

Mémoire présenté devant l'Université de Paris-Dauphine
pour l'obtention du Certificat d'Actuaire de Paris-Dauphine
et l'admission à l'Institut des Actuaire
le 31 janvier 2023

Par : Diletta KOUAM

Titre : Téléconsultation : évolutions de la pratique et impacts actuels et futurs sur la sinistralité en Santé

Confidentialité : Non Oui (Durée : 1 an 2 ans)

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

*Membres présents du jury de l'Institut
des Actuaire :*

*Membres présents du Jury du Certificat
d'Actuaire de Paris-Dauphine :*

Entreprise :

Nom : SIA Partners

Signature :

Sia Partners
21 rue de Berri - 75008 PARIS
RCS Paris 423 507 730
SASU Capital 200 000€

Directeur de Mémoire en entreprise :

Nom : Baptiste ANDRIEU

Signature :



*Autorisation de publication et de mise en ligne sur un site de diffusion de documents
actuaires (après expiration de l'éventuel délai de confidentialité)*

Secrétariat :

Bibliothèque :

Signature du responsable entreprise



Signature du candidat



Résumé

La téléconsultation est un poste de soins prenant de l'ampleur ces dernières années, de par, non seulement la crise sanitaire engendrée par la Covid-19, mais aussi des raisons médicales, économiques et sociales.

Cette innovation technique et les évolutions de consommation de soins associées représentent toutefois pour l'assureur une incertitude quant aux engagements qu'il porte.

Ainsi, dans la perspective d'améliorer la connaissance de ce risque, ce mémoire propose d'en étudier la sinistralité actuelle afin de prédire l'étendue et les conséquences futures pour le marché de l'assurance Santé en France.

Une première partie présentera brièvement le fonctionnement du système de santé français puis s'attardera sur le contexte de l'essor de la pratique de téléconsultation en France.

La suite du mémoire sera consacrée à l'analyse de la sinistralité de téléconsultation, en comparaison à celle de consultation, selon diverses approches.

1. L'analyse de la sinistralité historique sur la base des techniques usuelles de tarification, notamment la modélisation GLM et l'approche coût-fréquence. Cela reposera sur les données Open Source de la Sécurité Sociale et des statistiques de population de l'INSEE.
2. L'étude de la sinistralité future à l'aide, dans un premier temps, de séries temporelles modélisant les fréquences de recours à la consultation et à la téléconsultation. Dans un deuxième temps, l'élaboration et la calibration de scénarios d'évolution de la consommation permettront de challenger le caractère prospectif des modèles antérieurement étudiés.

L'objectif final sera de tirer des conclusions quant à la place, en termes de sinistralité, qu'occupera la téléconsultation dans le paysage assurantiel en santé dans le futur.

Mots-clés : Téléconsultation, Consultation, Sinistralité, Tarification, Projection, Séries temporelles.

Abstract

Teleconsultation is a healthcare industry that has been growing in recent years, not only because of the health crisis caused by COVID-19 but also for medical, economic and social reasons.

However, this technical innovation and the associated changes in healthcare consumption represent an uncertainty for the insurer in terms of its commitments.

Thus, with a view to improving knowledge of this risk, this paper proposes to study its current claim experience in order to predict the extent and future consequences for the French health insurance market.

The first part will briefly present the functioning of the French healthcare system and then focus on the context of the development of teleconsultation in France.

The rest of the paper will be devoted to the analysis of teleconsultation claim experience, in comparison to consultation one, according to various approaches.

1. The analysis of historical claims experience based on standard pricing techniques, including GLM modeling and the cost-frequency approach. This will be based on Sécurité sociale open source data and INSEE population statistics.
2. The study of future claims experience using, in the first instance, time series modeling of consultation and teleconsultation frequencies. Secondly, the development and calibration of consumption evolution scenarios will allow the challenge of the prospective character of the models previously studied.

The final objective will be to draw conclusions as to the place, in terms of claims experience, that teleconsultation will occupy in the future health insurance landscape.

Keywords : Teleconsultation; Consultation ; Claims experience ; Pricing ; Projection ; Time series.

Note de Synthèse

Les innovations technologiques, combinées aux nouveaux modes de vie, modifient en profondeur le rapport à la santé et ouvrent de nouvelles opportunités. Parmi ces dernières, la réalisation d'une consultation médicale à distance par écrans interposés, pratiquée par un médecin généraliste ou bien spécialiste, appelée plus couramment la téléconsultation, constitue un vecteur d'évolution majeur dans le cadre de la consommation de soins en santé.

Après des débuts timides, la crise de la Covid-19 a agi en tant qu'accélérateur d'adoption de la téléconsultation, qui, en l'espace de quelques semaines, a connu un véritable essor en France en 2020. Le graphique suivant, représentant la part mensuelle des dépenses en téléconsultation sur l'ensemble des consultations médicales réalisées pendant la première année de l'épidémie, illustre clairement ce propos.

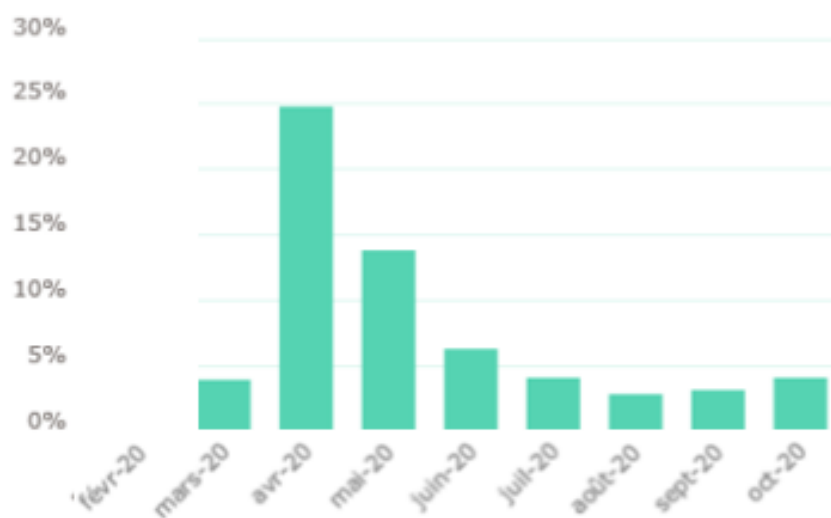


FIGURE 1 : Part (%) mensuelle des dépenses en téléconsultation sur l'ensemble des consultations médicales réalisées entre novembre 2019 et octobre 2020 (en date de paiement)

Cette pratique médicale innovante constitue un enjeu essentiel dans l'amélioration du système de santé actuel car pour autant que l'on dispose d'un smartphone ou d'une tablette, elle offre une réponse immédiate à des problématiques telles que la difficulté d'accès aux soins, dans les déserts médicaux par exemple, ou encore l'amélioration de la prise en charge et du suivi des patients.

Toutefois ces évolutions de consommation représentent une incertitude pour les assureurs et soulèvent donc le besoin d'en étudier l'essor et l'amplitude actuelle, dans un contexte de maîtrise des risques futurs.

C'est dans ce contexte que le mémoire propose d'étudier l'évolution de la consommation de ces actes de téléconsultation et de consultation traditionnelle.

Construction et analyses des bases de modélisation

Afin d'appliquer les techniques usuelles de tarification, il est nécessaire d'avoir des informations sur les sinistres, ainsi que des données sur l'exposition aux sinistres. Les premières seront regroupées dans une base de données appelée base « sinistres », les dernières se retrouveront dans une base nommée base « assurés ». Ainsi la première étape a été de construire ces deux bases de données.

En actuariat, il est courant d'exploiter des jeux de données de taille relativement importante. Partant de la base de données Open DAMIR, pour la création de la base « sinistres », mise à disposition par la Sécurité sociale, il a été inévitable d'être confronté à des problématiques de volumétrie et de complexité. A cet effet, le poids total des données brutes à traiter s'est élevé à 142,94 Go, ce qui correspond à environ 72 500 000 observations ($\sim 70\,000\,000$ lignes sur la consultation et $\sim 2\,500\,000$ lignes sur la téléconsultation).

Bien que le caractère agrégé de l'Open DAMIR permette l'anonymisation des données, il soulève un point d'attention quant à la détermination de la fréquence et de sa loi en présence d'informations uniformisées. Cette donnée est d'autant plus importante que l'étude de l'évolution temporelle de la consultation et de la téléconsultation est réalisée à partir de séries temporelles modélisant les fréquences de survenance de ces pratiques. C'est alors à ce moment qu'interviennent les données démographiques de l'INSEE. La base « assurés » s'appuie principalement sur les estimations annuelles de répartition de la population française par âge quinquennal, sexe et région, élaborées par l'INSEE.

Ainsi divers traitements ont été réalisés pour aboutir à ces bases de données. Ils sont résumés sur le schéma ci-après.

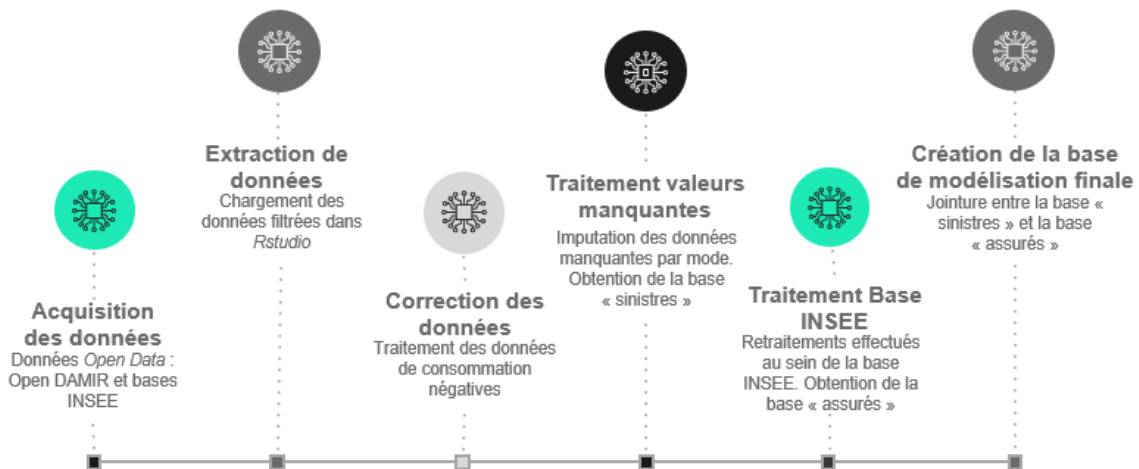


FIGURE 2 : Processus de transformation des données

La deuxième étape a été d'analyser les données traitées avant toute modélisation. Cette étape permet de comprendre les données en mesurant empiriquement l'impact de chaque variable tarifaire sur la consommation des téléconsultations comparée à celle des consultations. Il existe six facteurs de risque : année de remboursement, le sexe, l'âge, l'affiliation à la CMU, la région de résidence et la nature de l'acte de consultation considéré.

Tarification de prestations de consultation et de téléconsultation

Suite aux modélisations GLM du coût moyen et de la fréquence de téléconsultation et de consultation, calibrées suivant le principe de tarification « coût-fréquence », il a été possible de déterminer les primes pures des prestations de téléconsultation. Celles-ci sont présentées par année de remboursement et par nature de la prestation considérée.

Année	Poste de soins	Type de prestation	Fréquence moyenne	Coût moyen (en euros)	Sinistralité moyenne par prestation (en euros)	Sinistralité moyenne globale (en euros)
2018	Consultation traditionnelle	Consultation toutes spécialités	0,08	36,36	2,77	17,47
		Consultation généraliste	0,57	26,01	14,70	
	Téléconsultation	Téléconsultation toutes spécialités	0,53e-4	37,29	0,19e-2	0,25e-2
		Téléconsultation généraliste	0,2e-4	26,81	0,54e-3	
2019	Consultation traditionnelle	Consultation toutes spécialités	0,24	36,36	8,61	48,99
		Consultation généraliste	1,55	26,08	40,37	
	Téléconsultation	Téléconsultation toutes spécialités	0,94e-3	36,41	0,03	0,08
		Téléconsultation généraliste	0,18e-2	26,18	0,05	
2020	Consultation traditionnelle	Consultation toutes spécialités	0,19	36,36	7,22	43,17
		Consultation généraliste	1,33	26,08	34,47	
		Consultation lieu dédié et urgences	0,02	92,07	1,41	
		Consultation à domicile	0,32e-2	23,52	0,07	
	Téléconsultation	Téléconsultation toutes spécialités	0,08	35,12	2,64	9,34
		Téléconsultation généraliste	0,26	25,25	6,67	
		Téléconsultation lieu dédié et urgences	0,99e-3	23,83	0,02	
		Téléconsultation à domicile	0,18e-3	16,53	0,29e-2	

FIGURE 3 : Primes pures par année de remboursement et nature de la prestation.

Il en résulte qu'au fil des années l'écart de sinistralité entre les pratiques médicale de consultation et de téléconsultation se comble, car la différence entre les fréquences moyennes de consultation et de téléconsultation s'est amoindrie.

A titre d'illustration, il est observé que la fréquence moyenne de recours à une téléconsultation spécialiste (toutes spécialités confondues) a été multipliée x10 entre 2018 et 2019 et x100 entre 2019 et 2020 (approximativement), en atteignant sensiblement les mêmes ordres de grandeur des fréquences moyennes des prestations de consultation spécialistes (0,08 pour les téléconsultations toutes spécialités contre 0,19 pour les consultations toutes spécialités).

Ce phénomène est principalement du à l'accélération du développement de la fréquence de recours aux prestations de téléconsultation après apparition de la Covid-19.

Modélisation de la sinistralité future à l'aide des séries temporelles

L'objectif de cette section est de montrer l'existence d'un phénomène de « vases communicants » entre les évolutions de consommation des deux pratiques médicales en France, actuelles et futures, à l'aide de séries temporelles multivariées. L'étude s'est focalisée sur les fréquences d'utilisation car, comme susmentionné, les écarts entre les coûts moyens de l'ensemble des prestations de consultation et de téléconsultation sont négligeables comparés aux écarts de fréquence.

L'utilisation des séries temporelles multivariées en particulier est pertinente dans le contexte de l'étude car elle permet de mettre en évidence les effets de corrélation et de causalité entre différentes variables. Dans le cadre du mémoire, le modèle vectoriel autorégressif (VAR) de paramètre $p = 7$, a été retenu, à la suite d'une étude exhaustive des deux séries temporelles univariées.

Les étapes classiques du traitement statistique d'une série chronologique sont les suivantes :

- Analyse qualitative de la série et de ses caractéristiques fondamentales ;
- Choix et ajustement des modèles ;
- Projections du modèle.

La figure suivante représente les projections réalisées en appliquant le modèle $VAR(7)$ aux séries de fréquence de consultation et de téléconsultation.

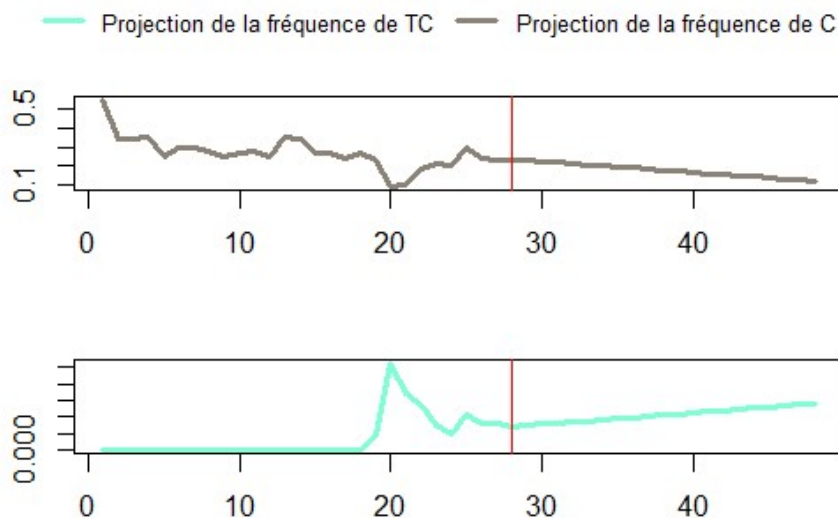


FIGURE 4 : Projections dans le futur des fréquences de survenance de la consultation et de la téléconsultation

Ces figures mettent en évidence la relation inverse de proportionnalité existant entre les évolutions futures des deux pratiques médicales. A partir des données historiques, le modèle prévoit une augmentation continue de la fréquence de consommation de téléconsultation, tandis que celle de consultation est vouée à diminuer dans le futur (selon le modèle). De plus, le modèle mesure une corrélation très négative (-0.933) entre les résidus des prédictions de la série de fréquence de téléconsultation et ceux de la série de fréquence de consultation.

Toutefois, il faut tenir en compte que ces résultats sont en grande partie dus à l'apparition de la Covid-19 en 2020, permettant à la téléconsultation de connaître un incroyable succès en France, de part le contexte exceptionnel de l'épidémie. Or, ce type d'évènements n'advient que très rarement,

ainsi il serait intéressant de poursuivre l'étude de l'évolution de la pratique médicale, une fois les effets de la Covid-19 atténués.

Par ailleurs, la structure du modèle VAR fournit des informations sur la capacité de prévision d'une variable ou d'un groupe de variables pour d'autres variables.

