec le modèle CART

Enfin, les box-plots des Figures 2.16 et 2.17 présentent les distributions des deux mesures de validation obtenues en implémentant le modèle CART sur 1 000 bases d'apprentissage différentes. Cela permettra par la suite de comparer les performances des différents modèles.

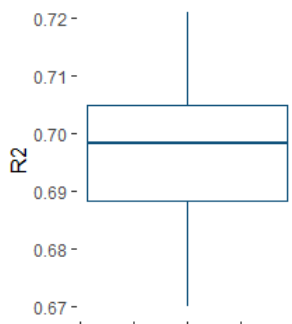


FIGURE 2.16 –  $R^2$  obtenus pour Les modélisations CART

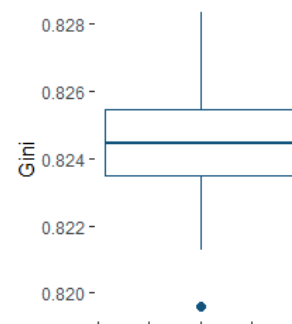


FIGURE 2.17 – Indices de Gini obtenus pour les modélisations CART

## Forêt aléatoire

Comme cela a été évoqué dans la partie théorique, afin de pouvoir juger de la qualité des prédictions du modèle CART qui mène à la création de la segmentation, il est utile de comparer ces prédictions avec celles d'autres modèles tels que les forêts aléatoires, le XG Boost ou encore le GLM présenté précédemment.

Une forêt aléatoire est implémentée sur la base des prestations dentaires. Pour ce faire, il est nécessaire de choisir le nombre d'arbres dans la forêt. Plus le nombre d'arbres est important, plus la prédiction est fiable mais plus le risque de sur-apprentissage augmente. Le choix se pose donc sur 500 arbres car ce nombre est un bon compromis qui permet un apprentissage correct mais évite autant que possible le sur-apprentissage. Il est également possible de choisir le nombre de variables qui seront aléatoirement sélectionnées pour construire chacun des arbres de la forêt. Le choix se porte ici sur 5 variables, cela peut paraître important étant donné que la base est constituée de 7 variables, mais choisir un plus petit nombre risquerait de mener à des arbres incohérents. En effet, comme l'étude du modèle CART l'a

prouvé, les importances des variables ne sont pas homogènes, ne sélectionner que 3 ou 4 variables pour créer un arbre pourrait mener trop souvent à des arbres ne prenant en compte ni le niveau de garantie ni le sous poste.

La forêt aléatoire modélisée de la sorte pour le poste dentaire permet d'obtenir un  $R^2$  de 68,75% et un indice de Gini de 0,828. Les moyennes sont présentées à la Figure 2.19 et restent toujours aussi proches.

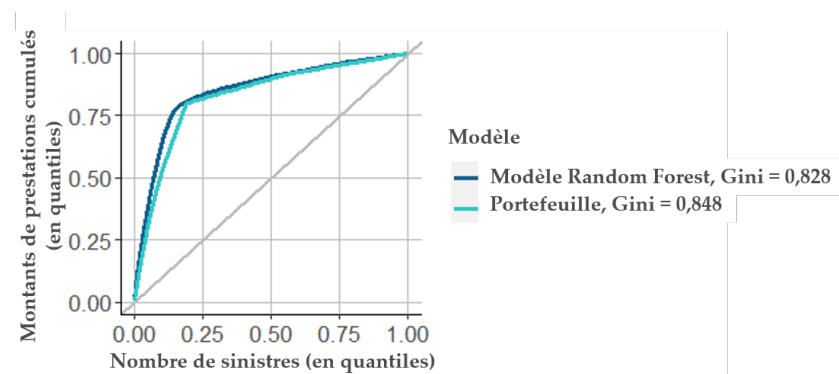


FIGURE 2.18 – Courbe de Lorenz associée au modèle Random Forest

PROTHÈSES DENTAIRES	Moyenne	AUTRES ACTES DENTAIRES	Moyenne
Base de test	170€	Base de test	10€
Valeurs prédites	169€	Valeurs prédites	10€

FIGURE 2.19 – Moyennes de la base de test et des prédictions obtenues avec le modèle Random Forest

La valeur du coefficient  $R^2$  obtenue avec la forêt aléatoire est inférieure à celle de l'arbre CART, ce qui est contre-intuitif. En effet, une forêt aléatoire est en quelque sorte une amélioration d'un arbre CART et donc les prédictions devraient être meilleures (ce qui implique un  $R^2$  plus élevé). Cependant, une des étapes à la base de l'algorithme d'une forêt aléatoire est de choisir aléatoirement les variables parmi lesquelles chaque nœud sera divisé en deux branches. Or, comme l'a montré l'analyse de l'arbre CART pour le poste dentaire, le sous poste joue un rôle majeur dans la segmentation. Ainsi, lorsque la variable sous poste n'est pas choisie pour la création d'un arbre CART, ce dernier est bien plus mauvais que celui effectué avec la variable sous poste. Les prédictions de la forêt aléatoire étant obtenues à partir de la moyenne des prédictions de chaque arbre CART, elles peuvent donc être plus mauvaises qu'un seul arbre CART qui utilise le sous poste.

Enfin, les Figures 2.20 et 2.21 présentent les box-plots des distributions du  $R^2$  et de l'indice de Gini qui seront utilisés afin de comparer les performances des différents modèles.

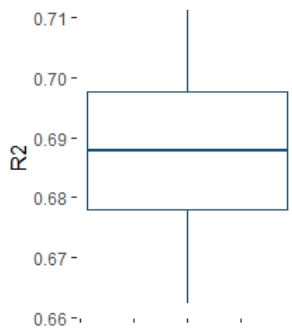


FIGURE 2.20 –  $R^2$  obtenus pour les modélisations Random Forest

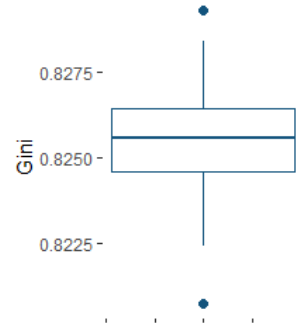


FIGURE 2.21 – Indices de Gini obtenus pour les modélisations Random Forest

## Extrem Gradient Boosting (XG Boost)

La méthode XG Boost peut également permettre de vérifier la qualité de prédiction de l'arbre CART. D'après la partie théorique de ce chapitre, le Gradient Boosting est, comme les forêts aléatoires, basé sur l'agrégation d'arbres. Cependant, la différence entre les deux méthodes est que la construction des arbres par Gradient Boosting prend en compte les erreurs faites par les arbres précédemment construits. Ainsi, cette méthode donne en général de meilleurs résultats que les forêts aléatoires.

L'étude théorique a mis en évidence quatre quantités qu'il est possible de contrôler afin de paramétrer le modèle. La fonction *train* du package *xgboost* permet de considérer plusieurs valeurs pour chaque paramètre en intégrant une phase de cross validation dans la création du modèle : en sortie du modèle se trouve le modèle implémenté avec les paramètres qui optimisent le *MSE* et donc le  $R^2$ . Évidemment, plus le panel de valeurs pour chaque paramètres est large, plus le temps mis par l'algorithme pour trouver le modèle optimal est important. Il faut donc trouver un équilibre entre un nombre suffisant de valeurs pour laisser une marge à l'algorithme, et un nombre restreint de valeurs pour que le temps d'implémentation reste raisonnable. Dans cette étude du poste dentaire, les paramètres suivants sont choisis :

- le nombre  $k$  d'échantillons pour la validation croisée est **15** ;
- le nombre d'itérations du boosting est compris entre **50** et **90**, par pas de 1 ;
- la profondeur maximale des arbres est comprise entre **3** et **6**, par pas de 1 ;
- le nombre de variables choisies aléatoirement est compris entre **70%** et **90%** du nombre total de variables dans la base de données, par pas de 1%.

Le modèle finalement obtenu présente un  $R^2$  de 70,41%, et un indice de Gini de 0,824. Les moyennes sont présentées à la Figure 2.23. Il apparaît donc que ce dernier modèle permet d'obtenir de meilleurs résultats en terme de  $R^2$  : la variance des données est mieux représentée. Cela vient conforter l'idée qu'en général, bien que beaucoup plus opaque en terme de compréhension, le modèle XG Boost est également plus performant.

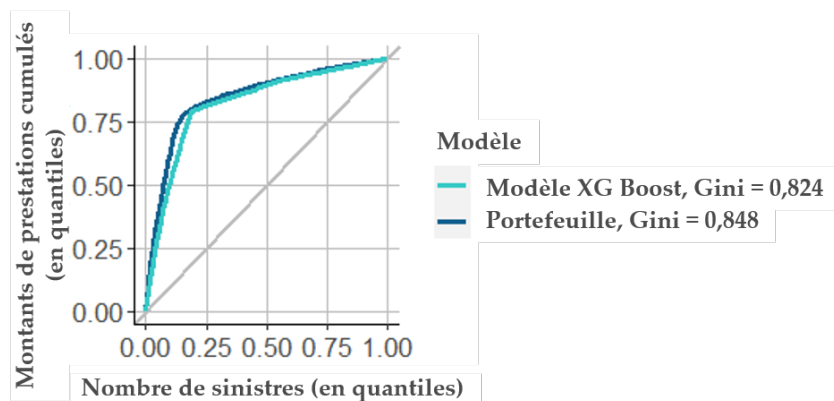


FIGURE 2.22 – Courbe de Lorenz associée au modèle XG Boost

PROTHÈSES DENTAIRES	Moyenne	AUTRES ACTES DENTAIRES	Moyenne
Base de test	170€	Base de test	10€
Valeurs prédites	170€	Valeurs prédites	10€

FIGURE 2.23 – Moyennes de la base de test et des prédictions obtenues avec le modèle XG Boost

Enfin, les Figures 2.24 et 2.25 présentent les box-plots des distributions du  $R^2$  et de l'indice de Gini. Ils sont ajoutés pour la comparaison des modèles.

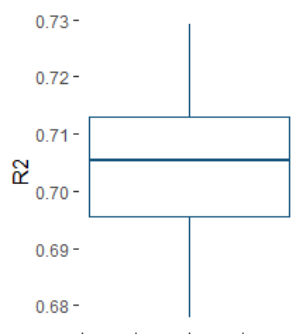


FIGURE 2.24 –  $R^2$  obtenus pour les modélisations XG Boost

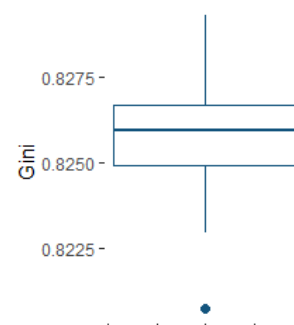


FIGURE 2.25 – Indices de Gini obtenus pour les modélisations XG Boost

## Comparaison des résultats et conclusion

Afin de pouvoir conclure quant à la validité de la segmentation obtenue par la modélisation CART, la comparaison des différents résultats est nécessaire.

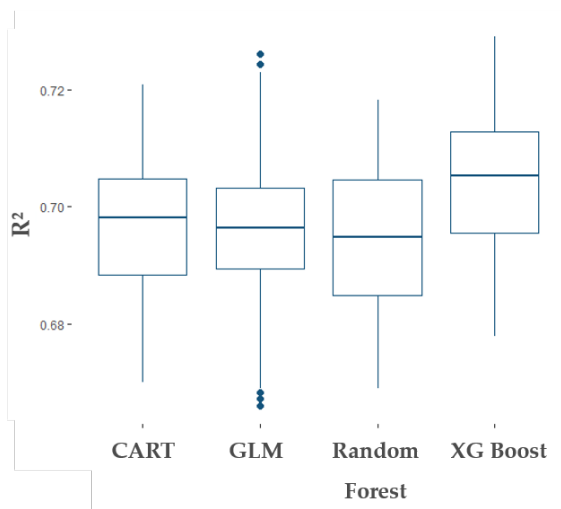


FIGURE 2.26 – Comparaison des  $R^2$  obtenus

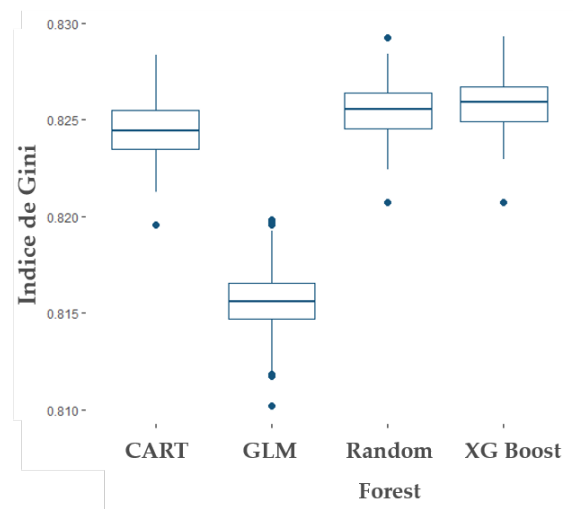


FIGURE 2.27 – Comparaison des indices de Gini obtenus

En ce qui concerne les valeurs du coefficient  $R^2$ , il apparaît que le modèle CART est performant : il est en moyenne supérieur aux modèles GLM et Random Forest, bien qu'également plus volatile que le modèle GLM. Les moins bons résultats du modèle Random Forest ont été expliqués lors de son analyse. Enfin, le modèle CART présente une moins bonne performance que le modèle XG Boost, ce qui était attendu compte tenu des performances habituelles de ce dernier modèle.

Pour rappel, l'intérêt d'avoir mis en place chacun de ces modèles, ainsi que de les comparer, est de conclure quant à la validité de la segmentation obtenue suite au modèle CART. Il apparaît donc que le modèle CART n'est pas le plus optimal au regard des deux indicateurs utilisés dans cette étude. Néanmoins le modèle XG Boost améliore en moyenne la valeur du  $R^2$  de 0,9% et celle de l'indice de Gini de 0,24%, les performances du modèle CART peuvent donc être validées pour le poste dentaire, et la segmentation obtenue par la modélisation CART est considérée comme satisfaisante.

## 5 Segmentation et analyse pour le poste audiologie

Pour rappel, une version détaillée de l'étude, dont le déroulé est identique à celle du poste dentaire, est disponible à l'Annexe E. Cette section présente de façon plus synthétique les résultats obtenus pour le poste audiologie, sans revenir sur tous les aspects techniques de la mise en œuvre de la méthode qui est la même que celle présentée pour le poste dentaire.

Contrairement au poste dentaire, il n'y a pas de sous poste en audiologie. Néanmoins, une autre difficulté survient : le nombre d'individus ayant consommé des soins pour ce poste est bien plus faible que pour les deux autres postes. Les résultats des différents modèles seront donc à rapporter à ce manque de données.

### Modèle linéaire généralisé (GLM)

Pour le poste audiologie, fixer une distribution classique serait une hypothèse incorrecte. En effet, à cause du manque de données, mais aussi du nombre restreint de garanties remboursées par l'assureur, la distribution des prestations présente plusieurs modes qui ne

peuvent pas être approchés par une distribution classique (telle que la loi Gamma, la loi Lognormale ou encore la loi de Weibull). La Figure 2.28 illustre ce problème avec la comparaison entre la distribution réelle des remboursements en audiologie et la distribution Gamma.

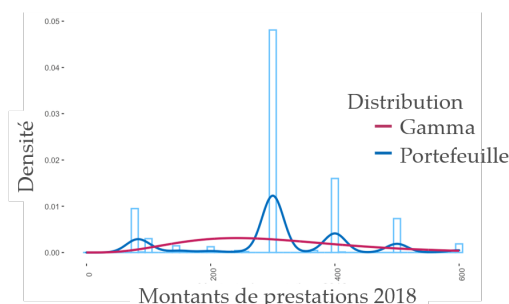


FIGURE 2.28 – Comparaison des distributions

Ainsi, pour le poste audiologie, aucun modèle GLM ne sera implémenté.

## Arbre CART

La même méthode que pour le poste dentaire est appliquée dans le but de modéliser les observations avec un arbre CART. Comme lors de la modélisation pour le poste dentaire, l'arbre CART étudié est celui pour lequel la valeur de  $R^2$  est proche de la moyenne des valeurs des 1000 itérations. L'arbre CART obtenu est moins profond, il ne permet d'obtenir que cinq profils de risque différents. La segmentation obtenue est la suivante :

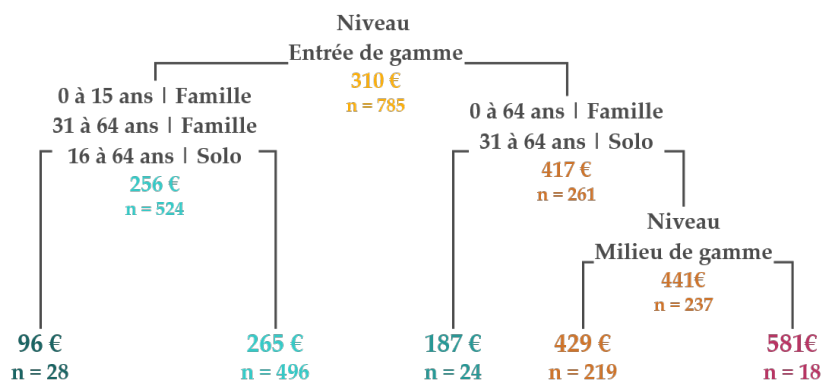


FIGURE 2.29 – Arbre CART obtenu

Cet arbre CART ne met en évidence que deux variables : le niveau de garantie et la variable qui couple l'âge et la gamme de garantie. La faible profondeur de cet arbre est principalement liée au manque de données ainsi qu'au fait qu'initialement, les profils de consommateurs de prothèses auditives ne sont pas nombreux : il y a par exemple un grand nombre de personnes âgées (la gamme, le niveau de garantie et l'âge varient donc peu). Par ailleurs, pour savoir ce qui explique les plus faibles remboursements chez les moins de 65 ans, il faut étudier la table de garanties proposée par l'assureur dans le domaine de l'audiologie. Il apparaît alors que le montant remboursé en entrée de gamme pour des garanties Solo ou Famille est trois fois plus faible que celui pour les Seniors. Le faible remboursement ici est donc lié à une prise en charge plus faible et peut être considéré comme un renoncement aux soins. Ce dernier constat vient néanmoins contredire l'hypothèse prise dans le Chapitre 3

de la Partie II qui considèrerait que la d eformation de fr equence dans la consommation des proth eses auditives serait progressive avec l' age de l'assur e.