A la suite de l'application aux données du test de causalité au sens de Granger, il résulte que la fréquence de consultation influence, au sens de Granger, la fréquence de téléconsultation. Néanmoins le contraire n'est pas vérifié.

Calibration de scénarios d'évolution de la sinistralité

Afin d'approfondir l'étude de l'évolution future de la sinistralité en téléconsultation, divers scénarios d'évolution de la consommation de consultations et de téléconsultations entre 2021 et 2070 ont été analysés. Ces scénarios ont été calibrés à partir des modèles de tarification et de séries temporelles précédemment étudiés afin de les challenger. L'objectif final est de tirer des éventuelles conjectures quant à la place qu'occupera la téléconsultation dans le paysage assurantiel en santé à l'horizon 2070.

La construction de scénarios repose sur les variables pivots préalablement identifiées. Les variables pivots sont les facteurs susceptibles d'affecter significativement le phénomène étudié. Dans le cadre du mémoire, des hypothèses de projection sur les paramètres suivants seront prises en compte :

- les expositions futures aux sinistres,
- le coût moyen des sinistres futurs,
- la fréquence de survenance des sinistres futurs.

La constitution des scénarios repose alors sur le croisement de ces hypothèses appliquées aux modèles de tarification et de séries temporelles précédemment utilisés. Ces scénarios sont présentés par les figures suivantes.

Stress Téléconsultation (TC)			
	Source de fréquence	Source de coût moyen	Hypothèses de projection
Scénario central	Résultats des modèles GLM	Résultats des modèles GLM	Prise en compte de l'inflation
Scénario séries temporelles	Prédictions du modèle VAR de séries temporelles	Résultats des modèles GLM	Prise en compte de l'inflation
Scénario pessimiste	Résultats des modèles GLM	Résultats des modèles GLM	<ul style="list-style-type: none"> • Prise en compte de l'inflation • Décroissance linéaire de la fréquence jusqu'à atteindre en 2070 la valeur enregistrée en 2019
Stress Consultation (C)			
	Source de fréquence	Source de coût moyen	Hypothèse de projection
Scénario central	Résultats des modèles GLM	Résultats des modèles GLM	Prise en compte de l'inflation
Scénario séries temporelles	Prédictions du modèle VAR de séries temporelles	Résultats des modèles GLM	<ul style="list-style-type: none"> • Prise en compte de l'inflation • Etablissement d'une valeur critique pour la fréquence de consultation par avis d'experts
Scénario tenant compte de la corrélation	Résultats des modèles GLM	Résultats des modèles GLM	<ul style="list-style-type: none"> • Prise en compte de l'inflation • Etude de l'influence de la corrélation empirique existant entre les fréquences de C et de TC

FIGURE 5 : Tableau récapitulatif des scénarios d'évolution étudiés lors des travaux

Le croisement de ces scénarios d'évolution permet ainsi de déterminer les primes pures annuelles de consultation et de téléconsultation entre 2021 et 2070. Afin d'exploiter ces résultats, il a été considéré que les agréger ou les multiplier ne serait pas cohérent. Finalement, il a été préféré d'analyser les divers couples de primes pures ayant comme expression générale $[PP_{2070}(C); PP_{2070}(TC)]$ et établis par croisement des scénarios de stress de la consultation et de la téléconsultation.

L'objectif principal de cette représentation est de prédire l'influence qu'aura la téléconsultation dans le paysage de l'assurance Santé à l'horizon 2070. A cet effet, la figure suivante présente la part de sinistralité portée par téléconsultation dans le panier de soins composé de consultations et de téléconsultation.

En pourcentage (%)	Scénario central	Scénario séries temporelles	Scénario pessimiste
Scénario central	15,5	37,2	3,4
Scénario séries temporelles	38,5	66,9	10,6
Scénario tenant compte de la corrélation	3,4	10,1	0,7

FIGURE 6 : Ratios de sinistralité portés par la téléconsultation en fonction des scénarios d'évolution

Celui-ci montre qu'il existe quatre niveaux caractérisant les projections de la proportion d'actes de téléconsultation parmi la consommation globale de consultations et de téléconsultations :

- Le premier niveau inclut les projections les plus défavorables à l'essor de la téléconsultation à l'horizon 2070. Elle comprend les ratios allant de 0,7 à 3,4 %, touchant trois croisements de scénarios d'évolution (sur neuf). L'hypothèse pénalisant le plus la croissance de la pratique de téléconsultation à l'horizon 2070 est alors celle de corrélation linéaire entre les fréquences de consultation et de téléconsultation.
- Le deuxième niveau comprend les ratios allant de 10,1 à 15,5 % . Les projections de trois croisements de scénarios d'évolution (sur neuf) font partie de ce niveau. Étant donné que la part du panier de soins prise par la téléconsultation valait 17,8 % en 2020, ces croisements de scénarios sont rapprochant le plus de la part de sinistralité actuelle de téléconsultation.
- Le troisième niveau est partagé par deux croisements et considère qu'en 2070, la sinistralité de téléconsultation occupera environ le tiers de la sinistralité globale de consultations et de téléconsultations (entre 37,2 et 38,5 %). Il s'agit du croisement du scénario central de consultation et du scénario intégrant les projections des modèles de séries temporelles pour la téléconsultation et vice versa.
- Enfin, la projection fournie par le croisement des scénarios tenant compte des prévisions des modèles de séries temporelles de la fréquence de consultation et de téléconsultation (ainsi que l'application de l'inflation) constitue le dernier niveau. Elle prévoit que la sinistralité en téléconsultation aura deux fois plus de poids et d'influence (66,9%) dans le panier d'assurance santé constitués de consultations et de téléconsultations que la pratique de consultation, à l'horizon 2070.

Conclusion

Ce mémoire a voulu étudier l'essor du risque téléconsultation en assurance Santé en proposant : tout d'abord la mise en place de modèles tarifaires prédictifs de la sinistralité en consultation et en téléconsultation, afin d'analyser leur relation dans le temps ; puis en projetant la sinistralité dans le futur afin de prédire l'influence qu'aura la téléconsultation dans le panier de soins constitué d'actes de consultation et de téléconsultation.

Ces études ont montré que la sinistralité globale de consultation et de téléconsultation a de grandes chances de doubler voir tripler dans le futur et ce, avec ou sans la prise en compte du taux d'inflation annuel. De plus, il est constaté que les écarts de sinistralité entre les deux pratiques médicales par rapport à la consommation globale de soins se creuseront probablement au cours du temps. Cependant, il est relevé que la place de la téléconsultation dans ce panier de soins, en termes de sinistralité, restera vraisemblablement inférieure à celle de consultation à l'horizon 2070.

Toutefois, compte tenu de la profondeur limitée de l'historique des données, il serait intéressant de poursuivre les travaux en intégrant dans l'étude les données de consommation de consultation et de téléconsultation datant de 2021, désormais disponibles sur le site de l'assurance maladie.

Synthesis note

Technological innovations, combined with new lifestyles, are profoundly changing the relationship with health and opening up new opportunities. Among the latter, remote medical consultations by a general practitioner or a specialist, more commonly known as teleconsultation, are a major vector of change in the consumption of health care.

After a timid start, the Covid-19 crisis has acted as an accelerator for the adoption of teleconsultation, which, in the space of a few weeks, has really taken off in France in 2020. The following graph, representing the monthly share of teleconsultation spending on all medical consultations performed during the first year of the epidemic, clearly illustrates this point.



Figure 7: Monthly share (%) of teleconsultation expenses of total medical visits performed between November 2019 and October 2020 (as of payment date)

This innovative medical practice is a key factor in the improvement of the current healthcare system because, as long as you have a smartphone or a tablet, it offers an immediate response to problems such as the difficulty of accessing care, in medical deserts for example, or the improvement of patient care and follow-up.

However, these changes in consumption represent an uncertainty for insurers and therefore raise the need to study their current growth and scope, in a context of controlling future risks.

It is in this context that the thesis proposes to study the evolution of the consumption of these teleconsultation and traditional consultation procedures.

Construction and analysis of the modeling bases

In order to apply the usual pricing techniques, it is necessary to have information on claims, as well as data on exposure to claims. The former will be gathered in a database called "claims" database, the latter will be gathered in a database called "insured" database. So the first step was to build these two databases.

In actuarial science, it is common to use relatively large data sets. Starting from the Open DAMIR database, for the creation of the "claims" database, made available by the Social Security, it was inevitable to be confronted with problems of volume and complexity. To this end, the total weight of the raw data to be processed amounted to 142.94 Go, which corresponds to approximately 72,500,000 observations (~70,000,000 lines on the consultation and ~ 2,500,000 lines on the teleconsultation).

Although the aggregated nature of Open DAMIR allows for the anonymization of the data, it raises a point of attention with regard to the determination of the frequency and its law in the presence of standardized information. This data is all the more important since the study of the temporal evolution of consultation and teleconsultation is carried out from time series modeling the frequencies of occurrence of these practices. This is where INSEE demographic data comes in. The "insured" database is mainly based on annual estimates of the distribution of the French population by five-year age, sex and region, compiled by INSEE.

Thus, various processing operations have been carried out to obtain these databases. They are summarized in the diagram below.

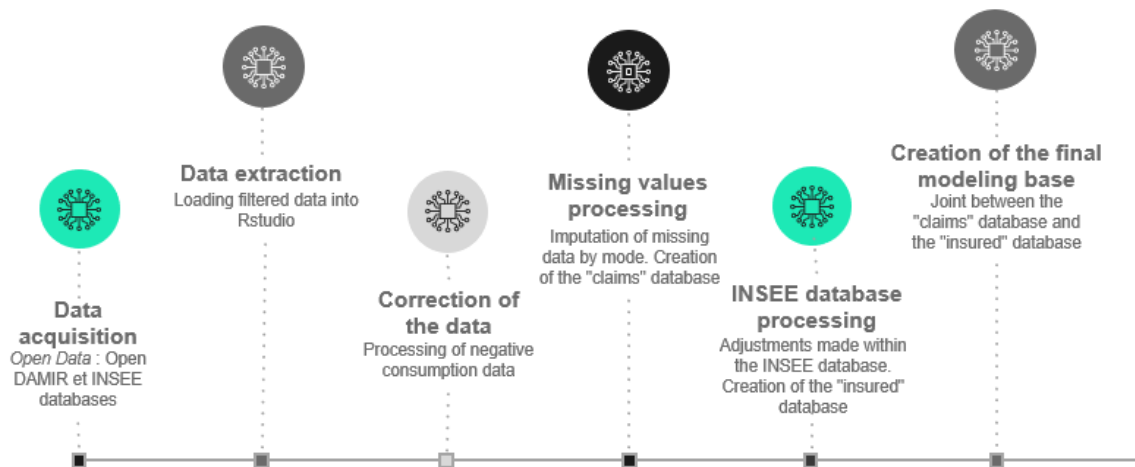


Figure 8: Data transformation process

The second step was to analyze the processed data before any modeling. This step allows us to understand the data by empirically measuring the impact of each tariff variable on the consumption of teleconsultations compared to consultations. There are six risk factors: year of reimbursement, gender, age, CMU affiliation, region of residence and the nature of the consultation procedure considered.

Pricing of consultation and teleconsultation services

Following the GLM modelling of the average cost and frequency of teleconsultation and consultation, calibrated according to the cost-frequency pricing principle, it was possible to determine the pure premiums for teleconsultation services. These are presented by year of reimbursement and by type of service considered.

Year	Medical service	Type of service	Average frequency	Average cost (in euros)	Average claims experience per benefit (in euros)	Overall average claims experience (in euros)
2018	Conventional consultation	All specialties consultation	0,08	36,36	2,77	17,47
		Generalist consultation	0,57	26,01	14,70	
	Teleconsultation	All specialties teleconsultation	0,53e-4	37,29	0,19e-2	0,25e-2
		Generalist teleconsultation	0,2e-4	26,81	0,54e-3	
2019	Conventional consultation	All specialties consultation	0,24	36,36	8,61	48,99
		Generalist consultation	1,55	26,08	40,37	
	Teleconsultation	All specialties teleconsultation	0,94e-3	36,41	0,03	0,08
		Generalist teleconsultation	0,18e-2	26,18	0,05	
2020	Conventional consultation	All specialties consultation	0,19	36,36	7,22	43,17
		Generalist consultation	1,33	26,08	34,47	
		Consultation dedicated place and emergencies	0,02	92,07	1,41	
		Home consultation	0,32e-2	23,52	0,07	
	Teleconsultation	All specialties teleconsultation	0,08	35,12	2,64	9,34
		Generalist teleconsultation	0,26	25,25	6,67	
		Teleconsultation dedicated place and emergencies	0,99e-3	23,83	0,02	
		Home teleconsultation	0,18e-3	16,53	0,29e-2	

Figure 9: Pure premiums by reimbursement year and type of service

As a result, over the years, the gap between the claims experience of consultation and teleconsultation medical practices has narrowed, as the difference between the average frequency of consultation and teleconsultation has decreased.

By way of illustration, it is observed that the average frequency of recourse to a specialist teleconsultation (all specialties combined) was multiplied x10 between 2018 and 2019 and x100 between 2019 and 2020 (approximately), reaching approximately the same orders of magnitude as the average frequencies of specialist consultation services (0.08 for all-specialty teleconsultations versus 0.19 for all-specialty consultations).

This phenomenon is mainly due to the acceleration of the development of the frequency of use of teleconsultation services after the appearance of Covid-19.

Modeling future claims experience using time series

The objective of this section is to show the existence of a phenomenon of communicating vessels between the evolution of consumption of the two medical practices in France, present and future, using multivariate time series. The study focused on the frequencies of use because, as mentioned above, the differences between the average costs of all consultation and teleconsultation services are negligible compared to the differences in frequency.

The use of multivariate time series in particular is relevant in the context of the study because it allows for the identification of correlation and causality effects between different variables. In the context of the thesis, the vector autoregressive model (VAR) of parameter $p = 7$, was retained, following an exhaustive study of the two univariate time series.

The classical steps of the static treatment of a time series are the following:

- Qualitative analysis of the series and its fundamental characteristics;
- Selection and fitting of the models;
- Model projections.

The following figure represents the projections made by applying the $VAR(7)$ model to the consultation and teleconsultation frequency series.

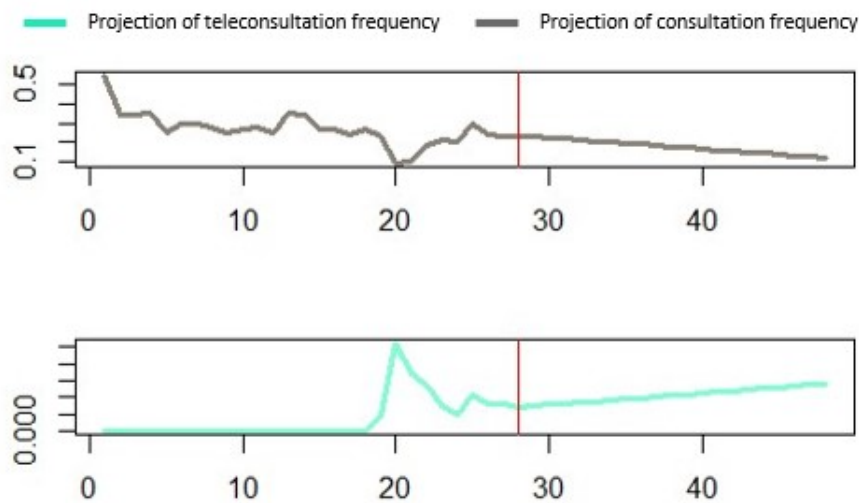


Figure 10: Future projections of the frequency of occurrence of consultations and teleconsultations

These figures highlight the inverse relationship of proportionality that exists between the future evolution of the two medical practices. Based on historical data, the model predicts a continuous increase in the frequency of consumption of teleconsultation, while that of consultation is bound to decrease in the future (according to the model). Furthermore, the model measures a very negative correlation (-0.933) between the residuals of the predictions of the teleconsultation frequency series and those of the consultation frequency series.

However, it should be taken into account that these results are largely due to the appearance of Covid-19 in 2020, allowing teleconsultation to be incredibly successful in France, due to the exceptional context of the epidemic. However, this type of event occurs very rarely, so it would be interesting to continue studying the evolution of teleconsultation practice once the effects of Covid-19 have subsided.

Furthermore, the structure of the VAR model provides information about the predictive ability of one variable or group of variables for other variables.

After applying the Granger causality test to the data, it follows that the frequency of consultation influences, in the Granger sense, the frequency of teleconsultation. However, the opposite is not true.

Calibration of scenarios of evolution of the claims experience

In order to further study the future evolution of teleconsultation claims, various scenarios for the evolution of consultation and teleconsultation consumption between 2021 and 2070 were analyzed. These scenarios were calibrated using the pricing and time series models previously studied in order to challenge them. The final objective is to draw possible conjectures as to the place teleconsultation will occupy in the health insurance environment by 2070.

The construction of scenarios is based on previously identified pivotal variables. The pivotal variables are the factors likely to significantly affect the phenomenon under study. In the context of the brief, projection assumptions on the following parameters will be taken into account:

- the future loss exposures,
- the average cost of future claims,
- the frequency of occurrence of future claims.

The scenarios are then constructed by combining these assumptions with the pricing and time series models used previously. These scenarios are presented in the following figures.

Stress test Teleconsultation (TC)			
	Frequency source	Average cost source	Projection hypotheses
Central scenario	Results of the GLM models	Results of the GLM models	Taking inflation into account
Time series scenario	Predictions of the time series VAR model	Results of the GLM models	Taking inflation into account
Pessimistic scenario	Results of the GLM models	Results of the GLM models	<ul style="list-style-type: none"> • Taking inflation into account • Linear decrease in frequency until the value recorded in 2019 is reached in 2070
Stress test Consultation (C)			
	Frequency source	Average cost source	Projection hypotheses
Central scenario	Results of the GLM models	Results of the GLM models	Taking inflation into account
Time series scenario	Predictions of the time series VAR model	Results of the GLM models	<ul style="list-style-type: none"> • Taking inflation into account • Establishment of a critical value for consultation frequency by expert opinion
Scenario taking into account correlation	Results of the GLM models	Results of the GLM models	<ul style="list-style-type: none"> • Taking inflation into account • Study of the influence of the empirical correlation between the frequencies of C and TC

Figure 11: Overview of the evolution scenarios studied during the work

The cross-tabulation of these evolution scenarios thus makes it possible to determine the annual pure premiums for consultation and teleconsultation between 2021 and 2070. In order to exploit these results, it was considered that aggregating or multiplying them would not be coherent. Finally, it was preferred to analyze the various pairs of pure premiums having as general expression $[PP_{2070}(C); PP_{2070}(TC)]$ and established by crossing the consultation and teleconsultation stress scenarios.

The main objective of this representation is to predict the influence that teleconsultation will have on the health insurance landscape by 2070. To this end, the following figure presents the share of claims carried by teleconsultation in the care basket composed of consultations and teleconsultation.

In percentage(%)	Central scenario	Time series scenario	Pessimistic scenario
Central scenario	15.5	37.2	3.4
Time series scenario	38.5	66.9	10.6
Scenario taking into account correlation	3.4	10.1	0.7

Figure 12: Loss ratios due to teleconsultation according to different scenarios

It shows that there are four levels characterizing the projections of the proportion of teleconsultation procedures among the overall consumption of consultations and teleconsultations:

- The first level includes the most unfavorable projections for the growth of teleconsultation by 2070. It includes ratios ranging from 0.7to3.4%, affecting three of the nine scenarios. The hypothesis that most penalizes the growth of teleconsultation practice by 2070 is the linear correlation between consultation and teleconsultation frequencies.
- The second level includes ratios ranging from 10.1to15.5%. The projections of three (out of nine) development scenarios are part of this level. Given that the share of the care basket taken up by teleconsultation was worth 17.8% in 2020, these scenario crossings are closest to the current teleconsultation claim share.
- The third level is shared by two crossovers and considers that in 2070, the teleconsultation claim rate will occupy about one third of the overall consultation and teleconsultation claim rate (between 37.2 and 38.5 %). This is the intersection of the central consultation scenario and the scenario incorporating the time series model projections for teleconsultation and vice versa.
- Finally, the projection provided by the intersection of the scenarios taking into account the time series model projections of the frequency of consultation and teleconsultation (as well as the application of inflation) constitutes the last level. It predicts that teleconsultation claims will have twice as much weight and influence (66.9%) in the health insurance basket of consultations and teleconsultations as consultation practice, by 2070.

Conclusion

This thesis aimed to study the development of teleconsultation risk in health insurance by proposing: firstly, the implementation of predictive pricing models for consultation and teleconsultation claims experience, in order to analyze their relationship over time; secondly, by projecting claims experience into the future in order to predict the influence that teleconsultation will have in the care basket made up of consultation and teleconsultation procedures.

These studies have shown that the overall claims experience for consultation and teleconsultation is likely to double or even triple in the future, with or without taking into account the annual inflation rate. In addition, it is noted that the differences in claims between the two medical practices in relation to the overall consumption of care will probably increase over time. However, it is noted that the place of teleconsultation in this care basket, in terms of claims experience, will probably remain lower than that of consultation by 2070.

However, given the limited depth of the data history, it would be interesting to continue the work by integrating into the study the consumption data for consultations and teleconsultations dating from 2021, which are now available on the Assurance Maladie website.

Remerciements

J'adresse tout d'abord mes remerciements à Baptiste ANDRIEU et Anne-Emmanuelle ADOU pour m'avoir accompagnée tout au long de la réalisation de ce mémoire. Je les remercie pour m'avoir fait confiance, pour leur expertise, leur investissement et leur soutien sans faille.

Ensuite, je souhaiterais témoigner toute ma reconnaissance à Céline HOUDAYER, Kris-Viime NIKIEMA et Arnaud BUZZI pour leur aide précieuse, leur bienveillance et leur expertise qui ont grandement contribué à améliorer mon travail.

Je remercie également tous les membres de l'équipe Actuariat de Sia Partners pour leur accueil chaleureux et leur mobilisation pour le bon déroulement de ce travail.

J'aimerais remercier l'ensemble du corps professoral du Master Actuariat de l'Université Paris Dauphine pour la qualité de l'enseignement dispensé ainsi que Christophe DUTANG, mon tuteur universitaire, pour son encadrement, ses relectures et ses conseils.

Finalement, je tiens à témoigner toute ma gratitude à ma famille et à mes proches pour leur encouragement et leur soutien moral inestimable.

Table des matières

Résumé	3
Abstract	4
Note de Synthèse	5
Synthesis note	13
Remerciements	21
Table des matières	23
Introduction	25
1 Contexte du mémoire	27
1.1 Organisation du système de santé en France	27
1.2 Zoom sur la téléconsultation	36
2 Traitement et analyse des données	51
2.1 Construction de la base « sinistres »	51
2.2 Construction de la base « assurés »	59
2.3 Création des bases de modélisation finales	64
2.4 Analyses statistiques des bases de données obtenues	65
3 Analyse de la tarification d'un acte de téléconsultation	77
3.1 Rappels sur la tarification	77
3.2 Théorie des Modèles linéaires généralisés	80
3.3 Modélisation de la sinistralité par une approche coût-fréquence	88

4 Modélisation de la sinistralité future	103
4.1 Modélisation à l'aide des séries temporelles	103
4.2 Calibration de scénarios d'évolution de la consommation	114
Conclusion	127
Bibliographie	129
Annexes	131
A.1 Annexe 1 : Référentiel d'équivalence	131
A.2 Annexe 2 : Modélisation GLM de la sinistralité en consultation	133
A.3 Annexe 3 : Influence des scénarios de projection sur les sinistralités futures	136

Introduction

Les innovations technologiques, les objets connectés et les possibilités véhiculées par la transformation numérique, combinés aux nouveaux modes de vie, modifient en profondeur le rapport à la santé et ouvrent de nouvelles opportunités. Parmi ces dernières, la réalisation d'une consultation médicale à distance par écrans interposés, pratiquée par un médecin généraliste ou bien spécialiste, appelée plus couramment la téléconsultation, constitue un vecteur d'évolution majeur dans le cadre de la consommation de soins en santé.

Après des débuts timides, la crise du Covid a permis à la téléconsultation de connaître un véritable essor en France en 2020, en devenant un outil essentiel pour l'accompagnement et le suivi des patients, en l'espace de quelques semaines, entre mars et avril 2020. Pour autant que l'on dispose d'un smartphone, d'une tablette ou d'un ordinateur, elle offre une réponse immédiate aux problématiques majeures liées aux pratiques médicales, comme la difficulté d'accès aux soins et l'amélioration de la qualité de la prise en charge et du suivi des patients. Le premier point englobe notamment la question des déserts médicaux, ainsi que le renoncement aux soins chez un spécialiste dû à un délai d'attente conséquent pour accéder à une consultation. Le dernier inclut, entre autres, la réduction des coûts, le désengorgement des services médicaux, ainsi que les recours abusifs aux urgences.

Cette nouvelle pratique et les évolutions de consommation de soins associées représentent toutefois pour l'assureur une incertitude quant à l'évolution de la consommation par postes de soins, notamment les postes de soins courants et de télé médecine. Ils soulèvent ainsi le besoin, dans un contexte de maîtrise des risques futurs via des techniques de tarification, d'en étudier l'essor et l'amplitude actuelle, afin d'en saisir l'étendue et les conséquences futures.

C'est dans ce contexte que le mémoire propose d'étudier l'évolution de la consommation de ces actes de soins. A cet effet, dans un premier temps la sinistralité historique des prestations de téléconsultation sera analysée à l'aide de modèles tarifaires usuels. Cette analyse s'effectuera en comparaison à la pratique de consultation, à partir des données de consommation mises à disposition par la Sécurité sociale. Dans un deuxième temps, les travaux se concentreront sur la modélisation de la sinistralité future à travers tout d'abord la mise en place de modèles de séries temporelles des fréquences de survenance de téléconsultation et de consultation, puis la calibration de scénarios de projection de la consommation aux modèles précédemment obtenus.

Chapitre 1

L'inscription de la téléconsultation au sein du système de santé français

Ce premier chapitre a pour objectif de présenter le contexte dans lequel se place cette étude. Il permettra au lecteur de se familiariser premièrement avec les notions de Sécurité sociale et de complémentaire santé. De surcroît, il sera question de développer la notion de téléconsultation afin de saisir les facteurs qui ont engendré son remarquable essor et l'enjeu que cette récente pratique médicale représente dans le cadre du système de soins.

1.1 Organisation du système de santé en France

1.1.1 Description du risque santé

La santé est l'un des risques sociaux couverts par la protection sociale, au même titre, par exemple, que la vieillesse, la retraite ou encore les accidents de travail. Pour prendre en charge ce risque, la France s'est dotée d'un système de couverture santé universelle, depuis 1945. Ce dernier peut être défini, selon l'Organisation Mondiale de la Santé (OMS), comme « l'ensemble des organisations, des institutions, des ressources et des personnes dont l'objectif principal est d'améliorer la santé [...] ». L'objectif du système de santé français est d'assurer un accès aux soins pour tous, sans pour autant engendrer de difficultés financières. Différentes entités, présentées ci-dessous, œuvrent pour assurer cette couverture sanitaire.

Acteurs du système de soins

En France, nous distinguons six grandes catégories d'acteurs du système de santé.

1. Les offreurs de soins, qui regroupent :

- les professions médicales et pharmaceutiques (médecins, pharmaciens, chirurgiens-dentistes, sages-femmes), ainsi que les auxiliaires médicaux (infirmières, masseurs-kinésithérapeutes, orthophonistes, etc.) aux statuts et aux modes d'exercices divers : libéraux, salariés ou mixtes ;
- les établissements de santé couvrant des champs d'intervention différents (hospitalier, médico-social), avec des statuts qui peuvent être publics ou privés, lucratifs ou non lucratifs ;
- les réseaux de santé pluridisciplinaires regroupant des médecins, des infirmières et d'autres profes-

sionnels (travailleurs sociaux, personnel administratif, etc.), développés pour favoriser l'accès aux soins, la coordination, la continuité ou encore l'interdisciplinarité des prises en charge (d'après l'article L6321-1 du Code de la santé publique) ;

- les structures de prévention : services de santé au travail, médecine scolaire, services de protection maternelle et infantile, structures de dépistage, etc.

2. Les producteurs de biens et services en santé, telle que l'industrie pharmaceutique, qui obéissent à des logiques économiques de marché.

3. L'État par le biais d'institutions publiques qui :

- organisent le système tant au niveau national (ministère en charge de la santé) que régional (agences régionales de santé – ARS) ou départemental (conseils départementaux dans le domaine de l'action sanitaire et sociale) ;
- conseillent ou aident à l'organisation : le Haut Conseil de la santé publique, la Santé publique France et l'Agence nationale sont autant d'appuis à la performance des établissements de santé et médico-sociaux ;
- contrôlent et orientent : le Parlement au travers des lois de financement de la Sécurité sociale et de l'Objectif national de dépenses d'assurance maladie (Ondam) ; les corps de contrôle, notamment la Cour des comptes, l'Inspection générale des affaires sociales (Igas).

4. Des bénéficiaires, c'est-à-dire des usagers, des bénévoles à l'hôpital, des associations du domaine médico-social qui font vivre la démocratie sanitaire, dont l'objectif est d'associer l'ensemble des acteurs de santé pour élaborer et mettre en œuvre la politique de santé, dans un esprit de dialogue et de concertation.

5. La Sécurité sociale

La Sécurité sociale est un service public qui a pour mission de protéger les personnes résidant en France contre les risques sociaux de la vie. Elle s'organise en 2 principaux régimes et des régimes spéciaux, qualifiés de régimes de base. Ces régimes sont obligatoires et couvrent chacune des catégories socio-professionnelles différentes :

- Le régime général : il regroupe depuis le 1er janvier 2018, les travailleurs salariés, les personnes bénéficiant de la Protection Universelle Maladie (anciennement CMU) et les travailleurs indépendants ;
- Le régime agricole : il regroupe les salariés et exploitants agricoles. Il est également connu sous le nom de Mutualité Sociale Agricole ;
- Les régimes spéciaux : ces régimes fonctionnent sur la base d'une solidarité restreinte à une profession (Régimes des marins, des militaires, des cultes, ...) ou à une entreprise (Régimes de l'Opéra de Paris, de la SNCF, de la RATP, ...) et proposent souvent des prestations plus intéressantes que celles du régime général.

Le régime général se compose lui-même de 5 branches, chacune représentant un risque social différent :

- La branche Famille, gérée par les Allocations familiales, vise à réduire les inégalités de niveau de vie entre les ménages ;
- La branche Retraite, gérée par l'Assurance Retraite, verse les pensions aux retraités de l'industrie, des services et du commerce. Présente dès leur premier emploi, elle suit les salariés tout au long de leur carrière et les aide à préparer leur retraite ;
- La branche Accidents du travail – maladies professionnelles (AT-MP), gérée par l'Assurance Maladie, l'assureur solidaire de plus de 85% de la population, couvre les risques professionnels auxquels sont confrontés les travailleurs ;
- La branche Recouvrement collecte auprès des entreprises, des travailleurs indépendants et des particuliers les cotisations et contributions sociales, au bénéfice des autres branches. Ces collectes permettent ainsi de financer l'ensemble des prestations (famille, maladie, retraite) ;
- La branche Maladie, également gérée par l'Assurance Maladie, assure la prise en charge des dépenses de santé de 60 millions d'assurés tout au long de leur vie quels que soient leurs ressources, leur situation ou leur état de santé. Elle couvre notamment les risques de maladie, de maternité, d'invalidité et de décès en plus du risque santé. De plus, elle agit en amont de la maladie et de ses complications, avec une offre de prévention sans cesse enrichie. Dans le milieu professionnel aussi, elle renforce ses dispositifs pour protéger les salariés les plus exposés ou les plus fragiles.

En guise de synthèse, le schéma ci-dessous permet d'appréhender le fonctionnement de la Sécurité sociale.

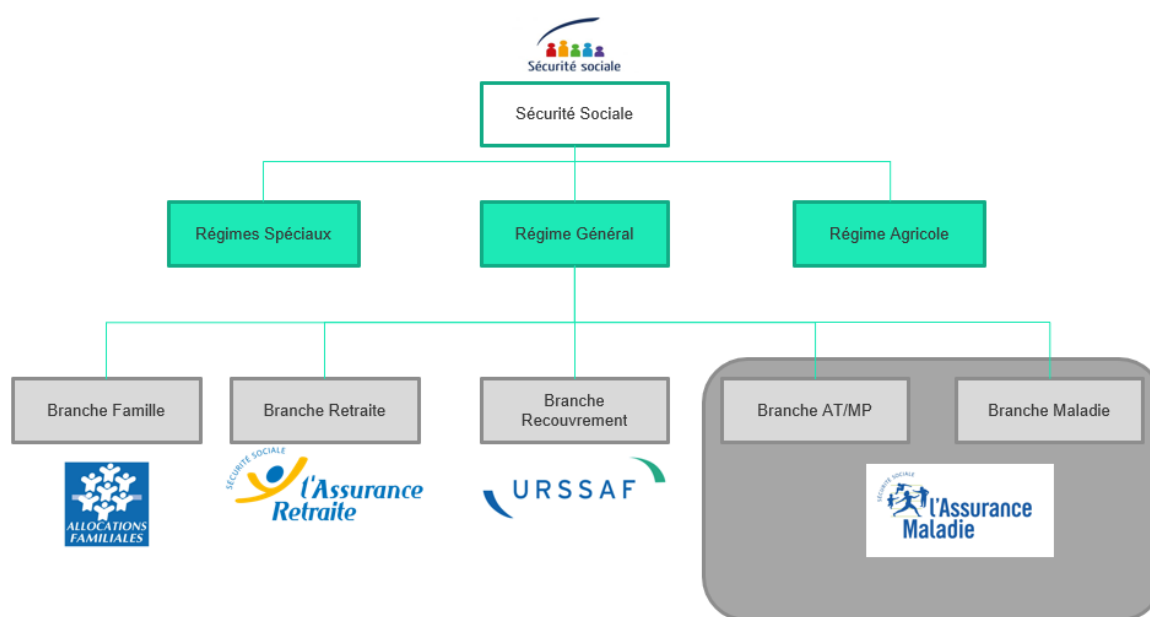


FIGURE 1.1 : Organisation de la Sécurité sociale

6. Assurance complémentaire santé :

La Sécurité sociale n'assure pas l'intégralité du risque Maladie. En effet, en contrepartie du versement d'une cotisation, l'assurance complémentaire en santé assure une partie de la dépense de soins et de biens médicaux non remboursés par l'assurance maladie obligatoire et permet donc de minimiser le reste à charge des assurés. Elle est souscrite individuellement ou collectivement à travers une entreprise et est fondée sur la liberté contractuelle. Actuellement, l'assurance maladie complémentaire assure plus de 95% de la population.