Les importances sont pr esent ees dans la Figure 2.30. Seules quatre variables apparaissent car les autres ont une importance inf erieure   0,01%. Cette figure met en avant l'importance des deux variables utilis ees dans la segmentation obtenue par le mod ele CART. Les profils seront donc tous pertinents du point de vue des variables qui les composent.

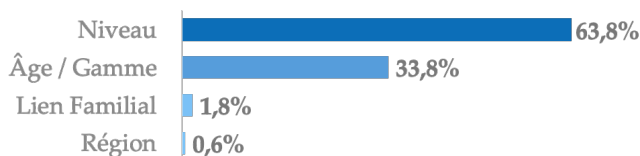


FIGURE 2.30 – Importance des variables dans la mod elisation CART

De plus, en compl ement de cette segmentation, l'analyse plus « grossi ere » issue de l' etude de ce mod ele est la suivante :

#### Tr es enclin   modifier sa consommation

Niveau de garantie : Entr ee de gamme  
 ge : particuli erement les moins de 65 ans  
RC moyen pour une proth ese auditive : 96  

#### Plut ot enclin   modifier sa consommation

Niveau de garantie : Milieu et haut de gamme  
 ge : moins de 65 ans  
RC moyen pour une proth ese auditive : 187  

#### Peu enclin   modifier sa consommation

Niveau de garantie : Milieu et haut de gamme  
 ge : plus de 65 ans  
RC moyen pour une proth ese auditive : 441  

Il appara t que la segmentation obtenue par mod elisation CART permet de confirmer les consid erations faites dans le Chapitre 3 de la Partie II : le niveau de garantie et l' age de l'assur e sont les principaux facteurs d'un potentiel changement de consommation face   cette r eforme « 100% sant e », mais elle met  galement en avant le potentiel renoncement aux soins chez les moins de 65 ans qui n'avait jusqu'alors pas  t  envisag e. Par ailleurs, le nombre d'assur es peu enclins   changer leur consommation, n'est que de 18 sur un total de 785 : il faut s'attendre   un impact important de la r eforme sur la fr equence de consommation des assur ees. Cette segmentation permet  galement   l'assureur d'avoir un meilleur aper u de son portefeuille pour les postes r eform es et ainsi de mieux piloter ses offres dans ce domaine. Encore une fois, un nombre de variable plus important initialement aurait peut- tre permis   la segmentation d'apporter plus d'information sur d' ventuels changements de comportement.



Par ailleurs, le  $R^2$  pour ce modèle vaut 61,50%, et la courbe de Lorenz ainsi que l'indice de Gini se trouvent dans la Figure 2.31. Cette fois-ci, les sinistres dont le remboursement est important (ceux compris entre les quantiles à 0% et 25%) sont correctement ordonnés par le modèle car les deux courbes sont proches. Les moyennes des valeurs observées et des valeurs prédites sont présentées à la Figure 2.32 et sont proches. Le modèle permet donc de prédire correctement l'ordre de grandeur des prestations.

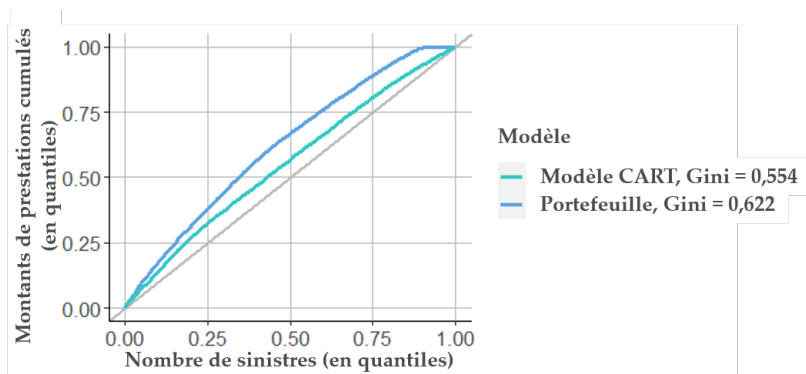


FIGURE 2.31 – Courbe de Lorenz associée au modèle CART

	Moyenne
Base de test	282€
Valeurs prédites	277€

FIGURE 2.32 – Moyennes de la base de test et des prédictions obtenues avec le modèle CART

Enfin, les box-plots des Figures 2.33 et 2.34 présentent les distributions des deux mesures de validation. Cela permettra par la suite de comparer les performances des différents modèles.

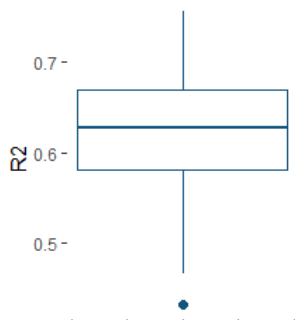


FIGURE 2.33 –  $R^2$  obtenus pour les modélisations CART

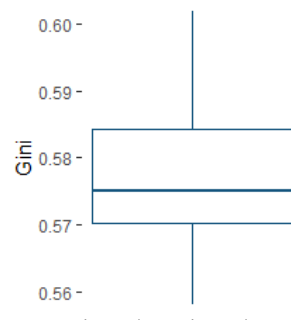


FIGURE 2.34 – Indices de Gini obtenus pour les modélisations CART

## Forêt aléatoire

Tout comme pour le poste dentaire, les modèles de forêts aléatoires et XG Boost sont utilisés afin de vérifier la qualité des prédictions du modèle CART qui mènent à la création de la segmentation.

Le nombre d'observations pour le poste audiologie étant plus faible que pour le poste dentaire, le nombre d'arbres composant la forêt aléatoire est réduit à 200. Cette fois, la valeur du  $R^2$  obtenue par forêt aléatoire est de 63,35%, ce qui est supérieur à celle obtenue avec la modélisation CART. L'indice de Gini est également amélioré avec une valeur de 0,577. L'écart entre les moyennes se réduit par rapport au modèle CART.

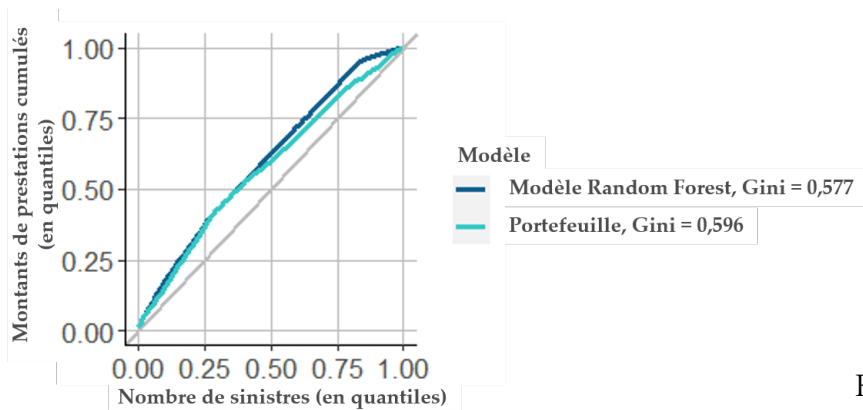


FIGURE 2.35 – Courbe de Lorenz associée au modèle Random Forest

	Moyenne
Base de test	282€
Valeurs prédites	279€

FIGURE 2.36 – Moyennes de la base de test et des prédictions obtenues avec le modèle Random Forest

Enfin, les box-plots des distributions des deux mesures de validation sont les suivants :

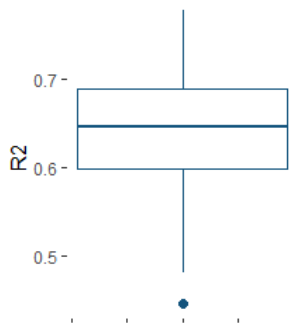


FIGURE 2.37 –  $R^2$  obtenus pour les modélisations Random Forest

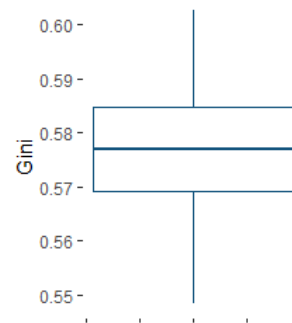


FIGURE 2.38 – Indices de Gini obtenus pour les modélisations Random Forest

## Extrem Gradient Boosting (XG Boost)

La méthode XG Boost peut également permettre de vérifier la qualité de prédiction de l'arbre CART.

Dans cette étude du poste audiologie, les paramètres suivants sont choisis :

- le nombre  $k$  d'échantillons pour la validation croisée est **5** ;
- le nombre d'itérations du boosting est compris entre **40** et **110**, par pas de 1 ;
- la profondeur maximale des arbres est comprise entre **3** et **5**, par pas de 1 ;
- le nombre de variables choisies aléatoirement est compris entre **50%** et **90%** du nombre total de variables dans la base de données, par pas de 1%.

Le modèle finalement obtenu présente un  $R^2$  de 64,30% et un indice de Gini de 0,582. Les moyennes sont présentées à la Figure E13. Ce dernier modèle permet d'obtenir de meilleurs résultats en terme de  $R^2$  : la variance des données est mieux représentée ; mais aussi en terme d'indice de Gini : l'hétérogénéité des observations est également mieux représentée.

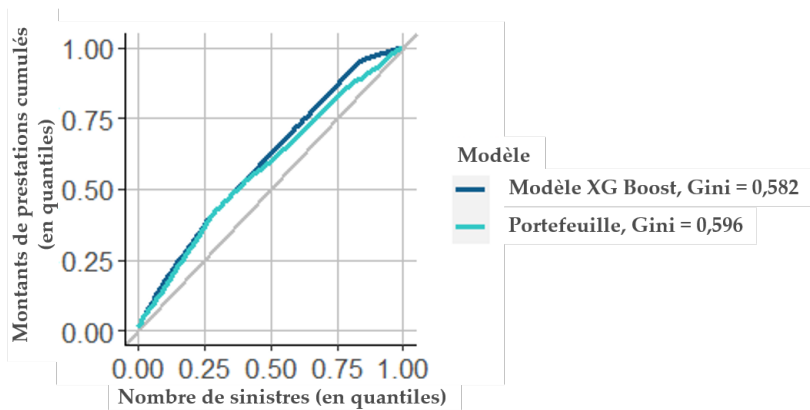


FIGURE 2.39 – Courbe de Lorenz associée au modèle XG Boost

	Moyenne
Base de test	282€
Valeurs prédites	280€

FIGURE 2.40 – Moyennes de la base de test et des prédictions obtenues avec le modèle XG Boost

Enfin, les box-plots des Figures 2.41 et 2.42 présentent les distributions des deux métriques de validation.

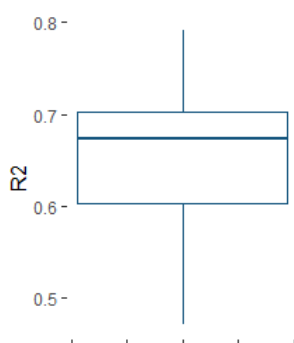


FIGURE 2.41 –  $R^2$  obtenus pour les modélisations XG Boost

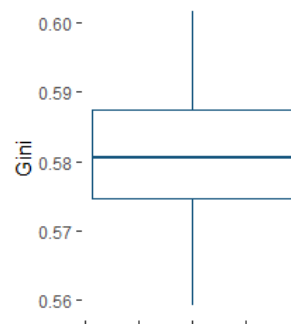


FIGURE 2.42 – Indices de Gini obtenus pour les modélisations XG Boost

## Comparaison des résultats et conclusion

Afin de pouvoir conclure quant à la validité des profils obtenus par la modélisation CART, la comparaison des différents résultats est nécessaire.

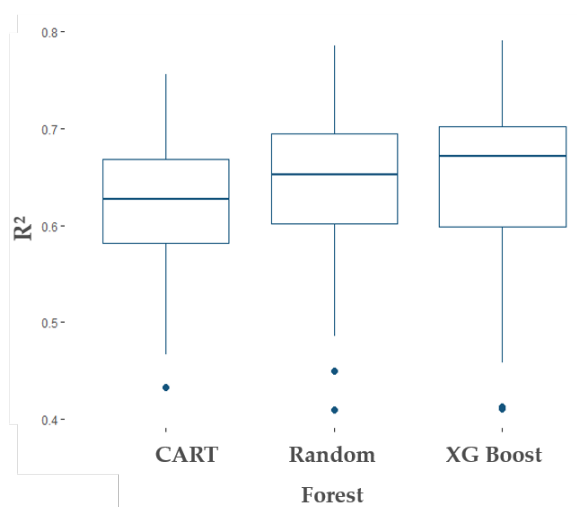


FIGURE 2.43 – Comparaison des  $R^2$  obtenus

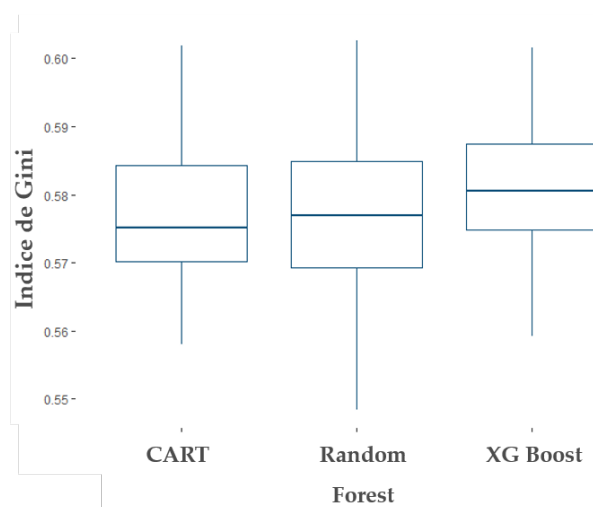


FIGURE 2.44 – Comparaison des indices de Gini obtenus

Les Figures 2.43 et 2.44 montrent que la performance des modèles augmente avec leur complexité. En ce sens, le modèle CART apparaît comme le moins performant au regard des deux mesures utilisées. Le modèle XG Boost est en moyenne meilleur d'environ 4% pour le coefficient  $R^2$ . Cela peut notamment s'expliquer par le trop faible nombre d'observations, ce qui implique que les méthodes itératives se basant sur un grand nombre d'arbres de décision soient plus précises. La segmentation obtenue par la modélisation CART reste relativement simple : elle ne fait intervenir que deux variables et les profils obtenus sont cohérents avec l'expérience du marché de la santé. Elle peut donc tout de même être considérée comme satisfaisante.

## 6 Segmentation et analyse pour le poste optique

Pour rappel, une version détaillée de l'étude, dont le déroulé est identique à celle du poste dentaire, est disponible à l'Annexe F. Cette section présente de façon plus synthétique les résultats obtenus pour le poste optique, sans revenir sur tous les aspects techniques de la mise en œuvre de la méthode qui est la même que celle présentée pour le poste dentaire.

Comme cela a été abordé précédemment lors de la sélection des variables, en optique le sous poste (verre ou monture) ne sera pas conservé.

### Modèle linéaire généralisé (GLM)

Le choix de la distribution utilisée pour modéliser à l'aide d'un GLM les données de remboursement optique se fait de la même manière que pour le dentaire, les montants de prestations étant nécessairement positifs et continus. La loi testée sera la loi Gamma. La Figure 2.45 montre que la distribution de Gamma n'est pas parfaite pour modéliser les prestations : deux pics à 50€ et 90€ sont mal représentés par la loi Gamma par exemple. Néanmoins, il s'agit de la loi la plus proche de la distribution observée du portefeuille et elle sera donc choisie pour le modèle GLM.

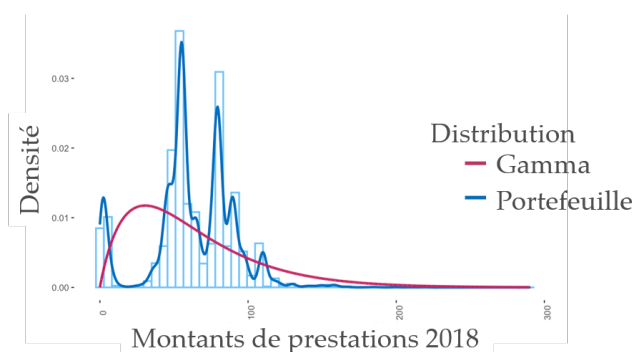


FIGURE 2.45 – Comparaison des distributions

Suite à la modélisation GLM, comme cela a été évoqué dans la partie théorique, une première segmentation « naïve » du portefeuille peut se dessiner en observant les coefficients du modèle pour les différentes modalités des variables. La Figure 2.46 représente ces coefficients pour le modèle GLM.

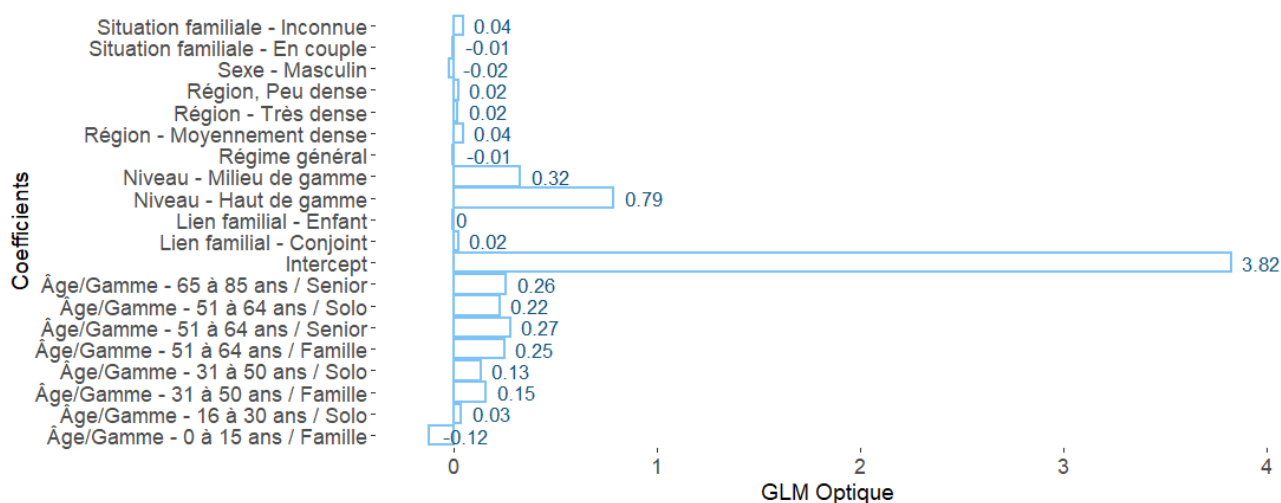


FIGURE 2.46 – Coefficients pour le modèle GLM implémenté

La valeur du coefficient pour l'Intercept est la plus élevée. Il apparaît ensuite par exemple que les prestations sont plus faibles (car le coefficient est négatif) pour les assurés ayant entre 0 et 15 ans dans la gamme Famille que pour les assurés ayant entre 16 et 30 ans dans la gamme Famille (qui sont modélisés dans l'Intercept). Il est alors possible d'identifier que les prestations entrée de gamme (présentes dans l'Intercept) sont inférieures aux prestations milieu et haut de gamme, ce qui est cohérent; ou encore que l'âge joue un rôle important dans la détermination du montant de prestations. Ce graphique permet d'avoir une première approche de segmentation selon le montant des remboursements prédits.