Il existe trois types d'Organismes d'Assurance Maladie Complémentaire (OCAM) ayant droit de proposer des couvertures complémentaires de santé :

- Les Mutuelles : ce sont des groupements de personnes à but non lucratif régis par le Code de la mutualité. Elles sont directement gérées par leurs assurés, chacun disposant d'une voix à l'Assemblée générale ou pour élire les délégués à l'Assemblée générale ;
- Les Sociétés d'assurance : elles sont constituées sous la forme de sociétés de capitaux permettant la réalisation de bénéfices et la rémunération d'actionnaires. Elles sont régies par le Code des assurances, qui lui n'encadre pas la représentation des assurés et des souscripteurs aux instances de gouvernance ;
- Les Institutions de prévoyance : ce sont des sociétés de personnes morales de droit privé, régies par le code de la Sécurité sociale. Elles sont caractérisées par une gestion paritaire et sont également à but non lucratif. Elles gèrent des contrats collectifs d'assurance de personnes, souscrits par les entreprises au bénéfice des salariés, dans le cadre exclusif des accords d'entreprises ou de branches professionnelles.

La figure 1.2 montre que la répartition du marché est inégale entre les trois organismes. Nous observons que les mutuelles dominent le marché que ce soit en nombre ou en chiffre d'affaires.

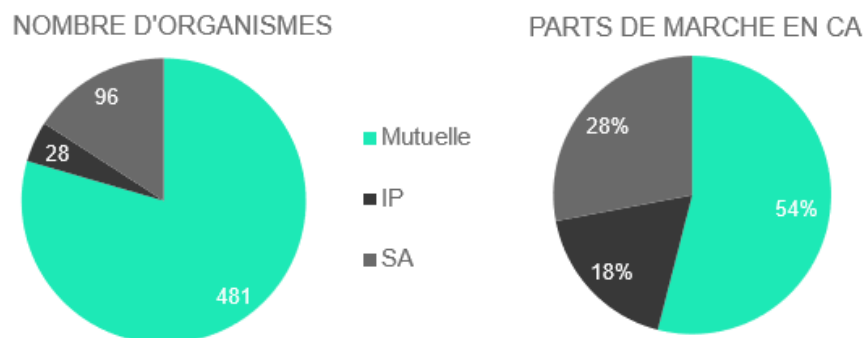


FIGURE 1.2 : Répartition du marché assurantiel entre les trois organismes

Synthèse

La multiplicité d'acteurs aux périmètres d'intervention et aux statuts très différents rend le système de santé français à la fois compétent dans sa capacité à répondre aux demandes et à couvrir les besoins de santé de la population, mais aussi particulièrement complexe à réguler. Le schéma ci-dessous a pour objectif de résumer les liens entre les principaux acteurs du système de soins, afin d'obtenir une vue d'ensemble de celui-ci.

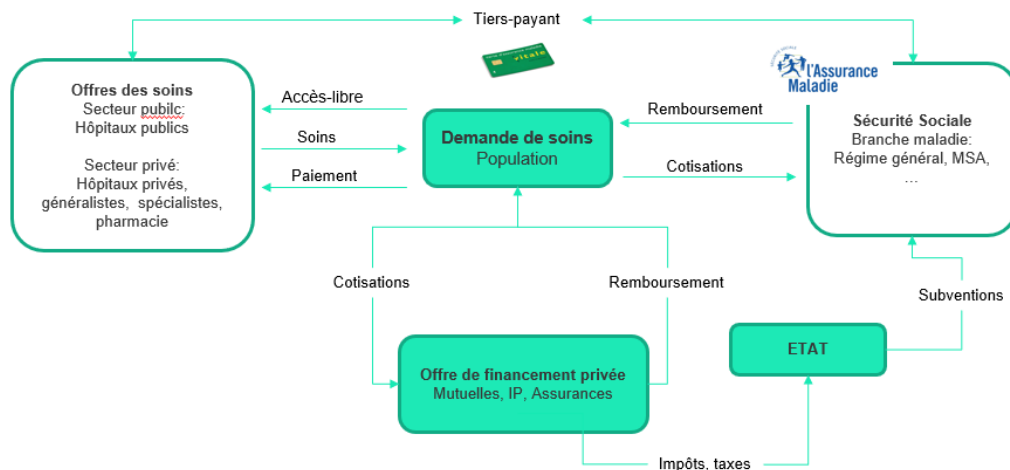


FIGURE 1.3 : Principaux acteurs du système de soins

1.1.2 Les principales sources de financement de la Sécurité sociale

En 2022, le montant prévisionnel des recettes de la Sécurité sociale, approuvé par le Projet de loi de financement de la sécurité sociale (PLFSS), s'élève à 438,2 milliards d'euros pour le régime général et le fonds de solidarité vieillesse. La figure suivante illustre la répartition des ressources de la Sécurité sociale.

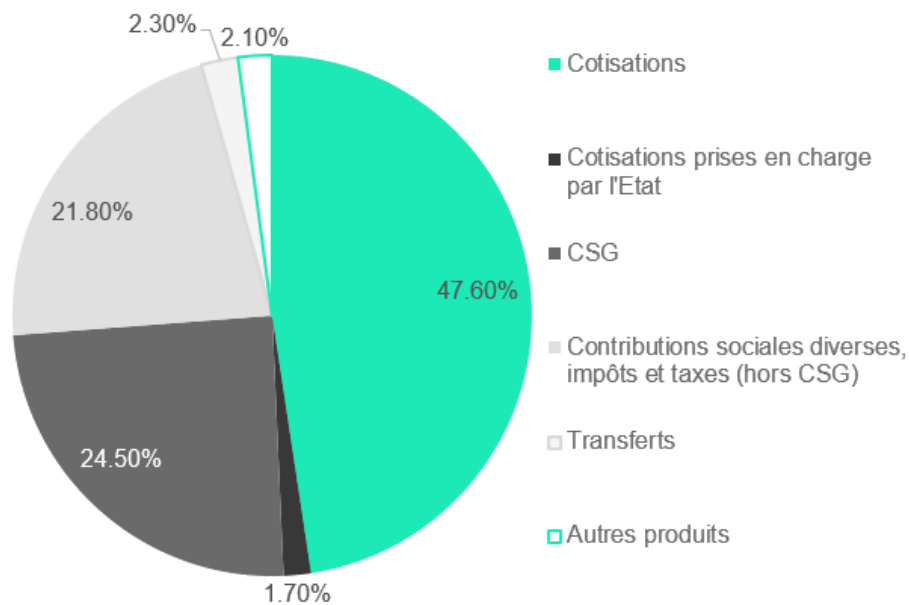


FIGURE 1.4 : Ressources de la Sécurité sociale *

Ainsi, on observe que le financement de la Sécurité sociale est principalement assuré par des cotisations sociales versées à l'Union de Recouvrement des cotisations de Sécurité Sociale et d'Al-

*Source : Commission des Comptes de la Sécurité sociale, juin 2021

locations Familiales (URSSAF). Historiquement, l'URSSAF a été créée dans le but de collecter les cotisations et les contributions destinées au financement de la Sécurité sociale. Cet organisme forme, avec l'ACOSS (Agence Centrale des Organismes de Sécurité Sociale), la branche du Recouvrement du régime général de la Sécurité sociale. Les cotisations sociales sont des versements obligatoires effectués par les non-salariés, les employeurs et leurs salariés afin d'acquiescer des droits aux prestations sociales. Cela concerne, par exemple, des cotisations d'assurance maladie-maternité-invalidité-décès ou d'assurance retraite.

La deuxième plus grande contribution est assurée par le financement fiscal, qui constitue une part croissante des ressources de la Sécurité sociale. Les « impôts et taxes affectés » sont des ressources fiscales affectées au financement des prestations sociales ou au financement de dépenses spécifiques.

Peuvent ainsi être distingués :

- l'impôt sur les produits (droits de consommation sur les alcools et le tabac, taxes sur les assurances automobiles, taxes sur les activités polluantes, ...);
- les impôts sur le revenu et le patrimoine. La Contribution Sociale Généralisée (CSG), est la principale ressource de ce type.

Les contributions publiques de l'État financent des dépenses de solidarité (tel que le Revenu Minimum d'Insertion (RMI) ou encore les Fonds de solidarité vieillesse), une partie des exonérations de cotisations employeurs pour les bas salaires, et subventionnent des régimes de façon permanente ou en cas de difficultés (par exemple les régimes de retraite de certaines professions dans lesquelles le nombre d'actifs cotisants est inférieur au nombre de retraités).

1.1.3 Le remboursement et les postes de soins

Le risque santé, selon l'Assurance Maladie, peut être subdivisé en différents postes de soins, chacun d'entre eux comprenant plusieurs actes médicaux, comme illustré dans la figure 1.5.

Pour chaque type d'acte référencé dans la Classification Commune des Actes Médicaux (CCAM), plusieurs éléments interviennent dans le calcul du remboursement des frais médicaux associés par la Sécurité sociale :

- La Base de Remboursement de la Sécurité Sociale (BRSS) ou Tarif Conventionnel servant à définir la somme remboursée par la Sécurité sociale après un acte médical. Cette dernière remboursera ainsi un pourcentage variable de la BRSS de 30 à 100% (taux de remboursement);
- La Participation Forfaitaire (PF) correspondant à un montant fixe de 1€ restant à la charge de l'assuré;
- Le Ticket Modérateur (TM) définit comme la part de base de remboursement restant à la charge de l'assuré après le remboursement de l'Assurance Maladie et la participation forfaitaire. Il est défini comme un pourcentage de la BRSS et dépend du type de dépense;
- Le montant de Frais Réels (FR) correspondant au coût total de l'acte;
- Lorsque les frais réels sont supérieurs à la base de remboursement de la Sécurité sociale, on définit le dépassement d'honoraires qui correspond à la différence entre ces deux montants;



FIGURE 1.5 : Postes de soins en Santé-Prévoyance

- Lorsque les frais réels sont supérieurs à la base de remboursement de la Sécurité sociale, on définit le dépassement d'honoraires qui correspond à la différence entre ces deux montants ;
- Et finalement le reste à charge correspondant au montant payé par l'assuré, déduction faite du ou des remboursements. C'est à ce stade qu'interviennent les assurances complémentaires santé.

La structure du processus de remboursement des frais de Santé en France est illustrée dans la figure suivante.

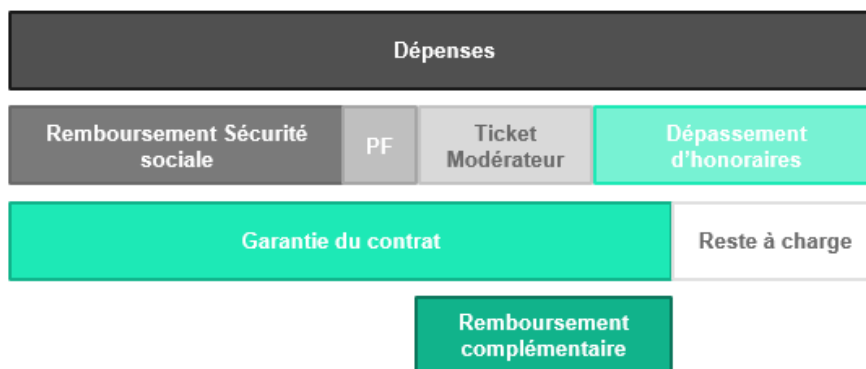


FIGURE 1.6 : Mécanisme de remboursement des garanties Santé

Selon le poste d'indemnisation, les remboursements du régime complémentaire peuvent s'exprimer :

- en pourcentage des frais réels,
- en pourcentage du remboursement de la Sécurité sociale,

- en pourcentage de la base de remboursement de la Sécurité social
- en pourcentage du ticket modérateur,
- sous la forme d'un forfait en euros par acte, semestriel ou annuel.

Enfin, il est important de préciser que les montants remboursés par l'Assurance Maladie dépendent également de la situation de l'assuré dans le parcours de soins coordonnés et du secteur conventionnel du médecin. En effet, si un assuré est hors du parcours de soins, c'est-à-dire s'il n'a pas déclaré son médecin traitant ou s'il consulte directement un médecin sans être orienté par son médecin traitant, depuis le 1er janvier 2009, le montant de ses remboursements est diminué. De plus, un médecin conventionné de secteur 1 applique le tarif conventionnel, alors qu'un médecin de secteur 2 pratique des honoraires libres. Dans ce cas, les dépassements d'honoraires ne sont pas remboursés par l'Assurance Maladie.

1.1.4 Prestations de solidarité

Depuis 1999, afin de faciliter l'accès aux soins des ménages aux revenus modestes, l'Etat a mis en place plusieurs dispositifs de solidarité leur permettant d'obtenir des assurances complémentaires santé peu coûteuses. Notamment, la Couverture Maladie Universelle Complémentaire, appelée CMU-C, était une couverture maladie complémentaire gratuite dont l'accès dépendait du plafond appliqué sur les ressources annuelles des ménages.

L'affiliation à cette complémentaire incluait : la prise en charge du ticket modérateur, la prise en charge du forfait journalier hospitalier sans limitation de durée, et l'exonération des différentes participations forfaitaires et franchises. Les personnes qui bénéficiaient de la CMU-C n'étaient pas soumises au dépassement d'honoraires et bénéficiaient de tarifs opposables pour l'optique, les prothèses dentaires et autres prestations médicales.

Les personnes dont le revenu était compris entre le plafond annuel de ressources de la CMU-C, et ce même plafond majoré de 35 %, pouvaient jouir d'une aide pour payer leur complémentaire santé en souscrivant un contrat labellisé, depuis 2015. Il s'agissait de l'Aide au paiement d'une Complémentaire Santé, nommée ACS. Cette aide permettait de réduire voire prendre en charge en intégralité le montant de la cotisation annuelle de la complémentaire. Dans le cadre d'un parcours de soins coordonnés, les médecins ne pouvaient pas appliquer de dépassements d'honoraires.

Suite au Projet de loi de financement de la sécurité sociale (PLFSS), le 1er novembre 2019, la CMU-C et l'ACS ont fusionné pour créer la Complémentaire Santé Solidaire (CSS) : les bénéficiaires de l'ACS ont maintenant des garanties identiques au panier de soins de la CMU-C. La CSS reste gratuite jusqu'aux plafonds actuels de la CMU-C et est soumise à participation financière pour les personnes disposant de revenus compris entre les plafonds actuels de la CMU-C et de l'ACS. Cette participation varie en fonction de l'âge. Avec cette réforme, le gouvernement a souhaité faciliter l'accès à l'ACS au plus grand nombre par un processus de mise en concurrence visant à offrir une plus grande lisibilité des offres de complémentaire santé et de meilleures garanties à des prix plus bas.

On peut souligner aussi l'efficacité du dispositif de complémentaire solidaire quant à la favorisation de l'accès aux soins. A cet égard, une étude de la Direction de la recherche des études de l'évaluation et des statistiques (DREES), publiée en 2015, révèle que les bénéficiaires de la CMU-C ont une probabilité plus forte de recourir à des soins dans l'année par rapport aux assurés dépourvus d'assurance complémentaire ou aux assurés bénéficiaires d'autres complémentaires, qui présentent un profil socio-économique et un état de santé équivalent.

Le bénéfice de la CMU-C a une incidence pour l'ensemble des postes de soins, à l'exception des courts séjours à l'hôpital (figure 1.7). Cela s'explique tout d'abord du fait que ce poste de soins est amplement pris en charge par le régime obligatoire de l'Assurance Maladie. De plus, le fait que les avantages de la CMU-C n'influent pas sur la dépense liée aux hospitalisations rentre en jeu, pour les mêmes raisons qu'il n'a pas d'effet sur la probabilité de recourir à ces soins.

À l'inverse, la figure 1.7 montre que pour l'optique et les soins dentaires, faiblement pris en charge par l'Assurance Maladie, l'effet de la CMU-C est accentué. A ce effet, la probabilité de recourir à ces soins augmente respectivement de 12 points et de 13 points. La probabilité est également nettement augmentée pour les soins de généralistes et de spécialistes, par rapport aux assurés couverts par d'autres complémentaires santé.