D'autre part, il est nécessaire d'observer la table des déviances en faisant un test ANOVA. D'après ce test, seul la variable Régime Social n'est pas significative pour un seuil de 5%. Cela signifie que la modélisation par un GLM n'utilise pas cette variable pour prédire les montants de prestations.

Par ailleurs, le modèle GLM présente un coefficient  $R^2$  de 23,30% et un indice de Gini valant 0,556.

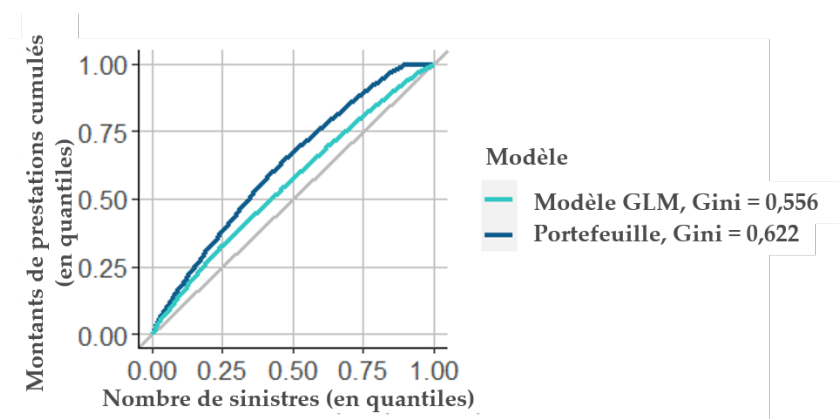


FIGURE 2.47 – Courbe de Lorenz associée au modèle GLM

La courbe de Lorenz du modèle GLM étudié est relativement éloignée de celle des observations du portefeuille, il est possible que les montants de remboursement ne soient pas parfaitement bien répartis par le modèle. Cependant, les deux courbes sont plus proches

entre le quantile à 0% et le quantile à 20% des sinistres (les plus coûteux donc), cela signifie que pour cette tranche l'ordre des prédictions respecte mieux celui des observations.

Les moyennes des valeurs prédites présentées dans la Figure 2.48 sont très proches des valeurs observées, le modèle conserve donc bien l'ordre de grandeur des montants de prestations.

	Moyenne
Base de test	62,35€
Valeurs prédites	62,40€

FIGURE 2.48 – Moyennes de la base de test et des prédictions obtenues avec le modèle GLM

Enfin, les box-plots des Figures 2.49 et 2.50 présentent les distributions du coefficient  $R^2$  et de l'indice de Gini, ce qui permettra par la suite de comparer les performances des modèles.

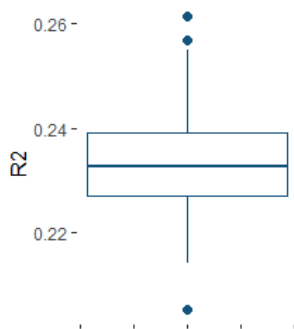


FIGURE 2.49 –  $R^2$  obtenus pour 1000 modèles

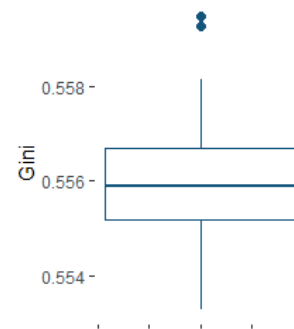


FIGURE 2.50 – Indices de Gini obtenus pour 1000 modèles

## Arbre CART

La méthode des arbres CART va être utilisée afin de créer des profils d'assurés ayant la même consommation de soins, en terme de remboursement complémentaire. La segmentation obtenue est la suivante :

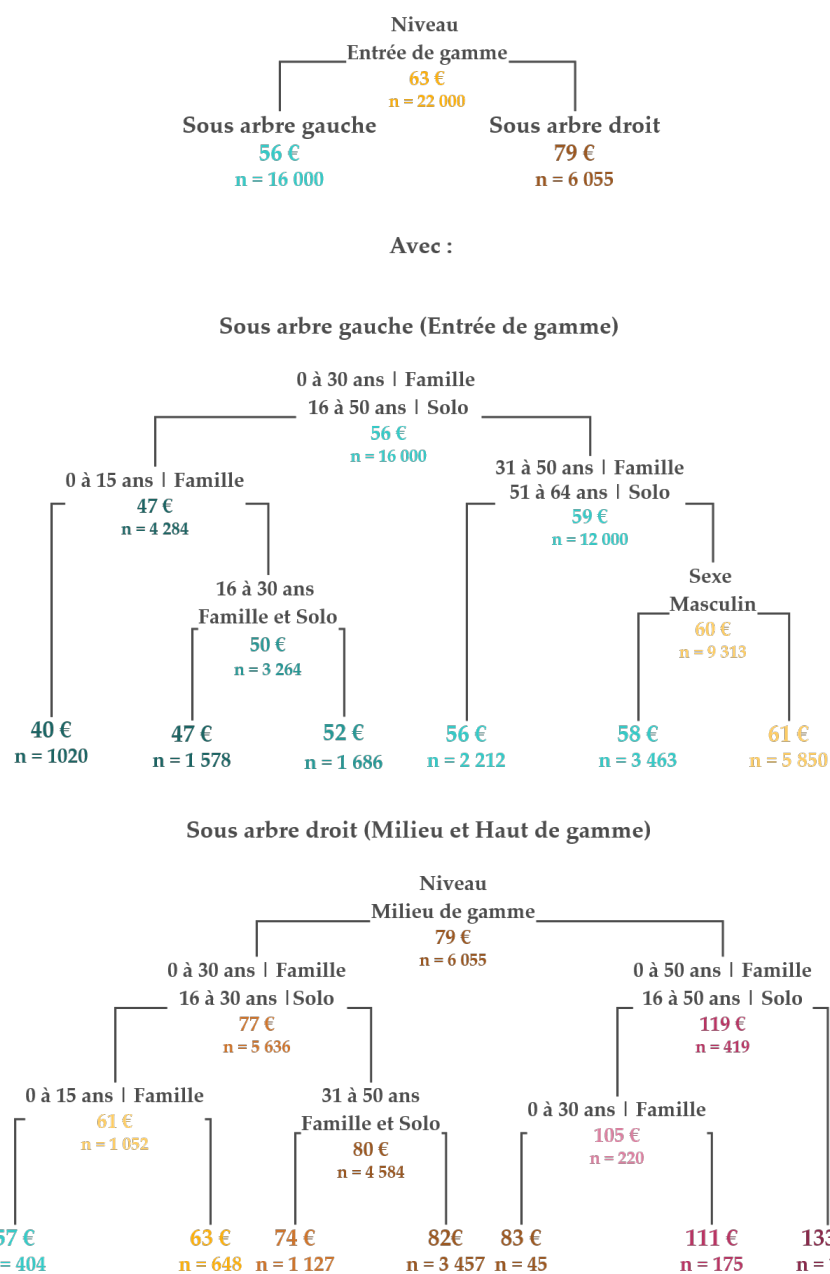


FIGURE 2.51 – Arbre CART obtenu

Cet arbre CART est plus profond que les deux arbres présentés pour les postes dentaire et audiologie, cela s'explique par le plus grand nombre de lignes (par rapport à l'audiologie) et l'absence du sous poste (par rapport au dentaire). Malgré sa profondeur, le modèle CART ne segmente le portefeuille quasiment qu'en utilisant le niveau de garantie ainsi que la variable liant l'âge et la gamme de garantie. Hormis ces deux variables, seul le genre de l'assuré apparaît une fois dans l'arbre. Il est alors possible de lire dans l'arbre présenté à la Figure 2.51 qu'en moyenne les remboursements augmentent avec l'âge de l'assuré, ce qui conforte l'hypothèse prise dans le Chapitre 3 de la Partie II. Cet arbre montre également que plus le niveau de garantie est élevé, plus le remboursement complémentaire l'est. Cela permet d'illustrer le renoncement aux soins. En effet, bien qu'en moyenne la correction des verres optiques soit de plus en plus complexe avec l'âge, les différences qui apparaissent pour une même tranche d'âge selon le niveau de garantie sont importantes, alors qu'il est raisonnable de considérer que les besoins des assurés sont les mêmes.

Les importances sont présentées à la Figure 2.52. Le constat reste le même que pour les postes audiologie et dentaire, l'importance des variables est très hétérogène, mettant en avant seulement les deux variables évoquées précédemment : le niveau de garantie et l'âge lié à la gamme. Il apparaît néanmoins que le lien familial, pourtant absent de l'arbre de la Figure 2.51, a une importance non négligeable. Cela s'explique par le fait que cette variable reste corrélée avec l'âge et la gamme de garantie et donc, bien qu'elle n'apparaisse pas dans l'arbre, elle offrirait une segmentation proche pour certains nœuds de l'arbre. Finalement, cette Figure 2.52 met en avant l'importance des deux variables utilisées dans la segmentation obtenue par le modèle CART. Les profils seront donc pertinents du point de vue des variables qui les composent.

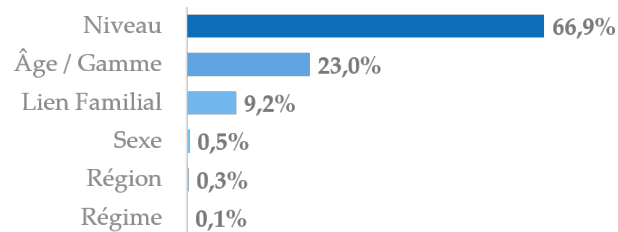


FIGURE 2.52 – Importance des variables dans la modélisation CART

De plus, en complément de cette segmentation l'analyse plus « grossière » issue de l'étude de ce modèle est la suivante :



Finalement, cette segmentation vient confirmer les hypothèses faites dans le Chapitre 3 de la Partie II : le niveau de garantie, ainsi que l'âge de l'assuré sont les principaux facteurs liés à la consommation des assurés. Elle permet également à l'assureur d'avoir un meilleur aperçu de son portefeuille pour les postes réformés et ainsi de mieux piloter ses offres dans ce domaine, elle montre par exemple que 98% des assurés ont un profil d'assuré enclin ou très enclin à modifier sa consommation de soins en optique. Encore une fois, un nombre de variable plus important initialement aurait peut-être permis à la segmentation d'apporter plus d'informations sur d'éventuels changement de comportement.

Par ailleurs, la valeur du  $R^2$  vaut ici 23,44%, ce qui est supérieur à la valeur obtenue pour le modèle GLM. Cela permet d'ores et déjà de supposer que la segmentation obtenue avec le modèle CART est satisfaisante en terme de capacité prédictive. La courbe de Lorenz du



modèle CART est très proche de celle observée pour le modèle GLM, les mêmes conclusions peuvent donc en être tirées. La Figure ?? présente les moyennes des valeurs prédites. Elles sont proches des valeurs observées, le modèle conserve donc bien l'ordre de grandeur des montants de prestations.

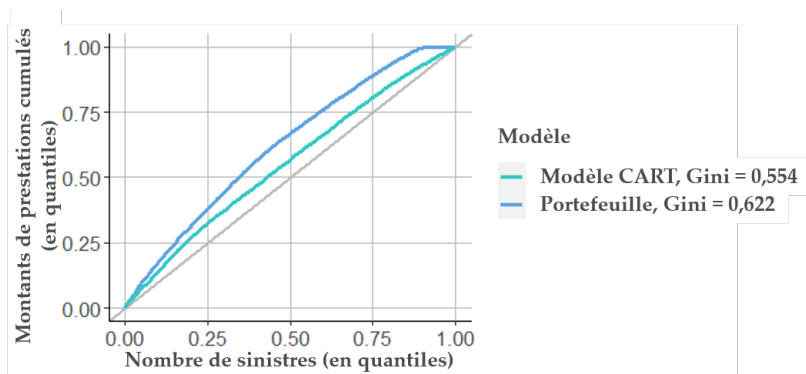


FIGURE 2.53 – Courbe de Lorenz associée au modèle CART

	Moyenne
Base de test	62,35€
Valeurs prédites	62,38€

FIGURE 2.54 – Moyennes de la base de test et des prédictions obtenues avec le modèle CART

Enfin, les box-plots des distributions des mesures de validation sont les suivants :

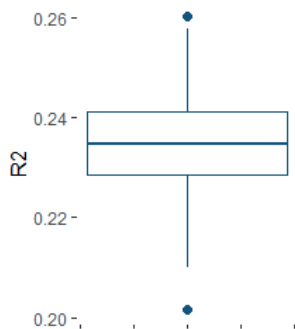


FIGURE 2.55 –  $R^2$  obtenus pour 1000 modèles

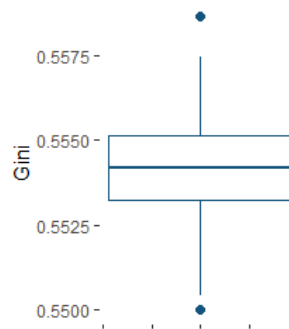


FIGURE 2.56 – Indices de Gini obtenus pour 1000 modèles

## Forêt aléatoire

Tout comme pour le poste dentaire, les modèles de forêts aléatoires XG Boost ainsi que le modèle GLM précédemment présenté sont utilisés afin de vérifier la qualité des prédictions du modèle CART qui mènent à la création de la segmentation.

Une forêt aléatoire de 400 arbres est alors implémentée, et le nombre de variables qui seront aléatoirement sélectionnées pour construire chacun des arbres de la forêt vaut 4.

La forêt aléatoire modélisée de la sorte pour le poste dentaire permet d'obtenir un  $R^2$  de 23,59%, ce qui est supérieur à la valeur obtenue avec le modèle CART, et un indice de Gini de 0,555. Les moyennes des valeurs observées ainsi que de celles prédites sont présentées à la Figure 2.62 et restent toujours aussi proches.

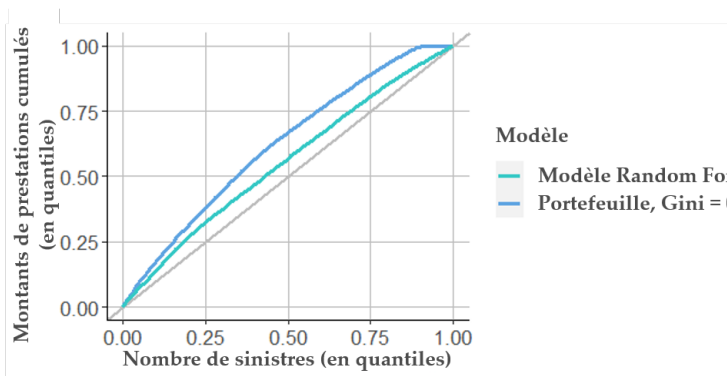


FIGURE 2.57 – Courbe de Lorenz associée au modèle Random Forest

	Moyenne
Base de test	62,35€
Valeurs prédites	62,38€

FIGURE 2.58 – Moyennes de la base de test et des prédictions obtenues avec le modèle Random Forest

Enfin, les Figures 2.59 et 2.60 présentent les box-plots des distributions du  $R^2$  et de l'indice de Gini.

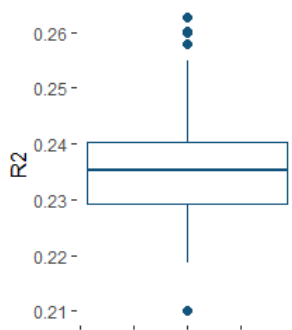


FIGURE 2.59 –  $R^2$  obtenus pour 1000 modèles

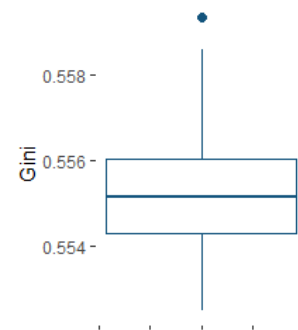


FIGURE 2.60 – Indices de Gini obtenus pour 1000 modèles

## Extrem Gradient Boosting (XG Boost)

La méthode XG Boost peut également permettre de vérifier la qualité de prédiction de l'arbre CART.

Dans cette étude du poste optique, les paramètres suivants sont choisis :

- le nombre  $k$  d'échantillons pour la validation croisée est **5** ;
- le nombre d'itérations du boosting est compris entre **80** et **110**, par pas de 1 ;
- la profondeur maximale des arbres est comprise entre **3** et **5**, par pas de 1 ;
- le nombre de variables choisies aléatoirement est compris entre **50%** et **90%** du nombre total de variables dans la base de données, par pas de 1%.

Le modèle obtenu présente un  $R^2$  de 23,76%, et un indice de Gini de 0,557. Les moyennes sont présentées à la Figure ???. Ce dernier modèle permet d'obtenir de meilleurs résultats en terme de  $R^2$  : la variance des données est mieux représentée ; mais aussi en terme d'indice de Gini : l'hétérogénéité des observation est également mieux représentée.

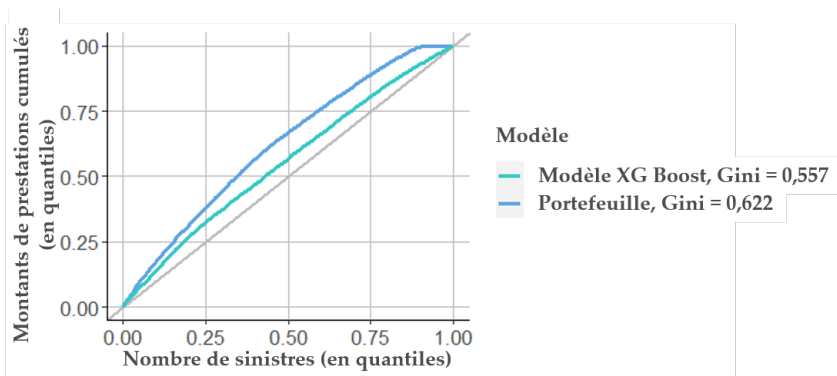


FIGURE 2.61 – Courbe de Lorenz associée au modèle XG Boost

	Moyenne
Base de test	62,35€
Valeurs prédites	62,38€

FIGURE 2.62 – Moyennes de la base de test et des prédictions obtenues avec le modèle XG Boost

Enfin, les box-plots des Figures 2.63 et 2.64 représentent les distributions des mesures de validation.