Poste de soins	Effet sur la probabilité de recourir aux soins dans l'année (en points de %)		Probabilité moyenne de recourir aux soins dans l'année pour les personnes sans complémentaires (en %)	Effet sur la dépense annuelle moyenne en cas de recours (en %)	
	Complémentaires hors CMU-C	CMU-C		Complémentaire hors CMU-C	CMU-C
Hospitalisation	ns	ns	10	ns	ns
Ambulance	+4	+5	85	+24	+34
Médecins	+6	+9	80	+19	+38
Généralistes	+8	+12	74	ns (*)	+43 (*)
Spécialistes	+10	+13	51	+32	+31
Pharmacie	+9	+11	76	+27	+43
Biologie	+9	+9	35	ns	ns
Optique	+12	+12	8	ns (*)	-66 (*)
Dentaire	+14	+13	28	+58	+57
Dentaire (y c. dépassements)				+72	+67
Soins conservateurs	+13	+11	28	+34	+36
Soins conservateur (y c. dépassements)				+35	+37
Soins prothétiques	ns	ns	5	ns	ns
Soins prothétiques (y c. dépassements)				ns	ns
Tous	+5	+6	85		

FIGURE 1.7 : Effet de la complémentaire CMU-C sur la probabilité de recourir aux soins et sur la dépense en cas de recours, par rapport à la non-couverture

On remarque également que les dépenses des bénéficiaires de la CMU-C ayant vu un généraliste dans l'année sont, toutes choses égales par ailleurs, plus élevées de 43 % que celles des personnes sans couverture complémentaire qui en ont vu un dans l'année.

Il a ainsi été observé que le système de santé français constitue un véritable enjeu économique et sociétal, au regard des conséquences directes pour la population. De plus, l'Etat est un acteur essentiel du système de santé. En effet, c'est lui qui définit les politiques de santé par le biais de lois majeures telles que la loi Kouchner du 4 mars 2002, relative aux droits des malades et à la qualité du système de santé, ou encore la loi HPST (pour Hôpital Patients Santé et Territoires) du 21 juillet 2009. L'Etat définit également les politiques de financement via la loi de financement de la Sécurité sociale, votée chaque année. Son objectif est de tout mettre en œuvre afin d'assurer la plus grande égalité face à l'accès aux soins, améliorer la qualité des soins, promouvoir la santé publique, et moderniser le système de santé français. Pour répondre à certains de ces défis, notamment l'égal accès aux soins sur l'ensemble du territoire, le numérique apporte de nouveaux usages comme la télémédecine, une pratique innovante pour les patients comme pour les professionnels de santé.

1.2 Zoom sur la téléconsultation

1.2.1 Présentation de la télé médecine

La télé médecine est un acte médical à distance utilisant les technologies de l'information et de la communication. Elle met en relation, entre eux ou avec un patient, un ou plusieurs professionnels de santé, parmi lesquels figurent nécessairement un professionnel de santé médical. Elle permet d'établir un diagnostic, d'assurer, pour un patient à risque, un suivi à visée préventive ou un suivi post-thérapeutique, de requérir un avis spécialisé, de préparer une décision thérapeutique, ou encore de prescrire des médicaments. La télé médecine impose les mêmes exigences et la même rigueur en termes de qualité de soins et devient une manière de soigner à part entière.

La télé médecine n'est pas une pratique nouvelle. Les États-Unis ont été l'un des pays pionniers : cela fait plus d'un siècle que les premières expériences dans ce domaine ont été menées. Voici un état des lieux de la télé médecine, et plus précisément, de l'historique de son apparition jusqu'à son utilisation de nos jours :

- 1920 : Utilisation de la télé assistance médicale par téléphone pour permettre les soins durant les croisières transatlantiques ;
- Après-guerre : Premières initiatives de télé médecine en France dans les Midi-Pyrénées sous la direction du Dr Louis Lareng (fondateur du SAMU) ;
- 1950 : Arrivée des transferts d'imageries médicales (radiologie) avec le travail du radiologue Jacob Gershon-Cohen ;
- 1959 : Essor de la télé psychiatrie, notamment à travers la prise en charge des vétérans de la guerre du Vietnam par vidéo ;
- 1964 : Début de la formation à distance par vidéo des professionnels de santé américains ;
- 1973 : Institutionnalisation de la discipline avec le premier Congrès International de Télé médecine ;
- 1998 : Reconnaissance de la télé médecine par l'Organisation Mondiale de la Santé (OMS) ;
- 2005 : L'OMS commence à réglementer l'e-santé et les télésoins à l'échelle mondiale ;
- 2009 : Légifération de la télé médecine par la Loi Hôpital, Patient, Santé et Territoire (HPST) en différenciant 5 actes de télé médecine : télé expertise, télé consultation, télé surveillance, télé assistance et régulation ;
- 2012 : Lancement du Pacte territoire santé 2012-2015 pour lutter contre les déserts médicaux. Parmi ses douze engagements, figure le développement de la télé médecine. Un second plan lui succède jusqu'en 2017 ; il a pour but de faciliter l'accès à la télé médecine pour les patients atteints de maladies chroniques et les soins urgents ;
- 2017 : Instauration de l'expérimentation de la télé médecine pour une durée de quatre ans à compter du 1er janvier 2014 par la loi de financement de la sécurité sociale (LFSS), à travers le dispositif ETAPES (Expérimentations de télé médecine pour l'amélioration des parcours en santé), qui encourage et soutient financièrement le déploiement de projets de télésanté. Initialement menées dans neuf régions, ces expérimentations s'étendent en 2017 à tout le territoire.

Ce dispositif est reconduit pour quatre ans par la LFSS en 2018 (art. 54) et porte sur la télésurveillance de certaines affections de longue durée (ALD). Ces expérimentations sont sujettes à l'évaluation de la Haute Autorité de santé (HAS), dont le rôle est de valider ou non celles-ci, en vue d'une généralisation de la télémedecine dans l'ensemble du territoire ;

- Septembre 2018 : Entrée la téléconsultation dans la nomenclature de la Sécurité sociale. L'Assurance Maladie annonce le remboursement des téléconsultations sous certaines conditions : la téléconsultation doit s'inscrire dans le respect du parcours de soins coordonné, entre le patient et son médecin traitant, à condition qu'il y ait eu au moins une consultation physique avec lui au cours des 12 derniers mois précédant la téléconsultation ;
- Septembre 2020 : Dans le contexte pandémique du Coronavirus, le Projet de loi de financement de la Sécurité Sociale 2021 (PLFSS) prévoit le remboursement intégral de toutes les téléconsultations jusqu'à la fin de l'année 2021, par l'Assurance Maladie ;
- 31 Décembre 2021 : Prolongement de la prise en charge intégrale de la téléconsultation par téléphone jusqu'à la mi-2022 ;
- 06 Avril 2022 : Publication d'une charte de la téléconsultation par l'Assurance maladie.

Le décret du 19 octobre 2010, pris en application de la loi HPST définit cinq pratiques médicales pouvant être exercées en télémedecine et leurs conditions d'exercice :

- La téléexpertise : avis sollicité à distance par un médecin auprès d'un ou de plusieurs professionnels médicaux en raison de leurs formations ou de leurs compétences spécifiques, sur la base des données médicales liées à la prise en charge d'un patient, avec son consentement. Elle permet d'obtenir rapidement l'avis d'un spécialiste, donc de réduire les délais de prise en charge et de suivi ;
- La télésurveillance médicale : interprétation à distance des données biologiques et/ou cliniques essentielles au suivi médical d'un patient, et le cas échéant, prise de toutes les décisions nécessaires à la prise en charge de ce patient ;
- La téléassistance médicale : assistance à distance d'un médecin réalisée par l'un de ses confrères pendant un acte médical ou chirurgical ;
- La régulation médicale : réponse médicale apportée par les centres 15 (SAMU) au titre des services d'aide médicale urgente et de la permanence des soins ambulatoires ;
- La téléconsultation : consultation réalisée par un médecin de toute spécialité médicale, à distance d'un patient, ce dernier pouvant être assisté ou non par un autre professionnel de santé (ex : médecin, infirmier, pharmacien...). La téléconsultation implique nécessairement l'utilisation de la vidéo-transmission, via un ordinateur, une tablette ou encore un smartphone. Celle-ci permet alors au médecin d'examiner tous symptômes permettant d'établir un diagnostic. Le patient, grâce aux technologies de la communication, aura également la possibilité de discuter avec son médecin.

Les travaux seront plus particulièrement centrés autour de cette dernière pratique médicale, qui, en l'espace de quelques années, a considérablement pris de l'ampleur dans le paysage médical.

1.2.2 Réglementation et organisation

Le 6 avril 2022, l'Assurance Maladie a publié une charte de bonnes pratiques de la téléconsultation listant les recommandations et obligations de l'acte médical conditionnant sa prise en charge par l'Assurance maladie.

Cette charte de bonnes pratiques s'adresse ainsi à tous les médecins, quelles que soient leur spécialité et leur statut (libéraux, salariés, activité mixte, etc), et s'organise en trois chapitres listant les recommandations et obligations sans lesquelles la prise en charge par l'Assurance maladie ne peut avoir lieu :

- les conditions requises pour la pratique de la téléconsultation : des enjeux de qualité et de sécurité des soins ;
- les principes conditionnant la prise en charge par l'Assurance maladie de la téléconsultation ;
- la spécificité des sociétés proposant des télécabines de téléconsultation et autres offreurs de solutions de télémédecine.

Tout patient, qu'il soit atteint d'une maladie chronique ou pas, peut a priori se voir proposer une téléconsultation. Néanmoins, le recours à la téléconsultation relève de la seule volonté du médecin (traitant, ou correspondant, selon les cas) qui doit évaluer de l'utilité et l'efficacité d'une prise en charge médicale à distance plutôt qu'un suivi traditionnel en face-à-face.

Le respect du parcours de soins coordonné est nécessaire pour ouvrir le droit au remboursement de la téléconsultation. En effet, celui-ci est structuré sur la base du dispositif du médecin traitant (ESAPCE Médecins), qui impose une orientation initiale du patient par son médecin traitant vers le médecin téléconsultant (si celui-ci n'est pas le médecin téléconsultant), et ce, pour assurer une prise en charge de qualité. Cependant, l'Assurance Maladie indique que les types de patients et de situations suivants constituent des exceptions à cette règle :

- « l'accès direct aux médecins spécialistes en ces domaines : gynécologie, ophtalmologie, stomatologie, chirurgie orale ou en chirurgie maxillo-faciale, psychiatrie ou neuropsychiatrie et pédiatrie ;
- les patients âgés de moins de 16 ans ;
- une situation d'urgence ;
- les patients ne disposant pas de médecin traitant ou dont le médecin traitant est indisponible dans un délai compatible avec leur état de santé. Dans ce cas, il peut être fait exception au parcours de soins et à l'obligation de connaissance préalable du patient par le médecin pratiquant la téléconsultation : le patient doit pour cela se rapprocher d'une organisation territoriale organisée avec des médecins volontaires, pour leur permettre :
 - d'être pris en charge rapidement compte tenu de leurs besoins en soins,
 - d'accéder à un médecin compte tenu de leur éloignement des professionnels de santé,
 - d'être en mesure dans un second temps de désigner un médecin traitant pour leur suivi au long cours et réintégrer ainsi le parcours de soins.

Ces organisations territoriales sont facilement identifiables pour les patients d'un territoire. L'information est disponible auprès de l'Assurance Maladie, de la Mutualité sociale agricole, du ministère

des Solidarités et de la santé, des agences régionales de santé (ARS), des professionnels de santé du territoire et des conseils départementaux et régionaux de l'Ordre des médecins ».

La connaissance du patient par le médecin téléconsultant est aussi nécessaire en tant que gage de qualité de la prestation médicale proposée. Il en découle alors que le patient doit avoir eu au moins une consultation physique avec ce médecin (cabinet, domicile du patient ou établissement de santé) au cours des douze derniers mois précédant la téléconsultation.

Concernant l'organisation d'une téléconsultation, le médecin assure la téléconsultation en informant le patient des modalités de réalisation de la téléconsultation. L'utilisation d'un site ou d'une application sécurisée pour protéger les données médicales, confidentielles et sensibles par nature, est requise. Les principales applications à ce jour sont Doctolib, Qare, ou encore Livi. Le patient peut être seul ou accompagné par un professionnel de santé lors de la téléconsultation, si son état le nécessite, depuis son domicile ou depuis un lieu dédié équipé, tel qu'une maison de santé pluriprofessionnelle, pharmacie équipée d'une cabine ou encore un établissement EHPAD. Ce type de consultation se déroule de la même manière qu'une consultation classique dans un cabinet médical.

Enfin, au sujet des tarifs, l'Assurance maladie déclare que les règles de facturation de la consultation de référence ou coordonnée définie à l'article 28.1 et 28.2 de la convention médicale s'appliquent aux téléconsultations. En effet, la téléconsultation est facturée par le médecin téléconsultant au même tarif qu'une consultation en face-à-face, soit entre 23 et 58,50 euros selon la spécialité et le secteur d'exercice du médecin. Le médecin qui accompagne, le cas échéant, le patient lors d'une téléconsultation réalisée par un autre médecin, peut facturer une consultation, dans les conditions habituelles, parallèlement à la facturation de la téléconsultation par le médecin téléconsultant. Comme une consultation habituelle, les médecins ont la possibilité de facturer un dépassement d'honoraires dans les conditions habituelles (secteur 2, etc.). De plus, les règles de prise en charge sont les mêmes que pour une consultation en présentiel, selon les mêmes taux de prise en charge qu'une consultation classique (70 %).

1.2.3 Téléconsultation, réponse aux défis du système de santé ?

Raison médicale

Premièrement la télémédecine est réputée pour être une innovation technique efficace. Au cours de la dernière décennie, plus de 8 500 études ont été menées pour en évaluer l'impact. L'étude nommée « Le recours à la téléconsultation et à la téléexpertise : quel impact économique attendre en France ? », réalisée par BENTATA (2022) pour l'Institut Economique Molinari, affirme que cette richesse en littérature a mis en avant l'impact positif de la télémédecine sur la santé des patients et de leur portefeuille. Néanmoins, même si des incertitudes demeurent quant à la bonne application de la télémédecine en général, l'usage de la téléconsultation pose de moins en moins de difficultés et son efficacité fait l'objet d'un large consensus.

Par exemple, dans une récente revue, l'Organisation de Coopération et de Développement Economique (OCDE) souligne que la télémédecine permet de fournir des services de soins à des patients n'ayant pas accès aux soins par les voies traditionnelles, comme par exemple les marins dans plusieurs pays européens, les patients en prison au Canada et aux Etats-Unis ou encore les populations originelles en Australie. Cette même revue met en avant le fait que dans la plupart des pays développés, les usages de la télémédecine se sont aussi développés et diversifiés avec succès. En effet, pour les patients diabétiques, la télésurveillance fournit des résultats équivalents aux consultations physiques, tout en limitant le nombre de visites médicales et de consultations d'urgence. Les mêmes conclusions sont faites sur la télésurveillance des patients souffrant de maladies respiratoires. De surcroît, la télésurveillance

diminue la mortalité et les risques d'hospitalisation des patients souffrant de maladies cardiaques comparé à une assistance traditionnelle et apporte les mêmes résultats qu'un suivi à domicile effectué par une infirmière spécialisée.

Bien que la téléconsultation n'ait pas été étudiée au même degré que la télémédecine, son utilisation a engendré un consensus parmi les récentes revues systématiques concernant son efficacité par rapport à la consultation traditionnelle pour traiter les premiers symptômes des patients, aussi bien pour des maladies nécessitant l'avis d'un médecin généraliste que pour des situations requérant l'intervention d'un spécialiste, notamment en dermatologie ou en cardiologie. Cela s'applique aussi pour le traitement des maladies mentales, le suivi des maladies chroniques, notamment des patients souffrant de diabète, le suivi des patients souffrant de maladies cardiaques et la provision de soins aux personnes âgées. Selon l'étude publiée par l'Institut Economique Molinari, ces résultats proviennent principalement du fait que la téléconsultation est adaptée à des examens ne requérant pas de contact physique avec le malade, ainsi que de la réduction des délais de réaction des patients et des médecins.

Pratique plébiscitée durant l'épidémie du Coronavirus

Après des débuts timides, la téléconsultation a connu un véritable essor en France en 2020, avec la pandémie de coronavirus, en devenant un outil essentiel pour l'accompagnement et le suivi des patients en l'espace de quelques semaines. L'important recours à la téléconsultation qui a fait suite au confinement généralisé provoqué par la première vague épidémique de la Covid-19 a mis en avant la réponse complémentaire que la téléconsultation apporte et sa cohérence avec les consultations en présentiel. Dans le contexte particulier de crise sanitaire, elle est un moyen efficace pour réduire les risques de contamination dans les salles d'attente des praticiens et donc réduire les risques pour les patients comme pour les médecins.

Les données de l'Open DAMIR, base de données Open Source de la Sécurité sociale extraite du Système National d'Information Inter Régimes de l'Assurance Maladie (SNIIRAM), rendent compte des dépenses nationales mensuelles de santé en France, notamment celles concernant les remboursements d'actes de téléconsultation effectuées par le régime général sous le poste de soins « soins de villes » pendant la période Covid-19. En effet, la figure 1.8 permet d'observer la part mensuelle des dépenses en téléconsultation sur l'ensemble des consultations médicales réalisées pendant la première année de l'épidémie.

Il peut être en déduit ainsi que :

- En plein confinement (avril 2020), un quart des consultations médicales, soit 4,4 millions d'actes pris en charge par le régime général étaient des téléconsultations ;
- L'usage de la téléconsultation a connu une croissance exponentielle à partir du confinement, passant de 37.000 actes en février 2020 à environ 4.350.000 actes en avril 2020, soit une multiplication par 120 en l'espace de deux mois ;
- Après le pic du 1er confinement, la part des dépenses en téléconsultation par rapport à la consommation totale de consultations s'est stabilisée entre 3 et 5% (soit a été multipliée par 30 par rapport à janvier 2020).