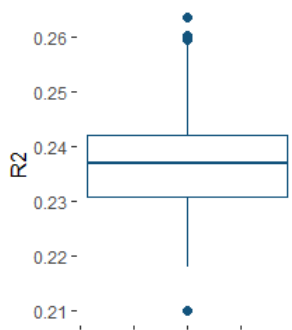


FIGURE 2.63 –  $R^2$  obtenus pour 1000 modèles

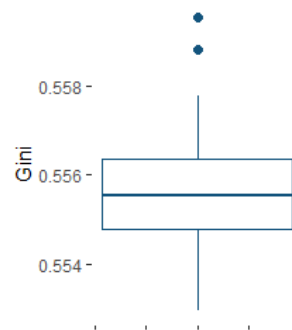


FIGURE 2.64 – Indices de Gini obtenus pour 1000 modèles

## Comparaison des résultats et conclusion

Afin de pouvoir conclure quant à la validité des profils obtenus par la modélisation CART, la comparaison des différents résultats est nécessaire.

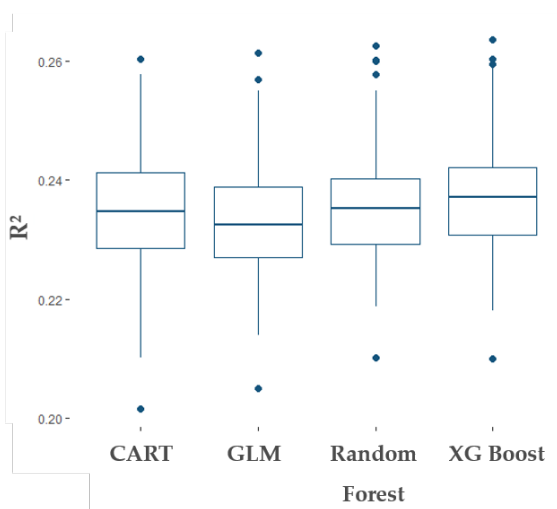


FIGURE 2.65 – Comparaison des  $R^2$  obtenus

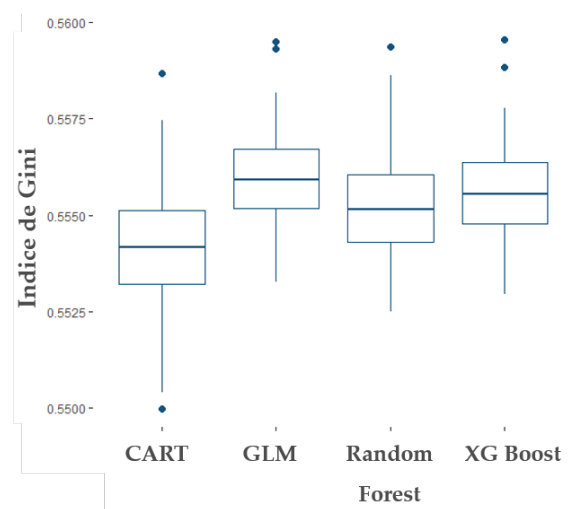


FIGURE 2.66 – Comparaison des Indices de Gini obtenus

---

La Figure 2.65 montre que le modèle GLM est le moins performant pour prédire correctement les montants de prestations, ce qui n'a rien d'étonnant compte tenu que les données du portefeuille ne suivent pas réellement une loi Gamma. Néanmoins, le modèle GLM est également celui qui présente les meilleurs résultats pour l'indice de Gini d'après la Figure 2.66. Cela peut s'expliquer par le fait que bien que la loi Gamma ne suive pas correctement les données, elle permet d'ordonner correctement les prédictions. Cette différence de résultat entre les deux mesures pour le modèle GLM met justement en avant la raison pour laquelle leur étude en parallèle est nécessaire : le  $R^2$  permet de savoir si la prédiction est proche de la valeur initiale et si le modèle rend bien compte de la variance des données, alors que l'indice de Gini informe sur la capacité à prédire les variables « dans le bon ordre ».

En dehors du modèle GLM, les Figures 2.65 et 2.66 montrent que la performance des modèles augmente avec leur complexité. Le modèle CART apparaît comme le moins performant au regard de l'indice de Gini, et seulement plus performant que le modèle GLM au regard de  $R^2$ . Le modèle XG Boost est en moyenne meilleur de plus de 1,3% pour le coefficient  $R^2$ , ce qui est raisonnable. La segmentation obtenue par la modélisation CART reste relativement simple : elle ne fait intervenir que deux variables et les profils obtenus sont cohérents avec l'expérience du marché de la santé. Elle peut donc tout de même être considérée comme satisfaisante.

## En résumé

Pour les trois postes, l'algorithme XG Boost est celui qui permet d'obtenir les meilleures valeurs pour les deux métriques utilisées dans l'étude. Néanmoins, l'amélioration qu'il apporte par rapport à un arbre CART est suffisamment faible pour pouvoir considérer que les profils extraits des arbres sont fiables.

Pour les trois postes étudiés, la segmentation obtenue suite à la modélisation CART est satisfaisante dans le sens où elle conforte la plupart du temps les hypothèses prises dans le Chapitre 3 de la Partie II. Néanmoins, en audiologie elle laisse supposer un renoncement aux soins chez les moins de 65 ans qui n'avait pas été anticipé.

Ainsi, les assurés les plus enclins à modifier leur consommation suite à la réforme « 100% santé » sont ceux ayant souscrit à une garantie entrée de gamme. Une fracture se fait également entre les assurés ayant moins de 65 ans et les autres : pour l'audiologie, les assurés de moins de 65 ans semblent être plus sujets à un renoncement aux soins, alors que c'est l'inverse pour l'optique (quel que soit le niveau de garantie souscrit). Alors que les assurés peu enclins à modifier leur consommation de soins sont ceux ayant souscrit à des garanties haut de gamme, ce qui est cohérent avec le propos même de la réforme.

De plus, bien que les segmentations obtenues n'apportent que peu d'informations supplémentaires sur les possibles motifs de renoncement aux soins chez les assurés, elles ont permis de vérifier les hypothèses émises lors de l'étude de la Partie II et d'en remettre en question certaines, notamment en audiologie. Elles offrent également à l'assureur une connaissance précise de son portefeuille, au vue des données disponibles. Ce dernier point est primordial pour lui permettre d'appréhender le mieux possible les impacts de la réforme.

Enfin, il n'est pas rare qu'une étude dépende principalement de la qualité des données mises à disposition, mais cela n'a pas empêché l'utilisation de cet outil comme outil de vérification d'hypothèses et d'aide à une meilleure connaissance du portefeuille étudié. Toutefois, l'outil de création et de validation d'une segmentation de portefeuille mis en place dans ce mémoire a vocation à être utilisé chez d'autres assureurs. Ainsi, il est possible que ces derniers puissent l'utiliser et en tirer de meilleures conclusions en fonction de la qualité de leurs données, ainsi qu'une segmentation qui apporterait bien plus d'informations sur le portefeuille. D'ailleurs, cet outil pourrait même être utilisé hors du cadre de la réforme « 100% santé », il suffit de choisir une autre variable autour de laquelle la segmentation doit être faite. Ces méthodes de Machine Learning très peu exploitées en santé peuvent aussi s'avérer être des outils très intéressants en termes de pilotage de portefeuille.

---

# CONCLUSION

Ce mémoire, dont l'objectif est de fournir aux assureurs un outil d'aide au pilotage de leur portefeuille lors du déploiement de la réforme « 100% santé », s'articule autour de deux principaux axes : l'étude de l'impact pur de la réforme et celle du comportement des assurés face à la mise en place de cette réforme.

Tout d'abord, la mesure de l'impact pur de la réforme « 100% santé » suite à une mise en As-if des prestations 2018 du portefeuille client a montré qu'il était de + 2,64% au total. Néanmoins, la réforme a des impacts très différents selon le poste de soins concerné. En effet, en audiologie, l'impact est de + 80% alors qu'en dentaire l'impact est plus faible (+ 32%) et qu'en optique, il est même négatif (- 6%). Cette hétérogénéité dans les résultats s'explique par les volontés du gouvernement qui sont différentes selon les postes de soins. En effet, afin de ne pas faire porter tout le poids de la réforme sur les remboursements de l'Assurance Maladie et des complémentaires santé, les plafonds limites de vente permettent de réduire les frais réels, et le plafonnement des remboursements pour le panier libre permet de réduire une partie des prestations de l'assureur, impliquant un impact négatif en optique. À l'inverse, en audiologie, le renoncement aux soins touche une très grande partie des assurés et le gouvernement a donc imposé une forte augmentation des remboursements malgré une réduction des prix de vente.

Ensuite, bien qu'il n'existe pas encore d'historique permettant de mesurer le changement de consommation des assurés, il est possible de poser des hypothèses de déformation de cette consommation afin d'anticiper l'impact global de la réforme « 100% santé » (ou du moins un intervalle d'impacts). Pour commencer, ces hypothèses sont fixées selon la gamme et le niveau de garantie pour chacun des postes réformés et sont déclinées en trois scénarios. L'impact global est alors compris entre +5,6% et +8,9% selon qu'il s'agisse du scénario minorant ou majorant. Le scénario central fixe l'impact global à +6,7%. D'autre part, une étude reposant sur la rentabilité de l'assureur a été effectuée : en considérant que l'assureur est capable d'assumer une perte de 15 points sur son S/P (en considérant les primes fixées), quels sont les taux de déformations qui mèneraient à un tel impact ? Les taux de déformations obtenus sont très largement supérieurs aux taux de renoncement aux soins observés en France par le gouvernement.

Enfin, le changement de comportement des assurés ne dépend pas uniquement du produit d'assurance qu'ils ont choisi, mais également des caractéristiques mêmes de l'assuré (son âge, son genre, la région où il vit ou encore le régime social auquel il appartient). Ainsi, afin de mieux appréhender le profil des assurés enclins à modifier leur consommation suite à la réforme « 100% santé », des segmentations par poste sont effectuées à l'aide d'arbres CART. L'arbre CART n'est en général pas le modèle le plus robuste, cependant il est ici un très bon compromis entre la volonté d'obtenir les prédictions les plus fiables possibles et celle d'utiliser un algorithme qui offre des profils en sortie. En effet, d'autres algorithmes tels que les forêts aléatoires ou XG Boost sont trop opaques pour avoir en sortie des profils identifiables. Ces derniers algorithmes ont néanmoins été utilisés afin de vérifier la qualité des prédictions des arbres CART et donc celle des profils obtenus. Il est apparu que les performances des arbres CART étaient suffisamment bonnes pour considérer les profils en découlant comme fiables. Cette partie de l'étude n'a révélé que peu de nouvelles informa-

---

tions, ce qui est principalement dû au faible nombre de variables disponibles, mais elle a permis d'une part de vérifier les hypothèses effectuées lors de l'étude des déformations de fréquence, et d'autre part de proposer à l'assureur une analyse plus poussée de son portefeuille, afin qu'il puisse mieux appréhender les conséquences de la réforme « 100% santé ».

Les résultats présentés tout au long de ce mémoire sont spécifiques à la base de données utilisée. En effet, il s'agit d'un portefeuille de contrats individuels, ce qui en fait un portefeuille plus vieillissant que la moyenne nationale. De plus, la base de données repose sur les prestations effectuées par l'assureur, il n'y a donc pas d'informations sur les assurés n'ayant pas consommé de soins médicaux, ce qui aurait pu permettre une approche différente. Ensuite, bien que la qualité des données ait été vérifiée dans l'étude, le nombre d'observations une fois les postes réformés isolés est parfois trop faible (comme en audiologie par exemple) et certaines variables caractéristiques de l'assuré (telle que la catégorie socioprofessionnelle) auraient pu permettre une segmentation plus précise. Ainsi, il serait possible de considérer un module supplémentaire dans l'outil développé : celui de l'ajout de variables externes. En effet, dans le domaine de la santé, certaines bases de données dites Opendata (telles que DAMIR par exemple) apportent de nombreuses statistiques qui auraient pu venir s'ajouter aux données de l'assureur (comme la CSP, le nombre moyen de médecins par région, etc.). Cela pourrait permettre d'obtenir une segmentation plus précise.

Enfin, il est nécessaire d'inclure cette étude dans le cadre plus large de l'actualité du système de santé en France. Tout d'abord, le droit de résiliation sans frais et à tout moment des contrats de complémentaires santé adopté le 27 mars 2019 est à étudier en parallèle de l'impact de la réforme « 100% santé ». En ouvrant le marché à la concurrence, cette loi aura pour conséquence d'empêcher les assureurs d'augmenter le tarif de leurs produits de santé. Ainsi, la hausse des tarifications nécessaire pour faire face à l'augmentation des remboursements liée à la réforme « 100% santé » se verra réduite. D'autre part, la crise sanitaire actuelle a complètement bouleversé le monde de la santé et le comportement des assurés. Le confinement qui a eu lieu en France au cours du second trimestre 2020 a drastiquement réduit la consommation des assurés dans le domaine de l'optique, de l'audiologie et du dentaire, ce qui implique donc probablement un impact moins important que celui calculé dans cette étude, du moins pour l'année 2020. C'est pourquoi l'étude de l'impact de la réforme « 100% santé » doit se poursuivre sur les prochaines années. Que ce soit par l'impact de la crise sanitaire qui se profile peu à peu, ou par l'obtention de nouvelles données sur la consommation des assurés, il sera indispensable de réévaluer de façon plus précise sur le long terme l'impact de la réforme « 100% santé » et la segmentation des portefeuilles assurés.

---

# Bibliographie

- [1] ASSURANCE MALADIE (6 juillet 2018) *Dossier de presse - Lutte contre le renoncement aux soins : un dispositif généralisé à l'ensemble du territoire métropolitain.*
- [2] BENTATA P. et BOUZOU N. (Octobre 2018) *Impacts économiques de la Réforme 100% santé sur les acteurs de l'optique en France.* Étude économique réalisée pour le compte du GIFO.
- [3] BOYER C. (2019) *Cours d'apprentissage statistique.* ISUP.
- [4] BREIMAN L., FRIEDMAN J., STONE C.J., & OLSHEN R. (1999) *Classification and regression trees.* CRC press.
- [5] CARTE BLANCHE (03 octobre 2018) *100% santé - La réforme et ses impacts.*
- [6] CHARPENTIER A. (2014) *Computational actuarial science with R.* CRC press.
- [7] CHEN T. & GUESTRIN C. (2016) *XGBoost : A Scalable Tree Boosting System.* eprint arXiv :1603.02754v3 [cs.LG].
- [8] DRESS (2016) *La complémentaire santé - Acteurs, bénéficiaires, garanties.*
- [9] DRESS (2016) *Les dépenses de santé en 2015 - Résultats des comptes de la santé en 2015.*
- [10] DRESS (2018) *Les dépenses de santé en 2017.*
- [11] DRESS (2019) *La complémentaire santé - Acteurs, bénéficiaires, garanties.*
- [12] GOWER J.C. (1971) *A general coefficient of similarity and some of its properties.* Biometrics.
- [13] INSTITUT DES ACTUAIRES (15 juin 2015) *Norme de Pratique Actuarielle 1.*
- [14] IRDES (Novembre 2011) *Questions d'économie de la santé n°170.*
- [15] JOURNAL OFFICIEL N°0010 (12 janvier 2019) *texte n010.*
- [16] LOI N° 89-1009 - ARTICLE 1 (31 décembre 1989)
- [17] MALOT-TULEAU C. (2006-2007) *Méthodes CART - Introduction à la sélection de variables.* Cours de Master MASS à l'Université de Nice Sophia-Antipolis.
- [18] MINISTÈRE DES SOLIDARITÉ ET DE LA SANTÉ (13 juin 2018) *Dossier de presse 100% Santé.*
- [19] MINISTÈRE DES SOLIDARITÉ ET DE LA SANTÉ (Décembre 2019) *Dossier de presse 100% Santé.*
- [20] NIANG K. (2017) *Cours d'analyse des données multidimensionnelles.* ISUP.
- [21] OCDE EUROSTAT OMS (2015) *System of Health Accounts.*
- [22] OTTOU P. (2017) *Méthodes d'apprentissage automatique appliquées au provisionnement ligne à ligne en assurance non-vie.* Mémoire de l'université de Dauphine.
- [23] PROGRAMME DU CONSEIL NATIONAL DE LA RÉSISTANCE (1944).
- [24] SANTÉ CLAIRE (26 février 2020) *Communiqué de presse.*
- [25] SANTÉ CLAIRE (Juillet 2020) *Observatoire des parcours de soins.*
- [26] SÉCURITÉ SOCIALE (2014) *Les Chiffres Clés de la Sécurité Sociale.*
- [27] SÉCURITÉ SOCIALE (01 septembre 2019) [www.securite-sociale.fr/home/dossiers/galerie-dossiers/tous-les-dossiers/100-sante-des-soins-pour-tous-1.html](http://www.securite-sociale.fr/home/dossiers/galerie-dossiers/tous-les-dossiers/100-sante-des-soins-pour-tous-1.html).
- [28] VAN DER MAATEN L. et HINTON G. (2008) *Visualizing data using t-SNE.* Journal of machine learning research, 9 (2008) 2579 - 2605
- [29] VIE PUBLIQUE (29 février 2016) *Les dépenses de santé et leur financement.*



---

# Annexes

---

# A | Listes A et B prothèses auditives

## Liste A :

- un système anti-acouphène ;
- une connectivité sans fil permettant un échange de données avec des dispositifs de communication sans fil (fonction télécommande et/ou Bluetooth) ;
- un réducteur de bruit du vent qui permet une atténuation des basses fréquences générées par les turbulences à l'entrée du ou des microphones ;
- une synchronisation binaurale, permettant de synchroniser les traitements du son entre l'oreille droite et gauche le cas échéant ;
- une directivité microphonique adaptative (le nul de captation induit par la directivité en fonction de la localisation de la source de bruit s'adapte automatiquement en fonction de l'azimut de la source) ;
- une bande passante élargie  $\geq 6\ 000$  Hz permettant de capter des sons sur une étendue de fréquences jusqu'à  $6\ 000$  Hz mesurée au coupleur 2cc selon la norme NF EN 60118-0 : 2015 ;
- une fonction « apprentissage de sonie » permettant l'enregistrement des modifications moyennes du volume apportée par l'utilisateur et d'appliquer ces changements soit automatiquement soit par l'intermédiaire de l'audioprothésiste ;
- un réducteur de réverbération assurant une gestion de la dégradation du signal liée aux réverbérations tardives (champs diffus) dans un local, au-delà de ce que peut permettre la directivité.

## Liste B :

- une bande passante élargie  $\geq 10\ 000$  Hz permettant de capter des sons sur une étendue de fréquences de 0 à  $10\ 000$  Hz ;
- au moins 20 canaux de réglages permettant une amplification du son différente sur 20 plages de fréquences non chevauchantes différentes ;
- un réducteur de bruit impulsionnel permettant d'augmenter le confort d'écoute du patient en réduisant les bruits de durée inférieure à 300 ms ;
- une batterie rechargeable et son chargeur branché sur secteur associé, permettant de s'affranchir de l'utilisation de piles traditionnelles.

---

# B | Description des variables composant la base de données de l'étude

## Base Prestations

- IdContrat : Identifiant du contrat
- IdPersonne : Identifiant du bénéficiaire du contrat concerné par la prestation (cela peut être l'assuré, son enfant, son conjoint ou tout autre bénéficiaire qu'il a déclaré dans le contrat).
- Clé actes : clé permettant d'identifier précisément à quel acte se réfère la ligne de prestation. Cette clé permet de lier la table prestation à la table correspondance des actes.
- Clé garantie : clé permettant de connaître la gamme et le niveau de la garantie du contrat.
- Date de soin : date à laquelle le soin a été effectué.
- Parcours de soins : le parcours de soins indique si l'assuré a suivi le parcours conventionné ou non, cela a une influence sur les remboursements des différentes institutions (Assurance Malade et complémentaires santé).
- Code acte : le code acte permet également d'identifier l'acte de la prestation.
- Détail acte : précisions sur l'acte de la prestations. Il s'agit d'une phrase qui permet à l'utilisateur de mieux comprendre la signification du code acte.
- Famille acte : regroupe plusieurs actes dans une même famille (comme par exemple les prothèses dentaires, ou les montures optiques).
- Nombre d'actes : nombre d'actes liées à la prestation (par exemple, s'il y a deux verres optiques identiques le nombre d'acte est à 2).
- Frais réels : montant payé pas l'assuré pour la prestation.
- Base de remboursement : base de remboursement de la Sécurité Sociale
- Taux de remboursement Sécurité Sociale : taux appliqué à la base de remboursement pour calculé le montant réellement payé par la Sécurité Sociale à l'assuré.
- Remboursement complémentaire : montant payé par la complémentaire santé à l'assuré.
- Date de paiement : date à laquelle le paiement de la complémentaire santé a été effectué.

## Base Correspondance des actes

- Clé actes : clé permettant d'identifier précisément à quel acte se réfère la ligne de prestation. Cette clé permet de lier la table prestation à la table correspondance des actes.
- Grand poste : audiologie, dentaire ou optique selon la prestation.
- Sous poste : détail du grand poste (par exemple lentilles ou prothèses auditives).
- Code acte : le code acte permet également d'identifier l'acte de la prestation.
- Détail acte : précisions sur l'acte de la prestations. Il s'agit d'une phrase qui permet à l'utilisateur de mieux comprendre la signification du code acte.

- 
- Famille acte : regroupe plusieurs actes dans une même famille (comme par exemple les prothèses dentaires, ou les montures optiques).
  - Bases de remboursement 2018 à 2023 : bases de remboursement de la Sécurité Sociale appliquées selon l'année.
  - Taux de remboursement Sécurité Sociale : taux appliqué à la base de remboursement pour calculé le montant réellement payé par la Sécurité Sociale à l'assuré.
  - Prix limite de vente de 2019 à 2023 : montant du plafonnement des actes mis en place par la réforme « 100% santé » selon l'année.
  - Panier : panier de soins auquel appartient l'acte (« 100% santé », maîtrisé ou libre).
  - Poids : poids de la ligne de prestation lorsque celle-ci est dupliquée.

### **Base Adhérent**

- IdContrat : Identifiant du contrat
- IdPersonne : Identifiant du bénéficiaire du contrat concerné par la prestation (cela peut être l'assuré, son enfant, son conjoint ou tout autre bénéficiaire qu'il a déclaré dans le contrat).
- Sexe : genre du bénéficiaire du contrat.
- Date de naissance : date de naissance du bénéficiaire du contrat.
- Situation familiale : situation familiale de l'assuré (en couple, célibataire, divorcé, veuf, inconnue).
- Type de bénéficiaire : détail sur le lien entre le bénéficiaire et l'assuré (enfant, conjoint, ...).
- Régime social : régime social de l'assuré (général, Alsace-Moselle, ...).

### **Base Contrat**

- IdContrat : Identifiant du contrat
- Date de début de contrat
- Date de fin de contrat

### **Base Garantie**

- IdContrat : Identifiant du contrat
- IdPersonne : Identifiant du bénéficiaire du contrat concerné par la prestation (cela peut être l'assuré, son enfant, son conjoint ou tout autre bénéficiaire qu'il a déclaré dans le contrat).
- Clé garantie : clé permettant de connaître la gamme et le niveau de la garantie du contrat.
- Gamme : gamme de garantie (Solo, Famille ou Senior).
- Niveau : niveau de garantie (entrée, milieu ou haut de gamme).

### **Base Adresse**

- IdPersonne : Identifiant du bénéficiaire du contrat concerné par la prestation (cela peut être l'assuré, son enfant, son conjoint ou tout autre bénéficiaire qu'il a déclaré dans le contrat).
- Code postal : code postal de résidence du bénéficiaire du contrat.

- 
- Ville : ville de résidence du bénéficiaire du contrat.
  - Type d'adresse : type d'adresse du lieu de résidence du bénéficiaire du contrat (résidence principale ou autre).

---

# C | Analyse des correspondances multiples

L'analyse des correspondances multiples (ou ACM) fait partie des méthodes dites factorielles. Ces dernières reposent sur le principe de la réduction du nombre de variables en les résumant par un petit nombre de composantes synthétiques appelées facteur. Le choix de la méthode factorielle dépend du type de variable. Des variables qualitatives requièrent l'utilisation de l'ACM. L'ACM est une méthode statistique exploratoire qui permet de décrire graphiquement l'information contenue dans de grands tableaux de données représentant  $p$  variables qualitatives sur un ensemble de  $n$  individus.

Tout comme le calcul des corrélations de Pearson, l'ACM nécessite la création du tableau disjonctif, appelé  $X = (X_1 | \dots | X_p)$ . Cependant, pour l'ACM, le tableau disjonctif ne suffit pas. Il faut également calculer le tableau diagonal des effectifs marginaux.

Ce tableau est une matrice diagonale par bloc,  $D = \begin{bmatrix} D_1 & & & \\ & D_2 & & \\ & & \dots & \\ & & & D_p \end{bmatrix}$ . Chaque bloc  $D_i$  est

lui-même une matrice diagonale contenant les différents effectifs des modalités d'une variable. Ces effectifs sont calculés en sommant les éléments de chaque colonne  $X_i$  du tableau disjonctif.

L'ACM requiert également la définition des notions suivantes :

— le tableau des profils lignes :  $\frac{X}{p}$

— le tableau des profils colonnes :  $XD^{-1}$

— le tableau de Burt<sup>3</sup> :  $B = X'X$

— le vecteur des coordonnées des catégories  $a = \begin{bmatrix} a_1 \\ \vdots \\ a_p \end{bmatrix}$ , qui est le vecteur propre de  $(XD^{-1})' \left( \frac{X}{p} \right)$ .

Ainsi,  $(XD^{-1})' \left( \frac{X}{p} \right) a = \frac{1}{p} D^{-1} X' X a = \frac{1}{p} D^{-1} B a = \mu a$  ce qui mène à  $\mu = \frac{1}{np} a' D a$

L'ACM de la matrice  $X$  repose sur la recherche des éléments propres du produit des profils lignes par les profils colonnes, soit  $(XD^{-1})' \left( \frac{X}{p} \right)$ . Il est possible de montrer qu'il est équivalent de faire une ACM sur le tableau de Burt  $B$ . Pour  $B$ , le tableau des profils lignes vaut  $(pD)^{-1} B$  et le tableau des profils colonnes vaut  $B(pD)^{-1}$ . Il s'agit alors de la recherche des éléments propres de  $\left( \frac{1}{p} D^{-1} B \right)^2$  : les vecteurs propres sont identiques et les valeurs propres sont égales à  $\mu^2$ .

---

3. Avec  $A'$  la matrice transposée de  $A$

---

L'interprétation se fait au travers de l'étude des contributions à un axe factoriel d'une modalité.

$$\mu = \frac{1}{np} a' D a = \frac{1}{np} \sum_{j=1}^{\sum m_i} n_j \cdot a_j^2$$

Avec

- $\sum_{i=1}^p m_i$  le nombre total de modalités du tableau ( $m_i$  étant le nombre de modalités de  $X_i$ )
- $n_j$  l'effectif de la modalité  $j$
- $(a_j)^2$  la coordonnée de la modalité  $j$  sur l'axe factoriel.

Ainsi, la contribution de la modalité  $j$  à un axe factoriel se calcule de la façon suivante :

$$CTR(j) = \frac{\frac{n_j}{np} \cdot a_j^2}{\mu}$$

Une modalité est considérée comme intéressante si  $CTR(j) > \frac{n_j}{np}$ .

Il est également possible de sommer les contributions des modalités d'une variable pour obtenir la contribution de ladite variable :

$$CTR(X_i) = \sum_{j=1}^{m_i} CTR(j) = \frac{1}{\mu} \sum_{j=1}^{m_i} \frac{n_j}{np} \cdot a_j^2$$

Grâce à la fonction *dudi.acm* du package *ade4* sous *R*, il est possible d'implémenter une ACM sur la base de données de prestations tout en pondérant les lignes par leur exposition totale (par l'argument *row.w*).

---

## D | Algorithme t-SNE

L'algorithme t-SNE (t-Distributed Stochastic Neighbor Embedding) a pour but de réduire la dimension d'espaces contenant de nombreuses variables pour obtenir au maximum trois dimensions. Il s'agit d'un algorithme très récent qui a été développé en 2008 par G. Hinton et L. van der Maater [28]. Il se base sur une matrice de distance  $D$  entre les individus de la base de données et une métrique (qui est initialement la distance euclidienne mais qui peut être modifiée).

Les distances de la matrice  $D$  sont converties en probabilités conditionnelles  $p_{j|i}$  (appelées score de similitude). Ainsi, la similitude de l'individu  $x_j$  par rapport à  $x_i$  peut être vue comme la probabilité que  $x_j$  soit voisin de  $x_i$  dans le cas où la distribution du voisinage de  $x_i$  était une loi normale centrée en  $x_i$  et de variance  $\sigma_i$ . Aussi, plus les individus  $x_i$  et  $x_j$  seront proches, plus la probabilité  $p_{j|i}$  sera élevée. Alors,  $p_{i|i} = 0$  et pour tout  $j \neq i$  :

$$p_{j|i} = \frac{\exp\left(\frac{-\|x_i - x_j\|^2}{2 \cdot \sigma_i^2}\right)}{\sum_{k \neq i} \exp\left(\frac{-\|x_i - x_k\|^2}{2 \cdot \sigma_i^2}\right)}$$

Une fois l'espace réduit à 2 ou 3 dimensions, la distribution des distances du voisinage est représentée par une loi de Cauchy (c'est-à-dire une loi de Student à 1 degré de liberté). La probabilité conditionnelle que l'individu  $y_j$  soit, dans le nouvel espace, voisin de  $y_i$  est alors :

$$q_{j|i} = \frac{\exp(-\|y_i - y_j\|^2)}{\sum_{k \neq i} \exp(-\|y_i - y_k\|^2)}, \text{ pour } j \neq i \quad \text{et} \quad q_{i|i} = 0$$

Étant donnée la construction de  $p_{j|i}$  et de  $q_{j|i}$ , ces deux valeurs sont égales si l'espace de faible dimension obtenu modélise correctement les similitudes entre les individus de l'espace de départ (de grande dimension). C'est pourquoi l'objectif de l'algorithme t-SNE est de minimiser l'écart entre  $p_{j|i}$  et  $q_{j|i}$ . Pour mesurer la qualité de  $q_{j|i}$ , l'algorithme utilise la divergence de Kullback-Leibler (notée  $KL$ ). La somme des divergences des individus (ou fonction de coût  $C$ ) est minimisée au travers d'une descente de gradient :

$$C = \sum_i KL(P_i \| Q_i) = \sum_i \sum_j p_{j|i} \cdot \log\left(\frac{p_{j|i}}{q_{j|i}}\right)$$

Où

- $P_i$  représente la distribution de probabilité conditionnelle des individus sachant  $x_i$
- $Q_i$  représente la distribution de probabilité conditionnelle des individus sachant  $y_i$

Le gradient se présente alors comme suit :

$$\frac{\delta C}{\delta y_i} = 2 \cdot \sum_j (p_{j|i} - q_{j|i} + p_{i|j} - q_{i|j}) \cdot (y_i - y_j)$$

L'initialisation de la descente de gradient se fait au travers de l'échantillonnage des individus à partir d'une distribution normale, centrée sur l'origine de l'espace et de faible variance.



---

Par ailleurs, il est possible d'accélérer l'optimisation tout en réduisant au maximum les erreurs de minima locaux. Pour ce faire, un terme est ajouté au gradient : le momentum. Plus précisément, afin de déterminer les changements de coordonnées des points de l'espace à chaque itération de la recherche de gradient, le gradient actuel est ajouté à une somme de gradients existants. Ainsi, à l'itération  $t$ , le gradient mis à jour est exprimé par :

$$\Upsilon^{(t)} = \Upsilon^{(t-1)} + \eta \frac{\delta C}{\delta \Upsilon} + \alpha(t) (\Upsilon^{(t-1)} - \Upsilon^{(t-2)})$$

Où

- $\Upsilon^{(t)}$  représente la solution de la descente de gradient à l'itération  $t$
- $\eta$  représente le taux d'apprentissage
- $\alpha(t)$  représente le momentum à l'itération  $t$

La variance  $\sigma_i$  de la loi Normale doit également être fixée. Pour ce faire, l'algorithme prend en paramètre la perplexité qui elle-même découle de l'entropie de Shannon  $H$  :

$$H(P_i) = - \sum_j p_{j|i} \cdot \log_2(p_{j|i}) \quad \text{et} \quad \text{Perp}(P_i) = 2^{H(P_i)}$$

Il est possible d'interpréter la perplexité comme une mesure de lissage sur le nombre de voisins. Les valeurs habituelles de perplexités sont comprises entre 5 et 50. L'algorithme est relativement robuste aux variations de perplexité.

---

# E | Analyse complète pour le poste audiologie

Cette section se réfère à la Section 5 de la Partie III. Elle présente l'étude complète de la même façon que le poste dentaire a été détaillé dans le corps de l'étude.

Dans cette annexe, c'est donc la segmentation du portefeuille par rapport à la consommation des assurés dans le poste audiologie qui est étudiée. Pour rappel, la segmentation est créée à partir des résultats d'un arbre CART, puis la qualité de prédiction de cet arbre est comparée à celle d'autres modèles afin de vérifier que les profils obtenus ont un sens. De plus, afin de consolider les résultats, chaque modèle sera implémenté 1 000 fois.

Contrairement au poste dentaire, il n'y a pas de sous poste en audiologie. Néanmoins, une autre difficulté survient : le nombre d'individus ayant consommé des soins pour ce poste est bien plus faible que pour les deux autres postes. Les résultats des différents modèles seront donc à rapporter à ce manque de données.

## Modèle linéaire généralisé (GLM)

Le modèle GLM est un modèle paramétrique et nécessite donc d'intuiter la distribution de la variable cible sachant les variables explicatives. Pour ce faire, il est possible de comparer graphiquement la distribution des données par rapport aux distributions usuellement utilisées dans des GLM. Néanmoins, comme cela est apparu lors de l'étude pour le poste dentaire, trouver une distribution classique proche de celle des coûts réels peut être très compliqué. Pour le poste audiologie, fixer une distribution classique serait une hypothèse incorrecte. En effet, à cause du manque de données, mais aussi du nombre restreint de garanties remboursées par l'assureur, la distribution des prestations présente plusieurs modes qui ne peuvent pas être approchés par une distribution classique (telle que la loi Gamma, la loi Lognormale ou encore la loi de Weibull). La Figure E1 illustre ce problème avec la comparaison entre la distribution réelle des remboursements en audiologie et la distribution de Gamma.

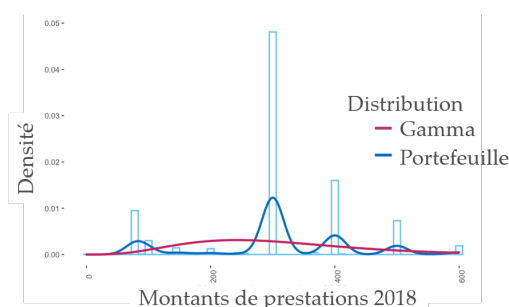


FIGURE E1 – Comparaison des distributions

Ainsi, pour le poste audiologie, aucun modèle GLM ne sera implémenté.