FIGURE 1.8 : Part (%) mensuelle des dépenses en téléconsultation sur l'ensemble des consultations médicales réalisées entre novembre 2019 et octobre 2020 (en date de paiement)

Ces actes ont été effectués très majoritairement par des médecins libéraux, surtout des généralistes (80% de l'ensemble des médecins) mais aussi des psychiatres, des pédiatres, des chirurgiens, des dermatologues et des gynécologues. La psychiatrie est la spécialité médicale la plus représentée avec plus d'un million d'actes réalisés entre janvier et octobre 2020. Toutes spécialités confondues, 28% des téléconsultations touchaient un suivi d'une maladie chronique.

Voici ci-après le détail sur l'utilisation de la téléconsultation en fonction de la spécialité des praticiens médicaux.

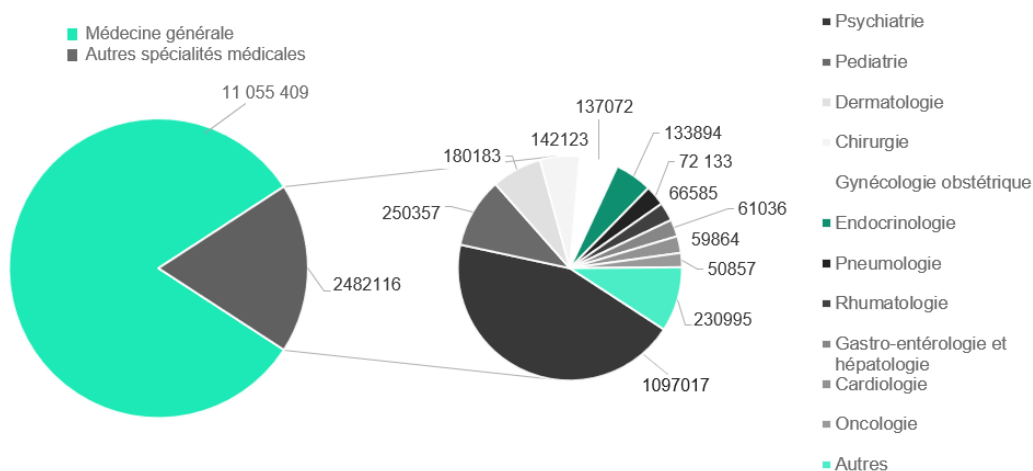


FIGURE 1.9 : Volume de téléconsultations réalisées de janvier à octobre 2020 et poids des spécialités médicales concernées (en date de paiement, hors sage-femme et IDE)

Par ailleurs, l'Institut Economique Molinari a observé que la période de confinement a modifié le profil des patients bénéficiant de la téléconsultation : l'âge moyen des patients ayant recours à la téléconsultation a considérablement augmenté, passant de 35 ans avant confinement à plus de 60

ans après confinement. Bien que ceci s'explique en partie par le fait que la Covid-19 touche plus gravement les patients les plus âgés, cette évolution démontre aussi que la téléconsultation pourrait se substituer à la visite en face à face pour satisfaire les besoins de l'ensemble des patients et non uniquement des plus jeunes. A cet effet, l'Institut Economique Molinari déclare, à travers son étude, que la téléconsultation peut efficacement se substituer à la consultation classique aussi bien pour des visites de routine que pour le suivi de patients atteints de maladies chroniques ou encore pour le suivi des patients souffrant de maladies mentales. Celle-ci pourrait ainsi aisément remplacer les consultations traditionnelles lorsque l'objet de la visite est particulièrement simple, comme c'est le cas par exemple du renouvellement d'ordonnances, de la signature d'un certificat médical ou encore de la prise de rendez-vous avec un médecin spécialiste.

Enfin, il est important de souligner que le développement rapide de la téléconsultation a également été favorisé par l'assouplissement de la réglementation dans le cadre de l'état d'urgence sanitaire instauré par la loi du 23 mars 2020 prévoyant jusqu'au premier juin 2022 :

- La prise en charge à 100% des actes de téléconsultation par l'assurance maladie, de même que pour les actes de téléexpertise, comme mentionné plus tôt ;
- Un assouplissement du respect du parcours de soins. Les consultations à distance ne nécessitent plus la connaissance préalable du médecin téléconsultant et l'orientation vers ce médecin par le médecin traitant n'est plus obligatoire. De plus, afin de lutter contre les disparités d'accès au numérique, la consultation par téléphone a été autorisée pour les femmes enceintes ainsi que pour les résidents en « zone blanche » (déserts médicaux), ne disposant pas d'outil vidéo, en ALD ou âgés de 70 ans et plus.

Néanmoins, outre son efficacité dans un contexte pandémique, d'autres facteurs ont participé au succès de la téléconsultation ainsi qu'à sa reconnaissance en tant que pratique médicale : elle se présente comme une solution contre les problématiques de renoncement aux soins, de déserts médicaux ainsi que de maîtrise des dépenses de santé. En effet, elle possède des externalités positives qui bénéficient à l'ensemble du système de santé.

Déserts médicaux

Selon une récente enquête réalisée par Iqvia France, 27 millions de Français se trouveront dans un désert médical d'ici à 5 ans. Le terme « Désert médical » correspond à l'incapacité ou la très grande difficulté pour les patients à avoir accès aux professionnels de santé sur un territoire, en raison du nombre limité ou de l'absence de médecins proches de leur lieu d'habitation. Cette situation engendre diverses conséquences, telles que :

- l'augmentation des "files d'attentes" avant de pouvoir consulter, au risque d'une dégradation de la situation médicale des patients, ainsi que de la qualité de leur prise en charge ;
- l'engorgement des urgences médicales causé soit par l'utilisation excessive des services d'urgence par les patients ne trouvant pas de professionnels de santé disponibles en consultation classique, soit parce que l'absence de consultations régulières aggrave les situations de santé ;
- des difficultés majeures pour les patients à trouver un nouveau médecin traitant, notamment lorsqu'un praticien part en retraite. En cas de non-remplacement ou de refus par les médecins en exercice de prendre de nouveaux patients, les assurés sont alors privés du parcours de soins coordonné, qui permet un meilleur remboursement des frais de santé et un meilleur suivi médical, notamment pour les personnes atteintes de maladies chroniques.

La notion de désert médical comprend une double dimension :

- une dimension spatiale d'accès à un professionnel de santé, c'est-à-dire sa proximité et le temps de trajet que doit mettre un patient pour pouvoir le consulter. Les indicateurs de temps d'accès au plus proche professionnel fournissent une information immédiatement parlante, fondamentale du point de vue de l'accessibilité physique ;
- une dimension temporelle caractérisant le temps d'attente avant de pouvoir obtenir un rendez-vous, qui soulève une problématique de disponibilité de l'offre de soins, et qui expose la tension entre offre et demande. La densité médicale est un bon indicateur pour cette dimension.

Ces deux dimensions sont illustrées sur les cartographies suivantes.

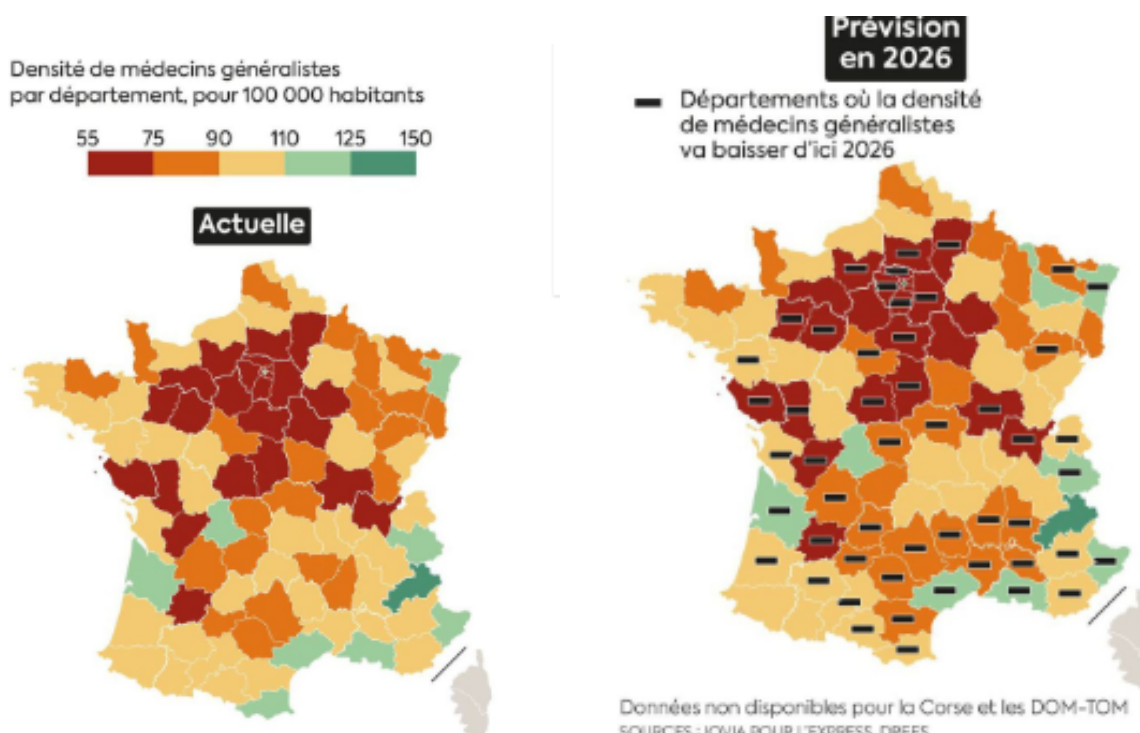


FIGURE 1.10 : Densité des médecins généralistes par département actuelle et projetée en 2026

Cette problématique a été abordée par les pouvoirs publics au travers des notions de zones. Le zonage est une approche de détermination de la meilleure répartition de l'offre médicale sur un territoire, employée par les agences régionales de santé. Elle a, dans un premier temps, été déterminée selon les dispositions prévues par l'article L1434-4 du Code de la santé publique en 2017, compte tenu de plusieurs indicateurs : l'accessibilité à un professionnel de santé ; les caractéristiques de la population considérée (âge, taux d'affection longue durée, nombre de bénéficiaires de la Complémentaire santé solidaire, etc.) ; le nombre des professionnels de santé, leur âge, leur activité, etc.

Cette nouvelle technologie s'appuie l'indicateur d'accessibilité aux soins : l'APL (Accessibilité Potentielle Localisée), développé par la direction de la Recherche, des Études, de l'Évaluation et des Statistiques (DREES) et l'Institut de recherche et documentation en économie de la santé (Irdes), conjointement. Cet indicateur synthétique permet de tenir compte à la fois de la proximité et de la disponibilité des médecins. Il combine donc les deux dimensions. Il intègre également une estimation de l'activité des médecins ainsi que des besoins en fonction de l'âge de la population locale et permet

des variantes tenant compte de l'âge des médecins ou encore de leur secteur d'activité. Cet indicateur peut être interprété de la façon suivante selon la DREES : « Un patient a dans son entourage plusieurs médecins accessibles (zone de recours, voir illustration). Plus le médecin est éloigné, plus son accessibilité est faible voire nulle au-delà d'un seuil. Chaque médecin accessible peut délivrer un certain nombre de consultations/visites aux patients susceptibles de le solliciter (patients de sa zone de patientèle). Pour la plupart des médecins, ce nombre est égal au nombre total d'actes que le médecin effectue dans l'année (prise en compte de l'activité réelle) divisé par le nombre de patients susceptibles de le solliciter. Ces derniers sont pondérés par leur âge (pour tenir compte des besoins de soins différenciés en fonction de l'âge) et par leur éloignement. Au final, pour connaître l'accessibilité d'un patient, on somme le nombre de consultations/visites que peuvent lui proposer les médecins accessibles. On obtient ainsi un nombre de consultations/visites accessibles pour chaque habitant. »

Une étude récente réalisée en 2018 par la DREES et parue en 2020 montre qu'en 2018 le manque d'accès aux médecins généralistes concernait 5,7% de Français, contre 3,8% en 2015. De plus, les territoires d'Outre-Mer (Guyane, Martinique et Guadeloupe) sont les plus concernés par cette sous-densité médicale. En métropole, ils sont suivis par les régions Île-de-France et Centre Val-de-Loire. A titre d'exemple, 14,6% de la population d'Île-de-France réside en zone sous-dense en termes d'accessibilité aux soins médicaux. Le tableau 1.11 présente le détail régional des communes sous-denses selon le seuil retenu.

L'augmentation des déserts médicaux s'explique d'un côté du fait que la demande de soins est en constante augmentation ; de l'autre, du fait que les départs à la retraite des médecins généralistes ne sont pas compensés, comme noté précédemment. De surcroît, certaines régions souffrent de leur faible attractivité : le nombre limité de commerces, d'écoles ou encore d'équipements sportifs ne poussent pas les jeunes médecins à s'installer. C'est pour lutter contre ces inégalités que le Ministère de la Santé a encouragé l'usage de la télémédecine. En octobre 2017, le Plan d'accès aux soins présenté par le Premier ministre et la ministre de la Santé, légifère un ensemble d'outils pour réformer les soins de proximité : au rang de ces outils figure justement la télémédecine. En effet, du fait de sa nature, elle évite au patient et/ou au professionnel de santé de se déplacer. Ainsi les facteurs d'éloignement et de temps de trajet vers le médecin ou encore de temps d'attente ne sont plus pertinents. La téléconsultation permet alors aux patients de réintégrer un parcours de soins n'ayant pas de médecin traitant et/ou de premier recours, en mettant à disposition du temps médical disponible. Le médecin qui téléconsulte rédige alors un compte-rendu qui est envoyé, sauf opposition de la part du patient, au médecin traitant lorsqu'il en existe un, afin d'assurer la continuité de l'information dans le parcours de soin habituel. On peut aussi rajouter que l'installation de cabines de téléconsultation dans les déserts médicaux favorise le développement du recours à cette nouvelle pratique médicale, et permet donc de réduire les inégalités.

*Source : SNIIR-AM 2015, CNAM-TS, populations municipales 2013, INSEE.

Région	Communes sous-denses au seuil de 2,5 consultations par an et		
	Population vivant dans une commune dont l'APL est inférieur au seuil	Part de la population régionale vivant dans une commune dont l'APL est inférieur au seuil	Part de la population vivant dans une commune dont l'APL est inférieur au seuil et résidant dans la région
Guadeloupe	76 867	19,1%	1,5%
Martinique	94 881	24,6%	1,8%
Guyane	70 219	28,8%	1,3%
La Réunion	7 226	0,9%	0,1%
Ile-de-France	1 188 170	9,9%	22,5%
Centre-Val-de-Loire	423 304	16,5%	8,0%
Bourgogne-Franche-Comté	356 216	12,6%	6,7%
Normandie	251 437	7,6%	4,8%
Hauts-de-France	203 097	3,4%	3,8%
Grand-Est	434 280	7,8%	8,2%
Pays-de-la-Loire	283 438	7,7%	5,4%
Bretagne	166 862	5,1%	3,2%
Nouvelle-Aquitaine	348 532	6,0%	6,6%
Occitanie	311 915	5,5%	5,9%
Auvergne-Rhône-Alpes	864 754	11,1%	16,3%
Provence-Alpes-Côte-d'Azur	161 459	3,3%	3,1%
Corse	48 801	15,2%	0,9%
Ensemble	5 291 458	8,1%	

FIGURE 1.11 : Répartition régionale des communes sous-denses en 2015 *

Lecture : en région Centre-Val-de-Loire, 423 304 personnes vivent dans une commune "sous-dense" au seuil de 2,5 consultations par an et par habitant, ce qui représente 16,5% de la population de la région. C'est ainsi 8,0 % de la population nationale vivant sous ce seuil qui réside en région Centre-Val-de-Loire.

Raisons économiques

L'Institut Economique Molinari, se basant sur la littérature récente sur la téléconsultation (et la téléexpertise médicale), affirme que le recours aux technologies de l'information et de la communication permettrait potentiellement de générer d'importantes économies sans compromettre la qualité de la prise en charge. Par exemple, une revue systématique publiée en 2015, recensant 35 études portant sur l'utilité-coût de la télémédecine observe que la celle-ci peut faire baisser les dépenses de santé pour certaines pathologies ou certaines pratiques comme par exemple la détection prénatale des maladies cardiaques congénitales, le suivi des patients en soins intensifs, l'ophtalmoscopie pour la détection de rétinopathie, ou encore les consultations pour les patients nécessitant un suivi pneumologique.

C'est un élément capital du processus de transformation du système de santé car celui-ci est, avec celui des retraites, le principal contributeur à la dette de la Sécurité sociale. En effet, cela est dû au fait que la progression des dépenses de santé dépasse celle de la richesse nationale, mais également à des déficits récurrents : 1,5 Md€ en 2019 et 30,4 Md€ en 2020 en raison de la crise sanitaire. Plus précisément, les revues observent que le recours à ces pratiques pourrait réduire les visites inutiles, désengorger les services d'urgence, réduire les dépenses de déplacement des patients et améliorer l'orientation des patients à travers leur parcours de soins (voir figure 1.12).

1. Baisse des dépenses de transport (téléconsultations)	293
2. Réduction des dépenses relatives aux consultations de généralistes et de spécialistes (téléconsultations + téléexpertises)	467
dont dermatologues, ophtalmologistes et ORL	47
dont cardiologues	50
dont endocrinologues et gastroentérologues	160
3. Economies générées dans les services d'urgences	284
dont réduction des visites (téléconsultations)	162
dont réduction des transferts de patients (téléexpertises)	122
Total (1+2+3)	1044

Source: Institut économique Molinari

FIGURE 1.12 : Synthèse des économies annuelles générées par la téléconsultation et la téléexpertise (millions d'euros par an)

En effet, l'Institut économique Molinari met en avant la baisse des dépenses de transport pour les patients optant pour la téléconsultation : 293 millions d'euros d'économie potentielle, pour l'année 2019. Cette revue apporte une estimation au niveau national de la distance moyenne qui sépare un habitant d'une pharmacie, d'un médecin généraliste ou d'un médecin spécialiste. Elle estime, à partir des données cartographiques les plus fines de l'Insee (200 mètres de côté), qu'en moyenne, un Français doit parcourir 2,45 kilomètres pour se rendre dans la pharmacie la plus proche, 6,25 kilomètres pour se rendre chez le généraliste le plus proche et enfin 19,85 kilomètres pour se rendre chez le spécialiste le plus proche.

Afin d'estimer l'impact de la téléconsultation sur les dépenses de déplacement des patients, deux cas de figures sont envisagés :

1. La téléconsultation a lieu au domicile du patient ;
2. La téléconsultation a lieu dans une pharmacie pour les assurés qui ne sont pas en mesure de suivre une téléconsultation en autonomie, soit par manque de compétences numériques, soit en raison de l'absence d'une connexion Internet performante.

En supposant que seulement 10% des consultations actuelles pourraient être effectuées efficacement en téléconsultation, une hypothèse forte et volontairement réductrice est effectuée. L'objectif est d'obtenir une estimation des économies minimales générées par la téléconsultation. Concernant les moyens de transport des patients, on suppose qu'ils sont identiques aux habitudes des actifs dans leurs déplacements domicile-travail, tels qu'estimés par l'Insee : pour les déplacements inférieurs à 5 kilomètres, 69,1% des déplacements effectués en voiture ou en moto, 18,8 % en transport en commun et le reste à pieds ou à vélo ; pour les déplacements supérieurs à 5 kilomètres, 82,2 % des trajets effectués en voiture ou moto, 16,3 % en transport en commun et le reste à pieds ou vélo.

Selon l'Open DAMIR, en 2019, 391 millions de consultations ont été effectuées, toutes spécialités confondues. Ainsi en prenant compte de la répartition entre consultations classiques et téléconsultations établie par l'Assurance Maladie, et en estimant les coûts de trajet selon les moyennes statistiques et en fonction du mode de transport, la revue montre que le recours à la téléconsultation aurait généré une économie de 325 millions d'euros si elles avaient eu lieu au domicile des patients et de 261 millions d'euros si les téléconsultations avaient eu lieu en pharmacies. La moyenne entre ces deux montants a alors été retenue. Des méthodes analogues sont alors utilisées afin d'obtenir 467 millions d'euros, correspondant au montant d'économies par an que pourrait être réalisé sur les dépenses relatives aux consultations généralistes et spécialisées, par l'utilisation de la téléconsultation et de la téléexpertise.

Concernant les dépenses des services d'urgence, la Cour des comptes estime le nombre annuel de visites aux urgences s'élève à 21 millions. De plus, le coût moyen d'une consultation simple aux urgences, de jour et hors week-end et jours fériés, s'élève à 115 euros pour l'hôpital. Ainsi, la dépense totale relative aux consultations aux urgences s'élève à 2,4 milliards d'euros. D'après Langabeer et al., la mise en place d'un système de téléconsultations pré-hospitalières destiné au filtrage des visites aux urgences et à la classification des patients selon la gravité de leur situation permettrait de réduire les consultations aux urgences de 6,7 %. Si l'on suppose que ces visites ne concernent que des consultations ne nécessitant pas d'hospitalisation et ayant lieu en journée, hors week-end et jours fériés, alors l'économie serait de 162 millions d'euros.

Renoncement aux soins

Les concepts de désert médical et de maîtrise des dépenses de santé sont intrinsèquement liés à celui du renoncement aux soins. Le renoncement aux soins est une notion qui vise à identifier les facteurs objectifs et subjectifs de non-recours à un soin ou à un praticien médical. Celle-ci est difficilement mesurable. En règle générale, on appréhende le renoncement aux soins au moyen d'enquêtes déclaratives.

Ainsi, en 2019, un sondage réalisé par l'Institut Opinionway pour le journal Les Échos et Harmonie Mutuelle révèle que 59 % des personnes interrogées affirmaient avoir renoncé à des soins médicaux au cours des douze mois précédant l'enquête, dont 29% pour des raisons financières. Parmi les répondants, les jeunes actifs de 25 à 34 ans étaient les plus nombreux (37%) à mentionner ce motif. Ce renoncement aux soins touchait premièrement les personnes disposant des revenus les plus modestes ou qui

connaissaient une baisse de leur pouvoir d'achat. Les soins auxquels les répondants renonçaient le plus étaient les soins dentaires (61%), le changement de montures (55%), les bilans de santé complets (34%) et enfin les consultations en cardiologie (22%).

Au-delà de cette étude, on peut trouver deux séries de facteurs qui expliquent le renoncement aux soins. Une première série est liée à l'environnement (absence de complémentaire santé, dépassements d'honoraires, augmentation de la participation financière des patients, organisation territoriale de l'offre...) et une seconde, aux individus (rapport personnel à la santé, possibilité de se tourner vers d'autres thérapies...).

Ainsi pour les raisons évoquées précédemment, on en conclut que la téléconsultation facilite et favorise le recours aux soins. Désormais utilisée quotidiennement par les praticiens, la téléconsultation est connue et appréciée par le grand public pour ses nombreux aspects pratiques.

1.2.4 Essor de la téléconsultation : boom durable ?

Il est légitime de se demander si la pratique de téléconsultation n'est efficace qu'en situation de crise sanitaire. En effet, le volume de téléconsultations a significativement baissé depuis le début du premier déconfinement, tout en restant à un niveau important. Depuis, la téléconsultation fluctue autour de 5 % du volume des consultations traditionnelles des médecins et dentistes libéraux, avec un pic à 8 % en novembre 2020 (Figure suivante).

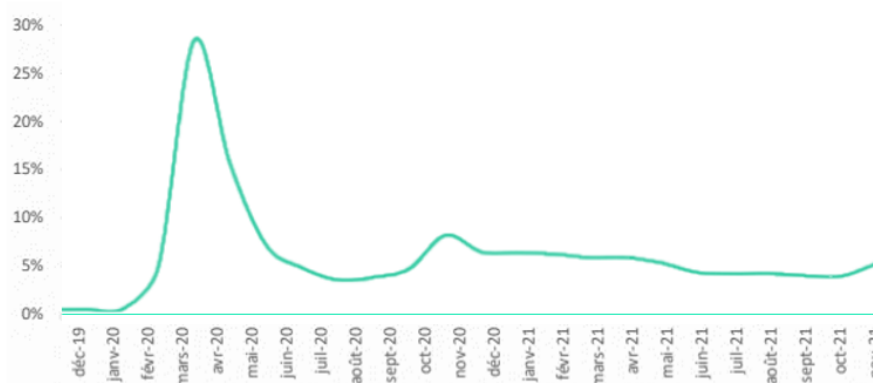


FIGURE 1.13 : Volume des téléconsultations comparé aux consultations, en face à face ou en ligne, des médecins libéraux

Cependant, dans ce chapitre on a observé toutes les applications possibles à partir de la téléconsultation, qu'elles soient techniques, sociales ou économiques. De plus, un sondage réalisé par ODOXA pour l'Agence du Numérique en Santé (ANS), publié en 2021, montre que la télé médecine en général, déjà plébiscitée en 2019, recueille une opinion positive avec 7 Français sur 10 qui s'en disent satisfaits. Selon cette même étude, 65% des Français accepteraient de recourir à une téléconsultation si leur médecin le leur proposait, contre 35% en novembre 2019. Par ailleurs, la téléconsultation est désormais adoptée en tant que pratique par les deux-tiers des médecins qui l'ont déjà pratiquée et 57% des médecins actuellement « réfractaires » pensent qu'ils l'utiliseront à l'avenir (10 pts de plus qu'en juin 2019).

Enfin, la régulation de la pratique effectuée par l'Assurance maladie à travers les différentes chartes des bonnes utilisations au cours du temps, souligne la reconnaissance de la téléconsultation en tant que prestation à part entière, qui s'installe au fur et à mesure dans le paysage médical français. Et c'est ainsi dans ce contexte que le mémoire a pour ambition de déterminer un modèle tarifaire de téléconsultation ainsi que d'étudier l'évolution de la consommation de d'actes de téléconsultations dans le temps, afin d'anticiper les éventuels avancements futurs.

Chapitre 2

Construction et analyses descriptives des bases de données

Ce chapitre a pour but de décrire le processus de manipulation des données utilisé pour parvenir aux bases nécessaires à la tarification de la consultation et de la téléconsultation ainsi qu'à l'étude de l'évolution temporelle de ces pratiques médicales.

Il s'articule en quatre parties :

- Construction de la base « sinistres » ;
- Construction de la base « assurés » ;
- Création des bases de modélisation finales ;
- Analyses statistiques des bases de données obtenues (afin d'établir les caractéristiques empiriques du risque).

2.1 Construction de la base « sinistres »

En actuariat, il est courant d'exploiter des jeux de données de taille relativement importante. Avec le développement des open data, les données sont de plus en plus disponibles à des coûts d'acquisition faibles et parfois inexistantes. Souvent, ce type de données soulève des défis en termes de gestion de volumétrie et de variété des données. A ces problèmes s'ajoute la complexité (vitesse d'exécution) des algorithmes à mettre en jeu pour collecter, stocker, traiter et analyser dans des délais courts et parfois en temps réel.

Partant de la base de données Open DAMIR, mise à disposition par la Sécurité sociale, pour construire la base "sinistres", il a été inévitable d'être confronté à ces problématiques.

2.1.1 Présentation de la base de données Open DAMIR

Depuis 20 ans, les divers progrès technologiques ont permis de stocker et de classer de grandes quantités de données. Les données de santé en font notamment partie. A cet effet, la base de données du Système

National d'Information Inter-Régimes de l'Assurance Maladie (SNIIRAM) a été créée par la loi de financement de la sécurité sociale (article L. 161-28-1). Ses objectifs principaux sont de contribuer :

- « A la connaissance des dépenses de l'ensemble des régimes d'assurance maladie par circonscription géographique, par nature des dépenses, par catégorie de professionnels responsables de ces dépenses et par professionnel ou établissement ;
- A la transmission en retour aux prestataires de soins d'informations pertinentes relatives à leur activité et leurs recettes, et s'il y a lieu à leurs prescriptions ;
- A la définition, à la mise en œuvre et à l'évaluation de politiques de santé publique (ce troisième objectif date de 2004). »

Il permet ainsi d'analyser des dépenses et des parcours de soins des assurés des régimes généraux.

Tandis que la Caisse Nationale de l'Assurance Maladie (CNAM) est responsable du recueil des données, du fonctionnement et du suivi de la base de données, la Commission Nationale de l'Informatique et des Libertés (CNIL) s'assure du respect des conditions d'accès en vertu du caractère confidentiel des données médicales. Ces données regroupent notamment :

- des fichiers des services médicaux permettant d'obtenir des informations sur les assurés,
- des répertoires de professionnels qui renseignent sur les prestataires de service médicaux,
- des feuilles de soins et des remboursements,
- des arrêts de travail et des indemnités journalières,
- des résumés de sortie hospitaliers transmis depuis 2007 par l'Agence Technique d'Information sur l'Hospitalisation (ATIH).

Néanmoins des informations personnelles, tels que les noms, prénoms, dates et lieux de naissance et numéros de Sécurité sociale des bénéficiaires et des prestataires, sont présentes dans ces fichiers. Celles-ci sont alors supprimées et remplacées par un numéro unique et anonyme correspondant à un numéro de chaînage issu d'une méthode d'anonymisation irréversible. Comme le précise le Rapport sur la gouvernance et l'utilisation des données de santé, publié en janvier 2014 par la DREES, l'anonymisation ne supprime pas le risque de ré-identification car l'intérêt de ce système d'information est sa précision et son exhaustivité. Concernant les données du SNIIRAM, ce rapport précise que les données individuelles qu'il contient sont « bien anonymes prises une par une, en ce sens qu'elles ne comportent pas l'identité des personnes, mais qu'elles ne peuvent pas être en accès libre parce qu'en croisant certaines informations qui y figurent, on peut identifier des personnes connues par ailleurs (des proches, des collègues ou des célébrités) ». Les différents niveaux d'accès aux données sont représentés par la figure 2.1.

*Source : ameli.fr

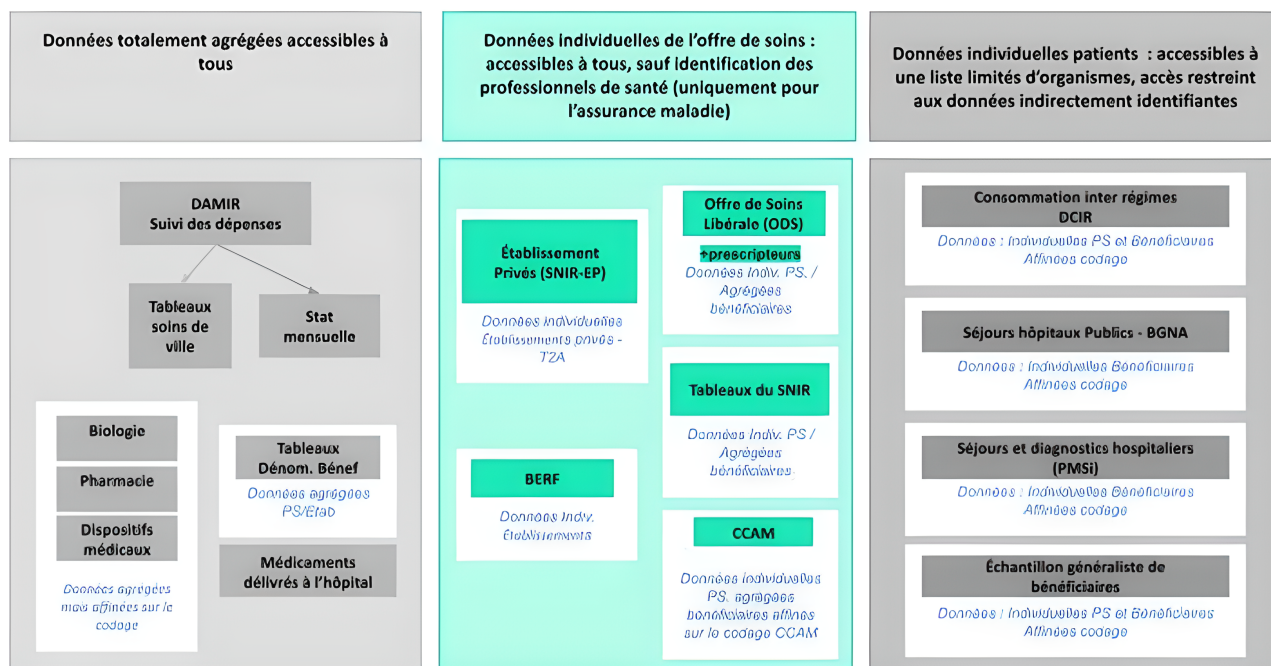


FIGURE 2.1 : Les trois niveaux d'accès des données du SNIIRAM *

La base de données Open DAMIR, appartenant aux données du SNIIRAM de niveau 1, sera exploitée dans le cadre de ce mémoire, afin de réaliser la base “sinistres”. L’Open DAMIR a pour but de décrire les dépenses de santé de la population française, mois après mois, depuis janvier 2009. Ses données concernent l’ensemble des prestations prises en charge par l’assurance maladie obligatoire (notamment les prestations de consultation et de téléconsultation) en fournissant des informations sur le patient, son prescripteur, le praticien de son soin et l’établissement de soin. Ces informations s’organisent en 55 variables et en 6 axes d’analyse détaillés dans la figure 2.2.

Il s’agit d’une source Open Data : l’accès y est public, directement sur le site de l’assurance maladie, et tous les fichiers mensuels peuvent se télécharger au format compressé .gz et peuvent être extraits au format .csv. Pour des raisons d’anonymat des patients mais aussi des professionnels de santé, les données sont proposées sous forme agrégée : une ligne d’un fichier est donc une somme des actes et des montants associés à un ensemble d’individus ayant une cinquantaine de caractéristiques communes (i.e. variables catégorielles telles que l’âge, le type d’acte, le sexe, etc.). Les données sont mises à jour tous les ans au mois de juin, et non tous les mois, ce qui explique que le mois de décembre 2020 soit le dernier mois disponible à date de l’analyse[†].