## Arbre CART

La méthode des arbres CART va être utilisée afin de créer des profils d'assurés ayant la même consommation de soins, en terme de remboursement complémentaire. L'arbre CART obtenu est le suivant :

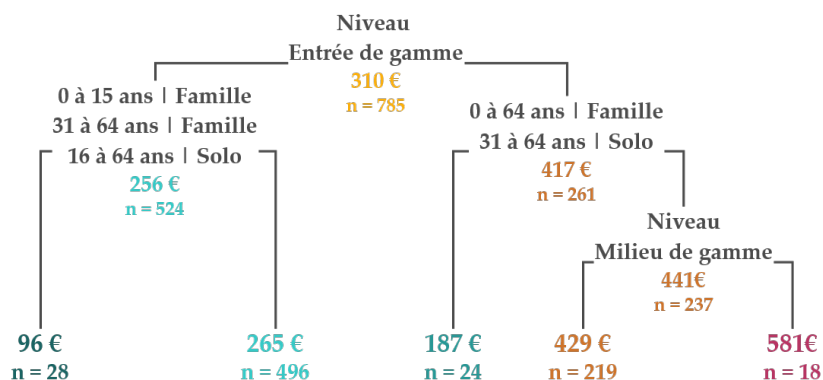


FIGURE E2 – Arbre CART obtenu

Cet arbre CART ne met en évidence que deux variables : le niveau de garantie et la variable qui couple l'âge et la gamme de garantie. La faible profondeur de cet arbre est principalement liée au manque de données ainsi qu'au fait qu'initialement, les profils de consommateurs de prothèses auditives ne sont pas nombreux : il y a par exemple un grand nombre de personnes âgées (la gamme, le niveau et l'âge varient donc peu). Par ailleurs, pour savoir ce qui explique les plus faibles remboursements chez les moins de 65 ans, notamment pour la garantie entrée de gamme (feuille située en bas à gauche de l'arbre de la Figure E2), il faut étudier la table de garanties proposée par l'assureur dans le domaine de l'audiologie. Il apparaît alors que le montant remboursé en entrée de gamme pour des garanties Solo ou Famille est trois fois plus faible que celui pour les Seniors. Le faible remboursement ici est donc lié à une prise en charge plus faible et peut donc être considéré comme un renoncement aux soins. Ce dernier constat vient alors contredire l'hypothèse prise dans le Chapitre 3 de la Partie II qui considérait que la déformation de fréquence dans la consommation des prothèses auditives serait progressive avec l'âge de l'assuré.

Cette modélisation permet d'offrir à l'assureur une segmentation cohérente de son portefeuille, bien que basée sur peu de variables.

Par ailleurs pour rappel, toutes les variables utilisées lors de la segmentation des individus n'ont pas la même importance. Cette importance permet de déterminer quelles sont les variables qui jouent un rôle principal dans les profils obtenus et lesquelles sont plus accessoires. Les importances des variables pour l'arbre CART présenté à la Figure E2 sont les suivantes :

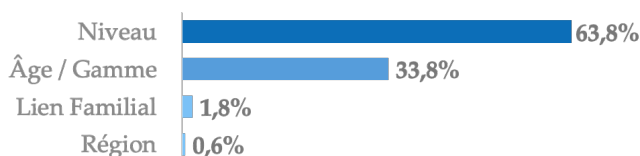


FIGURE E3 – Importance des variables dans la modélisation CART

---

Seules quatre variables apparaissent sur la Figure E3 car les autres ont une importance inférieure à 0,01%. Cette figure met en avant l'importance des deux variables utilisées dans la segmentation obtenue par le modèle CART. Les profils seront donc tous pertinents du point de vue des variables qui les composent.

De plus, en complément de cette segmentation il est possible de proposer à l'assureur une analyse plus « grossière », issue de l'étude de ce modèle. Contrairement à l'étude menée sur le poste dentaire, il semble que pour l'audiologie le montant des remboursements soit bien lié au renoncement aux soins. Ainsi, la segmentation proposée à la Figure E2 permet d'identifier différents profils d'assurés :

**Très enclin à modifier sa consommation**

Niveau de garantie : **Entrée de gamme**  
Âge : **particulièrement les moins de 65 ans**  
RC moyen pour une prothèse auditive : **96 €**

**Plutôt enclin à modifier sa consommation**

Niveau de garantie : **Milieu et haut de gamme**  
Âge : **moins de 65 ans**  
RC moyen pour une prothèse auditive : **187 €**

**Peu enclin à modifier sa consommation**

Niveau de garantie : **Milieu et haut de gamme**  
Âge : **plus de 65 ans**  
RC moyen pour une prothèse auditive : **441 €**

Finalement, la segmentation obtenue par modélisation CART permet de confirmer les considérations faites dans le Chapitre 3 de la Partie II : le niveau de garantie et l'âge de l'assuré sont les principaux facteurs d'un potentiel changement de consommation face à cette réforme « 100% santé », mais elle met également en avant le potentiel renoncement aux soins chez les moins de 65 ans qui n'avait jusqu'alors pas été envisagé. De plus, elle permet également à l'assureur d'avoir un meilleur aperçu de son portefeuille pour les postes réformés et ainsi de mieux piloter ses offres dans ce domaine. Encore une fois, un nombre de variables plus important initialement aurait peut-être permis à la segmentation d'apporter plus d'informations sur d'éventuels changements de comportement.

La méthode utilisée pour créer le modèle CART implique que la valeur du  $R^2$  est maximale pour la base d'apprentissage choisie, elle vaut ici 61,50%. La courbe de Lorenz ainsi que l'indice de Gini associé permettront de comparer les performances du modèle avec celles des autres modèles implémentés dans la suite du chapitre. D'après la Figure E4, la courbe de Lorenz du modèle CART est très proche de celle des observations du portefeuille, ce qui permet de considérer que les montants de remboursement sont correctement répartis par le modèle. Cependant, les deux courbes s'éloignent progressivement entre le quantile à 25% et le quantile à 100% des sinistres. Les sinistres dont le remboursement est important (c'est-à-dire ceux compris entre les quantiles 0% et 25%) sont correctement ordonnés puisque les deux courbes se chevauchent.

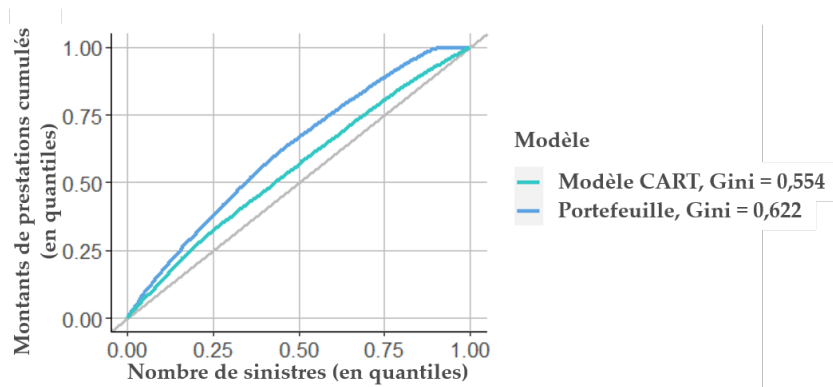


FIGURE E4 – Courbe de Lorenz associée au modèle CART

Par ailleurs, l'indice de Gini ne permet pas de savoir si les prédictions du modèle sont proches des valeurs observées mais seulement de savoir si l'ordre entre les montants des sinistres prédits se rapproche de celui des montants observés. Ainsi, pour s'assurer que les prédictions sont cohérentes en ordre de grandeur, les moyennes doivent être comparées, et ce pour les prothèses dentaires d'une part puis pour les autres actes d'autre part.

	Moyenne
Base de test	282€
Valeurs prédites	277€

FIGURE E5 – Moyennes de la base de test et des prédictions obtenues avec le modèle CART

Les moyennes des valeurs prédites sont proches des valeurs observées, le modèle conserve donc bien l'ordre de grandeur des montants de prestations.

Enfin, les box-plots des Figures E6 et E7 présentent les distributions des deux mesures de validation obtenues en implémentant le modèle CART sur 1 000 bases d'apprentissage différentes, cela permettra par la suite de comparer les performances des différents modèles.

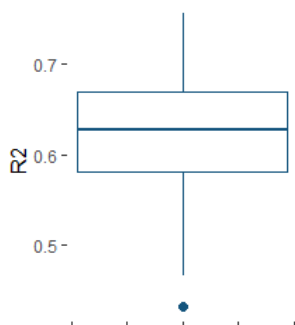


FIGURE E6 –  $R^2$  obtenus pour les modélisations CART

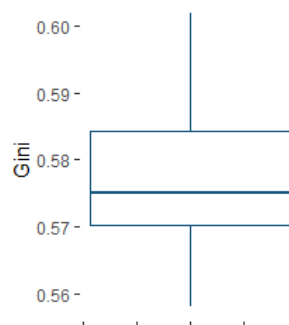


FIGURE E7 – Indices de Gini obtenus pour les modélisations CART

## Forêt aléatoire

Comme cela a été évoqué dans la partie théorique, afin de pouvoir juger de la qualité des prédictions du modèle CART qui mènent à la création de la segmentation, il est utile de

comparer ces prédictions avec celles d'autres modèles tels que les forêts aléatoires ou le XG Boost. Pour rappel, pour le poste audiologie, le modèle GLM ne sera pas étudié car il n'est pas raisonnable de poser l'hypothèse d'une distribution sur les données.

Une forêt aléatoire est implémentée sur la base des prestations d'audiologie. Pour ce faire, il est nécessaire de choisir le nombre d'arbres dans la forêt. Plus le nombre d'arbres est important, plus la prédiction est fiable mais plus le risque de sur-apprentissage augmente. Le choix se pose donc sur 200 arbres car ce nombre est un bon compromis qui permet un apprentissage correct mais évite autant que possible le sur-apprentissage. Il est plus faible que pour le poste dentaire, ce qui s'explique par un nombre de lignes beaucoup plus faible dans la base de données en audiologie. D'autre part, il est également possible de choisir le nombre de variables qui seront aléatoirement sélectionnées pour construire chacun des arbres de la forêt, le choix se porte ici sur 4 variables.

La forêt aléatoire modélisée de la sorte pour le poste audiologie permet d'obtenir un  $R^2$  de 63,35% et un indice de Gini de 0,577. Les moyennes sont présentées à la Figure E9 et restent toujours aussi proches. La valeur du coefficient  $R^2$  obtenue avec la forêt aléatoire est supérieure à celle obtenue avec le modèle CART. Cela s'explique principalement par le peu de données à disposition qui rend l'arbre CART plus instable.

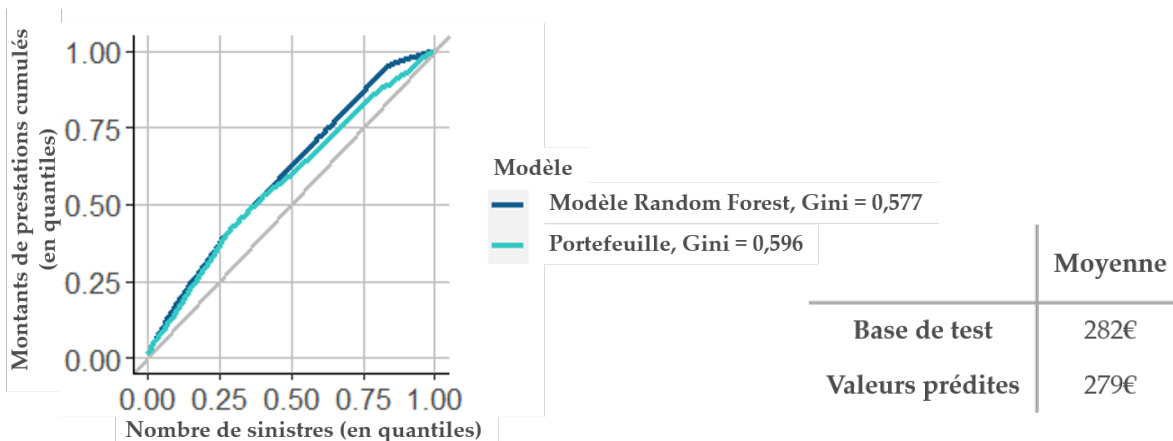


FIGURE E8 – Courbe de Lorenz associée au modèle Random Forest

FIGURE E9 – Moyennes de la base de test et des prédictions obtenues avec le modèle Random Forest

Enfin, les Figures E10 et E11 présentent les box-plots des distributions du  $R^2$  et de l'indice de Gini qui seront utilisés afin de comparer les performances des différents modèles.

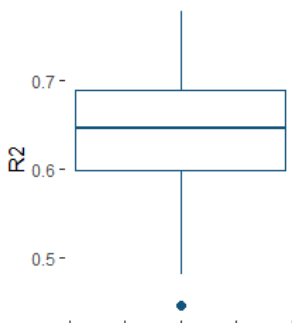


FIGURE E10 –  $R^2$  obtenus pour les modélisations Random Forest

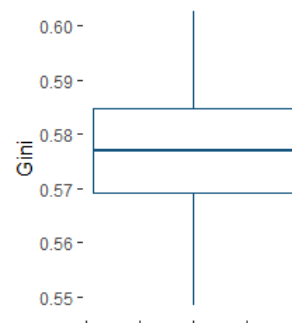


FIGURE E11 – Indices de Gini obtenus pour les modélisations Random Forest

## Extrem Gradient Boosting (XG Boost)

La méthode XG Boost peut également permettre de vérifier la qualité de prédiction de l'arbre CART. D'après la partie théorique de ce chapitre, le Gradient Boosting est, comme les forêts aléatoires, basé sur l'agrégation d'arbres. Cependant, la différence entre les deux méthodes est que la construction des arbres par Gradient Boosting prend en compte les erreurs faites par les arbres précédemment construits. Ainsi, cette méthode donne en général de meilleurs résultats que les forêts aléatoires.

L'étude théorique a mis en évidence quatre quantités qu'il est possible de contrôler afin de paramétrer le modèle. La fonction *train* du package *xgboost* permet de considérer plusieurs valeurs pour chaque paramètre en intégrant une phase de cross validation dans la création du modèle : en sortie du modèle se trouve le modèle implémenté avec les paramètres qui optimisent le *MSE* et donc le  $R^2$ . Évidemment, plus le panel de valeurs pour chaque paramètres est large, plus le temps mis par l'algorithme pour trouver le modèle optimal est important. Il faut donc trouver un équilibre entre un nombre suffisant de valeurs pour laisser une marge à l'algorithme, et un nombre restreint de valeurs pour que le temps d'implémentation reste raisonnable. Dans cette étude du poste audiologie, les paramètres suivants sont choisis :

- le nombre  $k$  d'échantillons pour la validation croisée est 5 ;
- le nombre d'itérations du boosting est compris entre 40 et 110, par pas de 1 ;
- la profondeur maximale des arbres est comprise entre 3 et 5, par pas de 1 ;
- le nombre de variables choisies aléatoirement est compris entre 50% et 90% du nombre total de variables dans la base de données, par pas de 1%.

Le modèle finalement obtenu présente un  $R^2$  de 64,30%, et un indice de Gini de 0,582. Les moyennes sont présentées à la Figure ?? . Il apparaît donc que ce dernier modèle permet d'obtenir de meilleurs résultats en terme de  $R^2$  : la variance des données est mieux représentée, mais aussi en terme d'indice de Gini : l'hétérogénéité des observations est également mieux représentée. Cela vient conforter l'idée qu'en général, bien que beaucoup plus opaque en terme de compréhension, le modèle XG Boost est également plus performant.

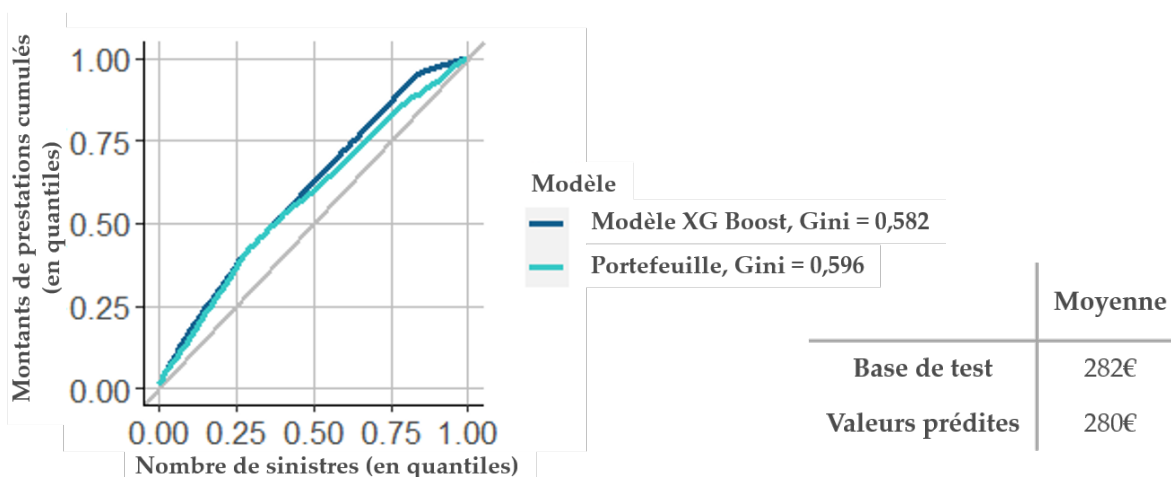


FIGURE E12 – Courbe de Lorenz associée au modèle XG Boost

FIGURE E13 – Moyennes de la base de test et des prédictions obtenues avec le modèle XG Boost

Enfin, les Figures E14 et E15 présentent les box-plots les distributions des deux métriques de validation. Ils sont ajoutés pour la comparaison des modèles.

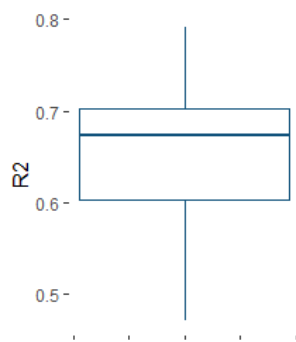


FIGURE E14 –  $R^2$  obtenus pour les modélisations XG Boost

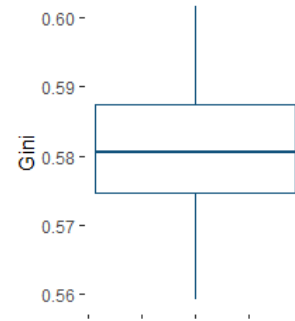


FIGURE E15 – Indices de Gini obtenus pour les modélisations XG Boost

## Comparaison des résultats et conclusion

Afin de pouvoir conclure quant à la validité des profils obtenus par la modélisation CART, la comparaison des différents résultats est nécessaire.

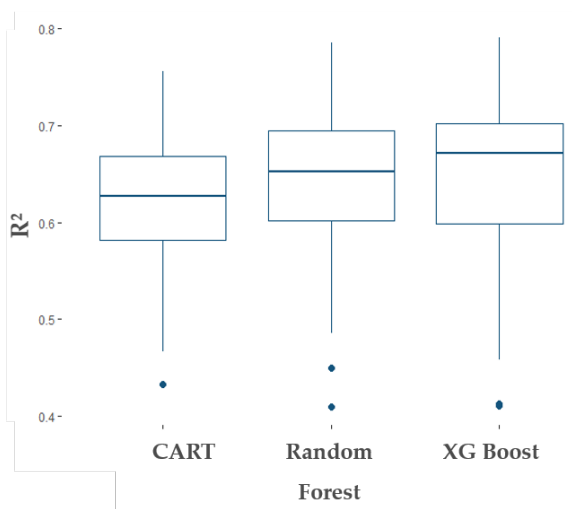


FIGURE E16 – Comparaison des  $R^2$  obtenus

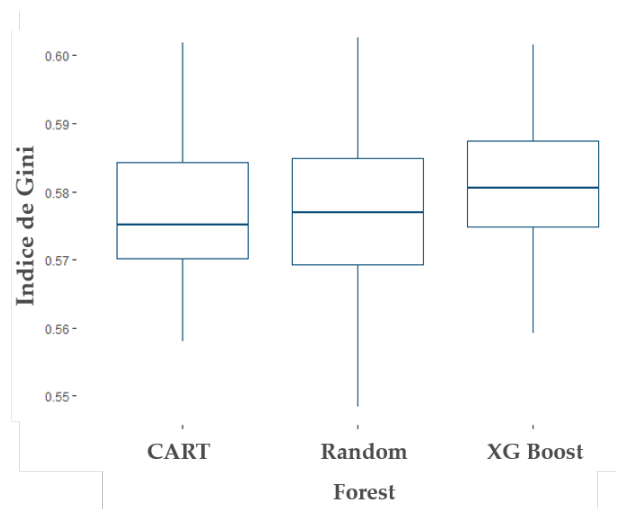


FIGURE E17 – Comparaison des coefficients de Gini obtenus

Les Figures E16 et E17 montrent que la performance des modèles augmente avec leur complexité. En ce sens, le modèle CART apparaît comme le moins performant au regard des deux mesures utilisées. Le modèle XG Boost est en moyenne meilleur d'environ 4% pour le coefficient  $R^2$ . Cela peut notamment s'expliquer par le trop faible nombre d'observations, ce qui implique que les méthodes itératives qui se basent sur un grand nombre d'arbres de décision soient plus précises. La segmentation obtenue par la modélisation CART reste relativement simple : elle ne fait intervenir que deux variables et les profils obtenus sont cohérents avec l'expérience du marché de la santé. Elle peut donc tout de même être considérée comme satisfaisante.



---

# F | Analyse complète pour le poste optique

Cette section se réfère à la Section 6 de la Partie III. Elle présente l'étude complète de la même façon que le poste dentaire a été détaillé dans le corps de l'étude.

Dans cette annexe, c'est donc la segmentation du portefeuille par rapport à la consommation des assurés dans le poste optique qui est détaillée. Pour rappel, la segmentation est créée à partir des résultats d'un arbre CART, puis la qualité de prédiction de cet arbre est comparée à celle d'autres modèles afin de vérifier que les profils obtenus ont un sens. De plus, afin de consolider les résultats, chaque modèle sera implémenté 1 000 fois.

Comme cela a été abordé lors de la sélection de variables, en optique le sous poste (verre ou monture) ne sera pas conservé.

## Modèle linéaire généralisé (GLM)

Le modèle GLM est un modèle paramétrique et nécessite donc d'intuiter la distribution de la variable cible sachant les variables explicatives. Pour ce faire, il est possible de comparer graphiquement la distribution des données par rapport aux distributions usuellement utilisées dans des GLM. Les montants de prestations étant nécessairement positifs et continus, la première loi testée sera la loi Gamma. Pour chacune des distributions Gamma, les paramètres de la loi sont calculés en fonction de l'espérance et de la variance des observations.

La Figure F1 montre que la distribution de Gamma n'est pas parfaite pour modéliser les prestations : deux pic à 50€ et 90€ sont mal représentés par la loi Gamma par exemple. Néanmoins, il s'agit de la loi la plus proche de la distribution observée du portefeuille et elle sera donc choisie pour les deux modèles GLM. En effet, il est rare que les données d'observation suivent parfaitement une loi connue, il est donc nécessaire de poser l'hypothèse lourde de la distribution puis de ne pas l'oublier lors de l'analyse des résultats par la suite.

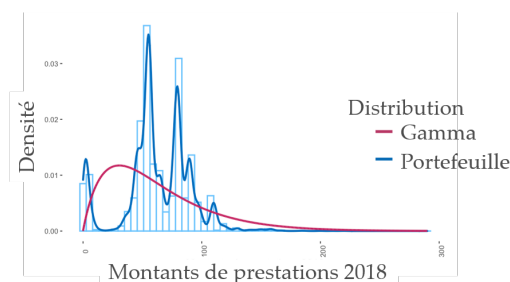


FIGURE F1 – Comparaison des distributions

Suite à la modélisation GLM, comme cela a été évoqué dans la partie théorique, une première segmentation « naïve » du portefeuille peut se dessiner en observant les coefficients du modèle pour les différentes modalités des variables. La Figure F2 représente ces coefficients pour le modèle GLM.

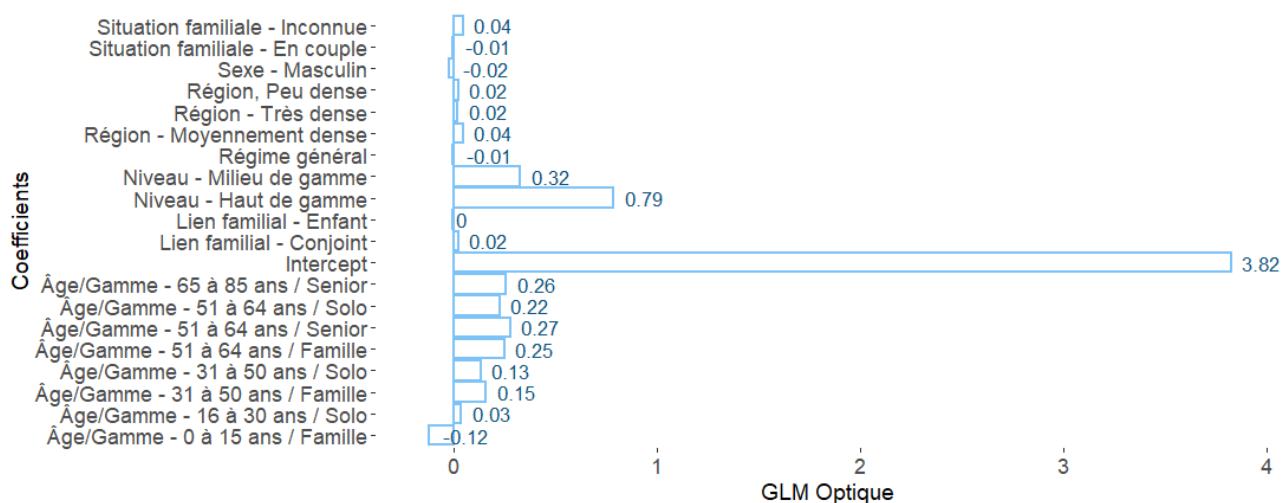


FIGURE F2 – Coefficients pour le modèle GLM

La valeur du coefficient pour l'Intercept est la plus élevée. Les valeurs des autres coefficients décrivent si le montant de prestation sera supérieur ou inférieur si la modalité de la variable prise dans le calcul de l'Intercept est remplacée par une autre. Ainsi par exemple les prestations sont plus faibles (car le coefficient est négatif) pour les assurés ayant entre 0 et 15 ans dans la gamme Famille que les assurés ayant entre 16 et 30 ans dans la gamme Famille (qui sont dans l'Intercept). Il est alors possible d'identifier que les prestations entrée de gamme (présentes dans l'Intercept) sont inférieures aux prestations milieu et haut de gamme, ce qui est cohérent ; ou encore que l'âge joue un rôle important dans la détermination du montant de prestations.

Ce graphique permet donc d'avoir une première approche de segmentation selon le montant des remboursements prédits.

D'autre part, il est nécessaire d'observer la table des déviations en faisant un test ANOVA. Ce test renvoie les p-valeurs de chaque variable et permet donc de savoir quelles variables sont significatives (*p-valeur inférieure à 0.05*) et lesquelles ne le sont pas (*p-valeur supérieure à 0.05*). Ainsi d'après ce test, seul la variable Régime Social n'est pas significative pour un seuil de 5%. Cela signifie que la modélisation par un GLM n'utilise pas cette variable pour prédire les montants de prestations. La courbe de Lorenz du modèle GLM étudié est relativement éloignée de celle des observations du portefeuille, il est possible que les montants de remboursement ne soient pas parfaitement bien répartis par le modèle. Cependant, les deux courbes sont plus proches entre le quantile à 0% et le quantile à 20% des sinistres (les plus coûteux donc), cela signifie que pour cette tranche l'ordre des prédictions respecte mieux celui des observations.

Par ailleurs, le modèle GLM présente un coefficient  $R^2$  de 23,30% et un indice de Gini valant 0,556.

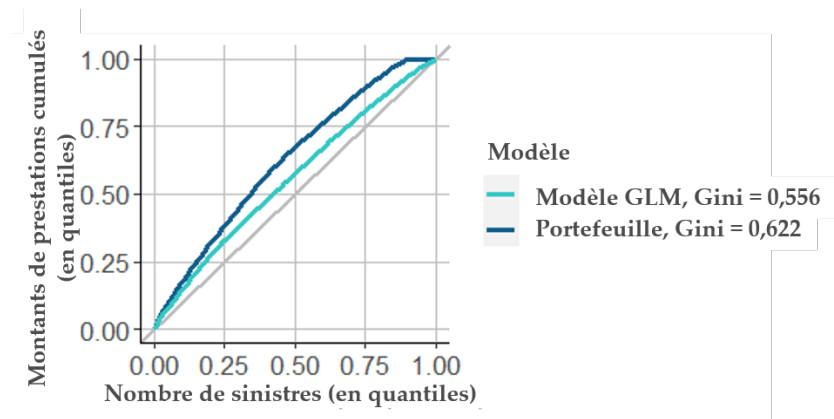


FIGURE F3 – Courbe de Lorenz associée au modèle GLM

Par ailleurs, l'indice de Gini ne permet pas de savoir si les prédictions du modèle sont proches des valeurs observées mais seulement de savoir si l'ordre entre les montants des sinistres prédits se rapproche de celui des montants observés. Ainsi, pour s'assurer que les prédictions sont cohérentes en ordre de grandeur, les moyennes doivent être comparées, et ce pour les prothèses dentaires d'une part puis pour les autres actes d'autre part.

	Moyenne
Base de test	62,35€
Valeurs prédites	62,40€

FIGURE F4 – Moyennes de la base de test et des prédictions obtenues avec le modèle GLM

Les moyennes des valeurs prédites sont très proches des valeurs observées, le modèle conserve donc bien l'ordre de grandeur des montants de prestations.

Enfin, comme cela a été décrit lors de la mise en place de la méthodologie de l'étude, le modèle GLM qui a été étudié ici est celui dont la valeur de  $R^2$  est proche de la moyenne des valeurs des 1 000 itérations. Par la suite, il est intéressant de connaître la distribution des valeurs de  $R^2$  et du coefficient de Gini pour ces 1000 itérations. Les box-plots des Figures F5 et F6 présentent ces distributions. Cela permettra par la suite de comparer les performances des différents modèles.

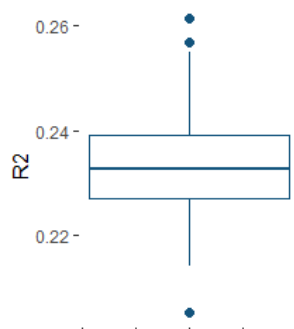


FIGURE F5 –  $R^2$  obtenus pour 1000 modèles

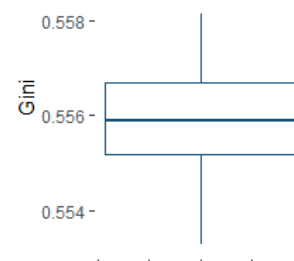


FIGURE F6 – Indices de Gini obtenus pour 1000 modèles

## Arbre CART

La capacité de segmentation des modèles GLM étant très limitée, la méthode des arbres CART va être utilisée afin de créer des profils d'assurés ayant la même consommation de soins, en terme de remboursement complémentaire. Tout comme pour le modèle GLM, l'arbre CART étudié est celui pour lequel la valeur de  $R^2$  est proche de la moyenne des valeurs des 1 000 itérations.

L'arbre CART obtenu est le suivant :

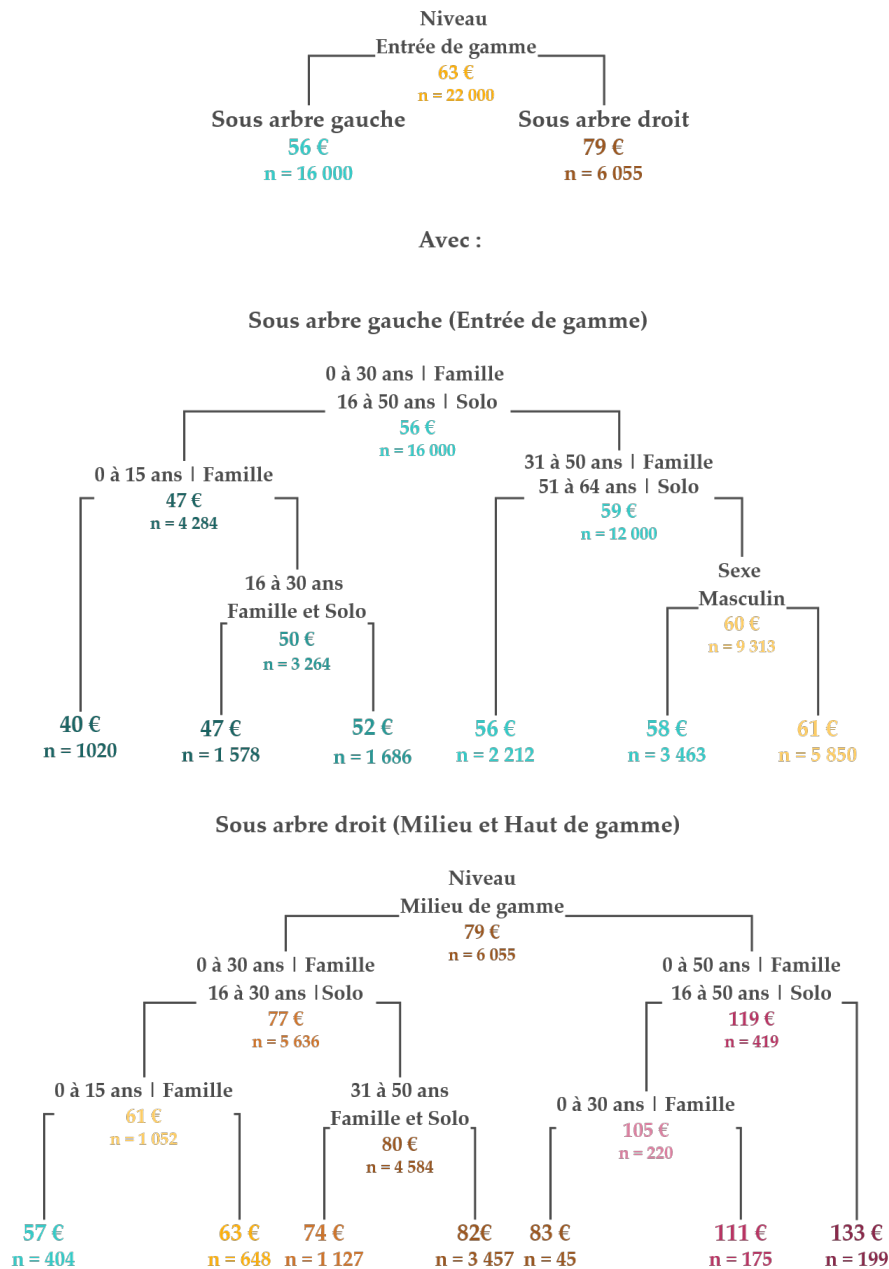


FIGURE F7 – Arbre CART obtenu

Cet arbre CART est plus profond que les deux arbres présentés pour les postes dentaire et audiologie, cela s'explique par le plus grand nombre de lignes (par rapport à l'audiologie) et l'absence du sous poste (par rapport au dentaire). Malgré sa profondeur, le modèle CART ne segmente le portefeuille quasiment qu'en utilisant le niveau de garantie ainsi que la variable liant l'âge et la gamme de garantie. Hormis ces deux variables, seul le genre de

l'assuré apparaît une fois dans l'arbre. Il est alors possible de lire dans l'arbre présenté à la Figure 2.51 qu'en moyenne les remboursements augmentent avec l'âge de l'assuré, ce qui conforte l'hypothèse prise dans le Chapitre 3 de la Partie II. Cet arbre montre également que plus le niveau de garantie est élevé, plus le remboursement complémentaire l'est. Cela permet d'illustrer le renoncement aux soins. En effet, bien qu'en moyenne la correction des verres optiques soit de plus en plus complexe avec l'âge, les différences qui apparaissent pour une même tranche d'âge selon le niveau de garantie sont importantes, alors qu'il est raisonnable de considérer que les besoins des assurés sont les mêmes.

Par ailleurs et pour rappel, toutes les variables utilisées lors de la segmentation des individus n'ont pas la même importance. Cette importance permet de déterminer quelles sont les variables qui jouent un rôle principal dans les profils obtenus et lesquelles sont plus accessoires. Les importances des variables pour l'arbre CART présenté à la Figure F7 sont les suivantes :

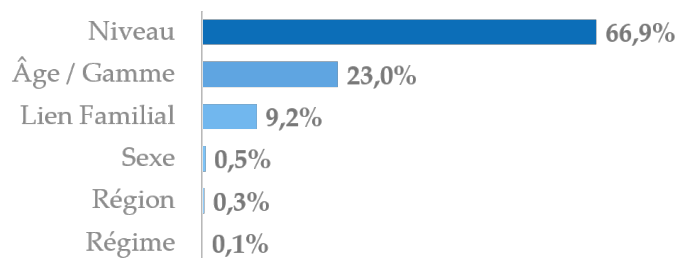


FIGURE F8 – Importance des variables dans la modélisation CART

Le constat reste le même que pour les postes audiologie et dentaire, l'importance des variables est très hétérogène, mettant en avant seulement les deux variables évoquées précédemment : le niveau de garantie, ainsi que l'âge lié à la gamme. Il apparaît néanmoins que le lien familial, pourtant absent de l'arbre de la Figure F7, a une importance non négligeable. Cela s'explique par le fait que cette variable reste corrélée avec l'âge et la gamme de garantie, et donc bien qu'elle n'apparaisse pas elle offrirait une segmentation satisfaisante. Finalement, cette Figure F8 met en avant l'importance des deux variables utilisées dans la segmentation obtenue par le modèle CART. Les profils seront donc pertinents du point de vue des variables qui les composent.

D'autre part, l'arbre CART ne prend pas en compte un grand nombre de variables. Cela peut provenir de la qualité des données, ainsi que du nombre d'observations et de variables relativement faible pour des modèles dits de « Machine Learning ». Aussi, avec des données de meilleure qualité les profils obtenus auraient pu être plus précis, mais il est préférable d'obtenir des profils moins précis et plus fiables que des profils liés à un sur-apprentissage.

De plus, en complément de cette segmentation il est possible de proposer à l'assureur une analyse plus « grossière », issue de l'étude de ce modèle. Contrairement à l'étude menée sur le poste dentaire, il semble que pour l'optique le montant des remboursements soit bien lié au renoncement aux soins. Ainsi, la segmentation proposée à la Figure E2 permet d'identifier différents profils d'assurés :

### Très enclin à modifier sa consommation

Niveau de garantie : **Entrée de gamme**  
Âge : **particulièrement les moins de 65 ans**  
RC moyen pour une prothèse auditive : **56 €**

### Plutôt enclin à modifier sa consommation

Niveau de garantie : **Milieu de gamme, particulièrement la gamme famille**  
Âge : **moins de 50 ans**  
RC moyen pour une prothèse auditive : **77 €**

### Peu enclin à modifier sa consommation

Niveau de garantie : **Haut de gamme**  
RC moyen pour une prothèse auditive : **119 €**

Finalement, la segmentation obtenue par modélisation CART permet de confirmer les considérations faites dans le Chapitre 3 de la Partie II : le niveau de garantie, ainsi que l'âge de l'assuré sont les principaux facteurs d'un potentiel changement de consommation face à cette réforme « 100% santé ». De plus, elle permet également à l'assureur d'avoir un meilleur aperçu de son portefeuille pour les postes réformés et ainsi de mieux piloter ses offres dans ce domaine. Encore une fois, un nombre de variables plus important initialement aurait peut-être permis à la segmentation d'apporter plus d'informations sur d'éventuels changements de comportement.

Par la méthode utilisée pour créer le modèle CART, la valeur du  $R^2$  est maximale pour la base d'apprentissage choisie, elle vaut ici 23,44%, ce qui est supérieur à la valeur obtenue pour le modèle GLM. Une comparaison plus complète sera faite à la fin de cette partie, une fois que les résultats seront obtenus pour tous les modèles, mais cette constatation permet d'ores et déjà de supposer que la segmentation obtenue avec le modèle CART est satisfaisante en terme de capacité prédictive. La courbe de Lorenz ainsi que l'indice de Gini associé sont présentés à la Figure F9, il apparaît que la courbe de Lorenz du modèle CART est très proche de celle observée pour le modèle GLM, les mêmes conclusions peuvent donc en être tirées. D'autre part, les moyennes des prédictions du modèle CART sont comparées à celles des observations à la Figure F10. Les valeurs prédites sont proches des valeurs observées, le modèle conserve donc bien l'ordre de grandeur des montants de prestations.

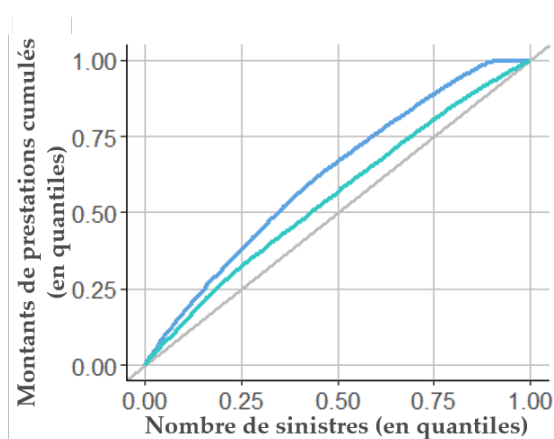


FIGURE F9 – Courbe de Lorenz associée au modèle CART

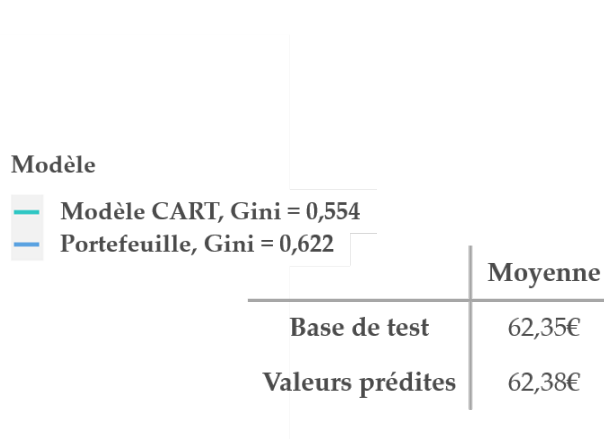


FIGURE F10 – Moyennes de la base de test et des prédictions obtenues avec le modèle CART

Enfin, les box-plots des Figures F11 et F12 présentent les distributions des deux mesures de validation obtenues en implémentant le modèle CART. Cela permettra par la suite de comparer les performances des différents modèles.

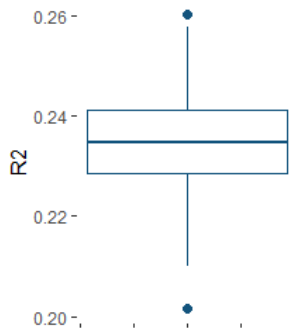


FIGURE F11 –  $R^2$  obtenus pour 1000 modèles

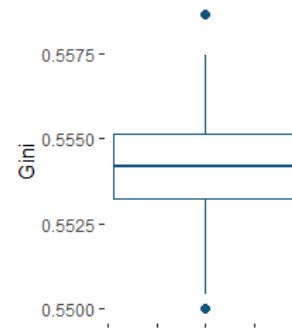


FIGURE F12 – Indices de Gini obtenus pour 1000 modèles

## Forêt aléatoire

Comme cela a été évoqué dans la partie théorique, afin de pouvoir juger de la qualité des prédictions du modèle CART qui mènent à la création de la segmentation, il est utile de comparer ces prédictions avec celles d'autres modèles tels que les forêts aléatoires, le XG Boost ou encore le modèle GLM présenté précédemment.

Une forêt aléatoire est implémentée sur la base des prestations d'optique. Pour ce faire, il est nécessaire de choisir le nombre d'arbres dans la forêt. Plus le nombre d'arbres est important, plus la prédiction est fiable mais plus le risque de sur-apprentissage augmente. Le choix se pose donc sur 400 arbres car ce nombre est un bon compromis qui permet un apprentissage correct mais évite autant que possible le sur-apprentissage. D'autre part, il est également possible de choisir le nombre de variables qui seront aléatoirement sélectionnées pour construire chacun des arbres de la forêt, le choix se porte ici sur 4 variables.

La forêt aléatoire modélisée de la sorte pour le poste optique permet d'obtenir un  $R^2$  de 23,59%, ce qui est supérieur à la valeur obtenue avec le modèle CART, et un indice de Gini de 0,555. Les moyennes sont présentées à la Figure F14 et restent toujours aussi proches.

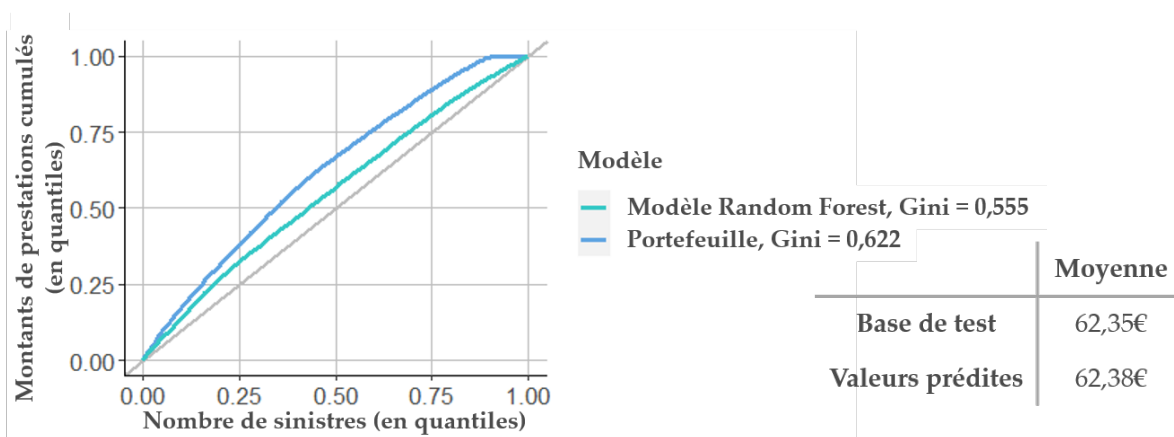


FIGURE F13 – Courbe de Lorenz associée au modèle Random Forest

FIGURE F14 – Moyennes de la base de test et des prédictions obtenues avec le modèle Random Forest

---

Enfin, les Figures F15 et F16 présentent les box-plots des distributions du  $R^2$  et de l'indice de Gini qui seront utilisés afin de comparer les performances des différents modèles.

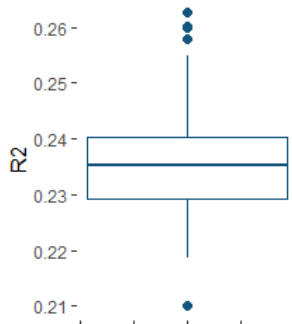


FIGURE F15 –  $R^2$  obtenus pour 1000 modèles

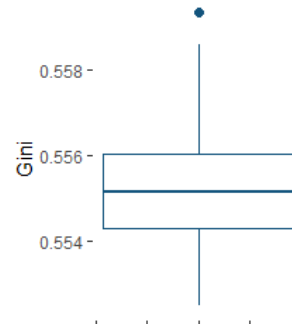


FIGURE F16 – Indices de Gini obtenus pour 1000 modèles

## Extrem Gradient Boosting (XG Boost)

La méthode XG Boost peut également permettre de vérifier la qualité de prédiction de l'arbre CART. Comme cela a été évoqué dans la partie théorique de ce chapitre, le Gradient Boosting est, comme les forêts aléatoires, basé sur l'agrégation d'arbres. Cependant, la différence entre les deux méthodes est que la construction des arbres par Gradient Boosting prend en compte les erreurs faites par les arbres précédemment construits. Ainsi, cette méthode donne en général de meilleurs résultats que les forêts aléatoires.

L'étude théorique a mis en évidence quatre quantités qu'il est possible de contrôler afin de paramétrer le modèle. La fonction *train* du package *xgboost* permet de considérer plusieurs valeurs pour chaque paramètre en intégrant une phase de cross validation dans la création du modèle : en sortie du modèle se trouve le modèle implémenté avec les paramètres qui optimisent le *MSE* et donc le  $R^2$ . Évidemment, plus le panel de valeurs pour chaque paramètres est large, plus le temps mis par l'algorithme pour trouver le modèle optimal est important. Il faut donc trouver un équilibre entre un nombre suffisant de valeurs pour laisser une marge à l'algorithme, et un nombre restreint de valeurs pour que le temps d'implémentation reste raisonnable. Dans cette étude du poste optique, les paramètres suivants sont choisis :

- le nombre  $k$  d'échantillons pour la validation croisée est **5** ;
- le nombre d'itérations du boosting est compris entre **80** et **110**, par pas de 1 ;
- la profondeur maximale des arbres est comprise entre **3** et **5**, par pas de 1 ;
- le nombre de variables choisies aléatoirement est compris entre **50%** et **90%** du nombre total de variables dans la base de données, par pas de 1%.

Le modèle finalement obtenu présente un  $R^2$  de 23,76%, et un indice de Gini de 0,557. Les moyennes sont présentées à la Figure F18. Il apparaît donc que ce dernier modèle permet d'obtenir de meilleurs résultats en terme de  $R^2$  : la variance des données est mieux représentée, mais aussi en terme d'indice de Gini : l'hétérogénéité des observations est également mieux représentée. Cela vient conforter l'idée qu'en général, bien que beaucoup plus opaque en terme de compréhension, le modèle XG Boost est également plus performant.



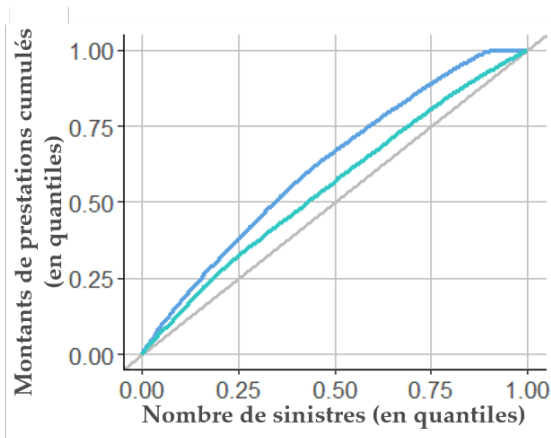


FIGURE F17 – Courbe de Lorenz associée au modèle XG Boost

Modèle

- Modèle XG Boost, Gini = 0,557
- Portefeuille, Gini = 0,622

	Moyenne
Base de test	62,35€
Valeurs prédites	62,34€

FIGURE F18 – Moyennes de la base de test et des prédictions obtenues avec le modèle XG Boost

Enfin, les Figures F19 et F20 présentent les box-plots des métriques de validation.

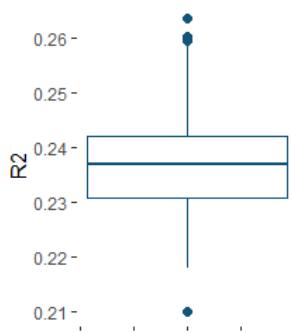


FIGURE F19 –  $R^2$  obtenus pour 1000 modèles

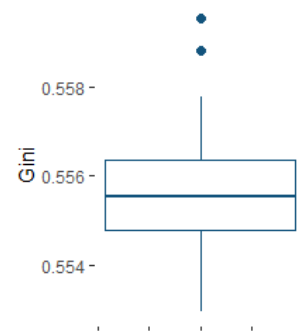


FIGURE F20 – Indices de Gini obtenus pour 1000 modèles

## Comparaison des résultats et conclusion

Afin de pouvoir conclure quant à la validité des profils obtenus par la modélisation CART, la comparaison des différents résultats est nécessaire.

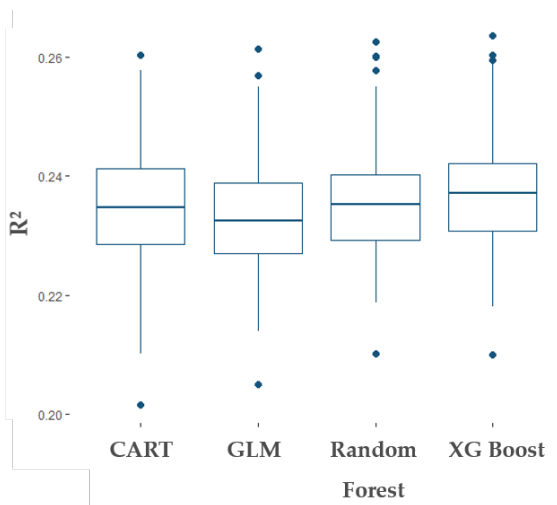


FIGURE F21 – Comparaison des  $R^2$  obtenus

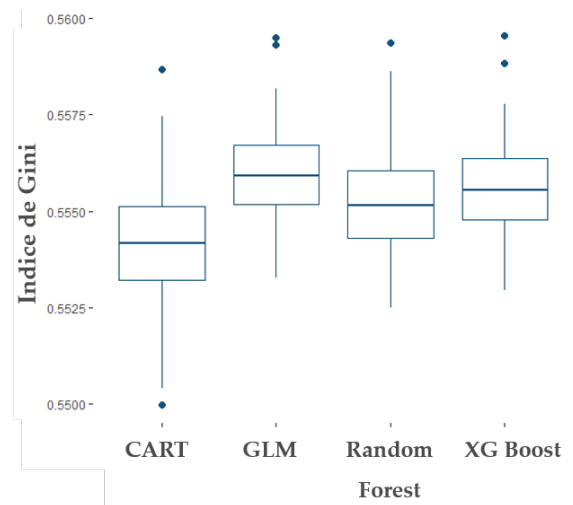


FIGURE F22 – Comparaison des coefficients de Gini obtenus

---

La Figure 2.65 montre que le modèle GLM est le moins performant pour prédire correctement les montants de prestations, ce qui n'a rien d'étonnant compte tenu que les données du portefeuille ne suivent pas réellement une loi Gamma. Néanmoins, le modèle GLM est également celui qui présente les meilleurs résultats pour l'indice de Gini d'après la Figure 2.66. Cela peut s'expliquer par le fait que bien que la loi Gamma ne suive pas correctement les données, elle permet d'ordonner correctement les prédictions. Cette différence de résultat entre les deux mesures pour le modèle GLM met justement en avant la raison pour laquelle leur étude en parallèle est nécessaire : le  $R^2$  permet de savoir si la prédiction est proche de la valeur initiale et si le modèle rend bien compte de la variance des données, alors que l'indice de Gini informe sur la capacité à prédire les variables « dans le bon ordre ».

En dehors du modèle GLM, les Figures 2.65 et 2.66 montrent que la performance des modèles augmente avec leur complexité. Le modèle CART apparaît comme le moins performant au regard de l'indice de Gini, et seulement plus performant que le modèle GLM au regard de  $R^2$ . Le modèle XG Boost est en moyenne meilleur de plus de 1,3% pour le coefficient  $R^2$ , ce qui est raisonnable. La segmentation obtenue par la modélisation CART reste relativement simple : elle ne fait intervenir que deux variables et les profils obtenus sont cohérents avec l'expérience du marché de la santé. Elle peut donc tout de même être considérée comme satisfaisante.