Par ailleurs, les axes géographiques sont limités par souci d’anonymat à 9 zones géographiques (regroupements de régions administratives) de 2009 à 2014 ; puis à 13 zones géographiques (proches des grandes régions administratives créées par la réforme territoriale de 2014) à partir de 2015.

[†]Les travaux ont commencés en Avril 2022

Axes d'analyse	Variables
Période	<ul style="list-style-type: none"> ❖ Année de soins ❖ Mois de soins
Bénéficiaires	<ul style="list-style-type: none"> ❖ Nature de prestation ❖ Nature d'assurance ❖ Nature de l'accident du travail ❖ Type d'enveloppe ❖ Complément d'acte ❖ Motif d'exonération du Ticket Modérateur ❖ Taux de remboursement ❖ Code secteur privé/public ❖ Type de prise en charge forfait journalier ❖ Indicateur TAA privé/public ❖ Code qualificatif parcours de soins (sortie) ❖ Nature du destinataire de règlement affiné ❖ Type de remboursement
Prescripteur	<ul style="list-style-type: none"> ❖ Catégorie du prescripteur ❖ Nature d'activité du prescripteur ❖ ZEAT du prescripteur ❖ Région du prescripteur ❖ Statut juridique du prescripteur ❖ ZEAT d'implantation de l'établissement ❖ Région d'implantation de l'établissement ❖ Catégorie de l'établissement
Exécutant	<ul style="list-style-type: none"> ❖ Catégorie de l'exécutant ❖ Spécialité médicale de l'exécutant ❖ Nature d'activité de l'exécutant ❖ ZEAT de l'exécutant ❖ Région de l'exécutant ❖ Statut juridique de l'exécutant ❖ Mode de fixation des tarifs de l'établissement ❖ ZEAT d'implantation de l'établissement ❖ Région d'implantation de l'établissement ❖ Catégorie de l'établissement ❖ Discipline de prestation de l'établissement ❖ Mode de Traitement de l'établissement
Dépenses	<ul style="list-style-type: none"> ❖ Coefficient global de la prestation préfiltré ❖ Dénombrement de la prestation préfiltré ❖ Quantité de la prestation préfiltrée ❖ Montant du dépassement de la prestation préfiltré ❖ Montant de la dépense de la prestation préfiltrée ❖ Montant versé/remboursé préfiltré ❖ Coefficient global ❖ Dénombrement ❖ Quantité ❖ Montant du dépassement ❖ Montant de la dépense ❖ Montant versé/remboursé ❖ Base de remboursement

FIGURE 2.2 : Récapitulatif des variables disponibles de l'Open DAMIR

2.1.2 Présentation des traitements réalisés

Étape 1 : Choix des variables et des modalités nécessaires à l'étude

Afin de sélectionner les variables pertinentes à l'étude des actes de consultation et de téléconsultation, le descriptif Excel des variables de l'Open DAMIR, intitulé "Lexique Open DAMIR" a été exploité. Le choix s'est porté sur les modalités et les variables détaillées dans la figure 2.3.

	Variables	Libellés	Modalités (Variables catégorielles)
Période de modélisation	FLX_ANN_MOI	Année et mois de remboursement	❖ 201807 à 202012
Profil d'assurés	BEN_SEX_COD	Sexe	❖ 1 : masculin ❖ 2 : féminin
	AGE_BEN_SNDS	Age	❖ 0 : 0 à 19 ans ❖ 20 : 20 à 29 ans ❖ 30 : 30 à 39 ans ❖ 40 : 40 à 49 ans ❖ 50 : 50 à 59 ans ❖ 60 : 60 à 69 ans ❖ 70 : 70 à 79 ans ❖ 80 : 80 ans et plus
	BEN_RES_REG	Région de résidence	❖ 5 : Régions et Départements d'outre-mer ❖ 11 : Ile-de-France ❖ 24 : Centre-Val de Loire ❖ 27 : Bourgogne-Franche-Comté ❖ 28 : Normandie ❖ 32 : Hauts-de-France - Nord-Pas-de-Calais-Picardie ❖ 44 : Grand Est ❖ 52 : Pays de la Loire ❖ 53 : Bretagne ❖ 75 : Aquitaine-Limousin-Poitou-Charentes ❖ 76 : Languedoc-Roussillon-Midi-Pyrénées ❖ 84 : Auvergne-Rhône-Alpes ❖ 93 : Provence-Alpes-Côte d'Azur et Corse
	BEN_CMU_TOP	Affiliation CMU	❖ 0 : non CMU ❖ 1 : CMU
Consommation de sinistres	PRS_NAT	Nature de la prestation	(Voir figure suivante)
	PRS_PAI_MNT	Coût total de la prestation (frais réels)	
	PRS_ACT_QTE	Quantité d'actes concernés	

FIGURE 2.3 : Variables choisies pour la base « sinistres »

La période de modélisation choisie s'étend de septembre 2018 à décembre 2020, car les données relatives à la pratique de téléconsultation ne sont enregistrées qu'à partir de cette date. De plus, il est important de noter que ces mois correspondent à des mois de règlement du remboursement, et ne correspondent donc pas forcément aux mois de délivrance de la prestation. Une variable spécifique identifie clairement ces derniers.

A titre d'exemple, un acte de téléconsultation réalisé par un médecin généraliste ayant été effectué au cours du mois de novembre 2018, pourra être réglé en mars 2019.

D'autre part, les modalités de la variable « PRS_NAT » - décrivant la nature de la prestation au sein de l'Open DAMIR - ont été choisies selon un référentiel d'équivalence entre les prestations de consultation et de téléconsultation, que l'on peut retrouver dans l'annexe 1.

Il a semblé nécessaire de construire un tel référentiel afin de comparer les évolutions de consom-

mation entre les deux pratiques médicales.

Ce référentiel a été établi à partir des correspondances entre les descriptifs de chaque acte de consultation et de téléconsultation existant dans la base de l'Open DAMIR.

Ces correspondances nous ont permis de distinguer trois catégories de prestations de consultation :

- Catégorie 1 : actes de consultation dont l'équivalence est connue ;
- Catégorie 2 : actes de consultation sans équivalence ;
- Catégorie 3 : actes de consultation dont l'équivalence est inconnue.

Ainsi ont été exclues de l'étude les prestations de consultation n'ayant pas d'équivalent ou dont l'équivalent est inconnu.

La figure 2.4 présente les actes de téléconsultation et de consultation retenus lors de l'étude.

Téléconsultation	Consultation
1191 : Téléconsultation toutes spécialités	<ul style="list-style-type: none"> ❖ 1112 : Consultation cotée CS ❖ 1113 : Consultation cotée CNP ❖ 1117 : Consultation des spécialistes cotée C2 ❖ 1118 : Consultation des psychiatres cotée C2,5 ❖ 1132 : Majoration consultation endocrinologie ❖ 1148 : Rémunération forfaitaire par consultation pour le suivi des personnes âgées ❖ 1922 : Consultation ❖ 1933 : Majoration consultation régulée de nuit ❖ 1934 : Majoration consultation régulée milieu de nuit
1192 : Téléconsultation généraliste	<ul style="list-style-type: none"> ❖ 1109 : Consultation spécialiste médecine générale ❖ 1110 : Consultation médecine générale ❖ 1111 : Consultation cotée c
3103 : Téléconsultation lors d'un soin	<ul style="list-style-type: none"> ❖ 1098 : Consultation CCMU 3
3104 : Téléconsultation dans un lieu dédié	<ul style="list-style-type: none"> ❖ 1954 : Participation assuré consultations et soins externes (C2S) ❖ 1955 : Participation assuré consultations et soins externes (régime local) ❖ 2217 : Forfait consultation urgence hôpital ❖ 2336 : Forfait pour consultation en centre médico-psycho pédagogique
3105 : Téléconsultation à domicile	<ul style="list-style-type: none"> ❖ 1153 : Majoration consultation pour insuffisance cardiaque après hospitalisation ❖ 1154 : Majoration consultation suivi après hospitalisation patients à forte comorbidité

FIGURE 2.4 : Équivalence entre les actes de téléconsultation et de consultation retenus

Ainsi, afin de comparer les évolutions de consommation de la consultation et téléconsultation généraliste, il suffira de réaliser des moyennes pondérées (pondération sur les quantités d'actes consommés selon la nature de la prestation) de la fréquence et du coût moyen des différentes prestations de consultation équivalentes à la prestation de téléconsultation généraliste.

Étape 2 : Extraction des variables et des modalités nécessaires à l'étude

En vue d'extraire les variables et les modalités choisies, le langage R, développé par R CORE TEAM (2022), a été privilégié. Toutefois, au cours de cette étape des problématiques liées à la volumétrie ont été rencontrées.

En effet, à partir de septembre 2018, un fichier représentant les dépenses mensuelles de santé de la population française possède en moyenne un poids de 5,105 Go (Gigaoctets). Or, la période de modélisation s'étend de septembre 2018 à décembre 2020, ainsi le poids total des données à traiter est de 142,94 Go, ce qui correspond à environ 72 500 000 observations ($\sim 70\,000\,000$ lignes sur la consultation et $\sim 2\,500\,000$ lignes sur la téléconsultation). Rapidement, il a été observé qu'il aurait été coûteux en temps et en mémoire (sachant que l'espace mémoire sur Rstudio est limité) de charger ces fichiers un par un, n'ayant pas la possibilité de stocker les données dans un serveur local.

Ainsi la solution trouvée afin d'automatiser le chargement des données dans Rstudio a été de déterminer un module d'importation comprenant un filtre qui sélectionne les variables et les modalités pour la variable « PRS_NAT » choisies, en amont du chargement. Ainsi, en partant de deux bases vides de type data.table (l'une pour les actes de téléconsultation et l'autre pour ceux de consultation), le module permet à chaque fichier mensuel de l'Open DAMIR, d'être chargé dans ces bases en étant déjà filtré. De plus, à l'aide des fonctions `gc()` et `rm()` cela est fait sans alourdir l'espace mémoire de Rstudio car elles permettent respectivement de libérer de l'espace mémoire à chaque étape de la boucle et de supprimer le fichier de l'espace mémoire une fois enregistré dans la base.

Étape 3 : Évaluation de la qualité des données

L'étape suivante a été de tester la qualité des données de consommation sur la téléconsultation et la consultation ainsi extraites de l'Open DAMIR. La revue de celles-ci a mis en évidence la présence de données aberrantes ainsi que de données manquantes. Il a donc été nécessaire de traiter ces deux problématiques afin d'obtenir une base sans anomalies retraçant la consommation de consultations et de téléconsultations de 2018 à 2020, agrégée selon les caractéristiques des bénéficiaires, les actes médicaux et la période de remboursement des soins.

Traitement des données aberrantes

Ces données correspondent aux observations (lignes de la matrice des données) de la base possédant des coûts et/ou des quantités négatives ou nulles. Il s'agit en réalité d'opérations de régularisation des remboursements de la Sécurité sociale aux assurés, qui ne correspondent donc pas à des consommations effectives. De plus, cette anomalie ne concerne qu'un faible pourcentage de l'ensemble des lignes de la base ($\sim 1,2\%$) et $\sim 0,8\%$ des prestations, en terme de dépenses. Ainsi ces lignes ont été supprimées.

Traitement des données manquantes

On dit qu'une observation comporte une donnée manquante s'il existe une ou plusieurs modalités pour lesquelles la valeur n'est pas renseignée. La répartition de ces anomalies selon les variables et le nombre de lignes concernées est illustrée dans la figure suivante.

Nombres de lignes concernées	Age	Région	Affiliation CMU	Nombre de données manquantes par ligne
1 040 726				0
1 285 469				1
50 846				1
75 481				2
438				1
975				2
1				2
17				3
Total données manquantes pour chaque variables	1 431	126 345	1 361 942	1 489 718

FIGURE 2.5 : Répartition des valeurs manquantes - Actes de téléconsultation

Lecture : sur 1 040 726 lignes, il n'y a aucune donnée manquante. Sur 75 481 lignes, les variables CMU et région ne sont pas précisées à la fois. Au total, la variable CMU n'est pas mentionnée pour 1 361 942.

Ces irrégularités concernent les variables catégorielles suivantes : âge, région de résidence et affiliation à la CMU. Cette dernière est celle avec le plus de données manquantes (elle est à l'origine de 1361942 valeurs manquantes, soit $\sim 90\%$ de l'ensemble des valeurs manquantes) suivi de la variable région puis de la variable âge. Toutefois, il a été décidé de ne pas supprimer la variable CMU, afin d'obtenir une segmentation supplémentaire de la population d'assurés.

Plusieurs techniques existent afin de traiter les données manquantes : il peut être envisagé de les supprimer de la base de données ou encore les remplacer par des valeurs artificielles (on parle d'imputation). Afin de limiter la perte d'informations et de consistance du jeu de données, il a été décidé de remplacer les valeurs manquantes de chaque modalité par le mode de cette variable explicative.

En statistique, le mode, ou valeur dominante, est la valeur la plus représentée d'une variable quelconque dans une population donnée. Une répartition peut être unimodale ou plurimodale (bimodale, trimodale...), si deux ou plusieurs valeurs de la variable considérée émergent également, voire sans aucun mode (distribution uniforme) si toutes les valeurs de la variable considérée émergent également. Dans le cas d'une répartition en classes d'amplitudes égales, la classe modale désigne celle qui a le plus fort effectif. La convention est d'appeler mode le centre de la classe modale. Si les classes sont d'amplitudes diverses, il convient de relativiser pour désigner ce paramètre. La classe modale est alors celle qui a la plus forte densité.

Afin de s'assurer de la stabilité de la méthode utilisée, il a semblé utile de comparer les distributions du coût total de la prestation (« PRS_PALMNT ») et de la quantité d'actes consommés (« PRS_ACT_QTE »), avant et après reconstitution des données. Ces variables constituent un enjeu majeur dans les travaux car elles renseignent sur la consommation des actes de téléconsultation et de consultation.

Compte tenu de des résultats observés sur les figures 2.6 et 2.7, il peut être conclu que le principe d'imputation par mode s'adapte bien aux données.

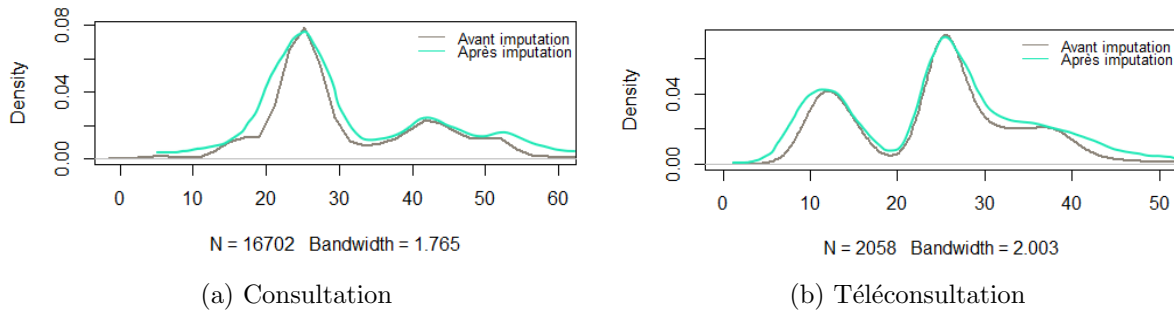


FIGURE 2.6 : Comparaison des distributions de coût moyen avant et après imputation

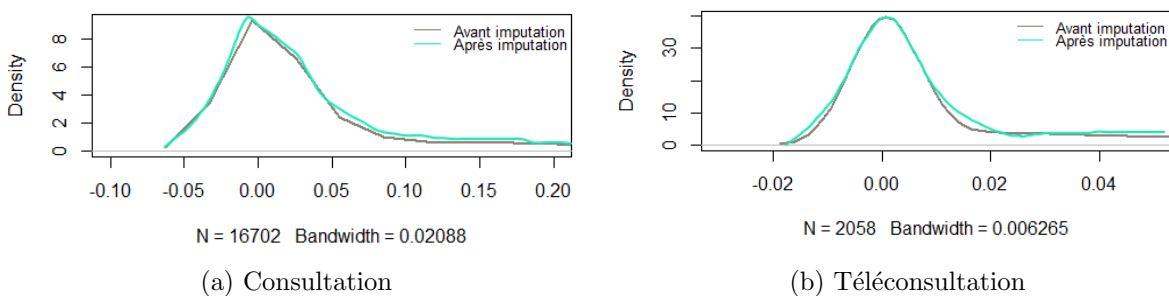


FIGURE 2.7 : Comparaison des distributions de fréquence avant et après imputation

Étape 4 : Création de la variable « ANNÉE »

Enfin, la dernière étape a été de rajouter une variable « ANNÉE » ne représentant que les années de remboursement des soins à la différence de la variable « FLX_ANN_MOI » qui représente l'année et le mois de remboursement des soins. Cela permet non seulement de joindre cette base à la base assurés qui est annuelle (période de modélisation allant de 2018 à 2020) mais aussi d'agréger des groupes d'observations du jeu de données en des observations uniques, selon l'année et le profil d'assurés, en sommant sur les montants et les quantités des sinistres.

2.2 Construction de la base « assurés »

Bien que le caractère agrégé de l'Open DAMIR permette l'anonymisation des données, il soulève un point d'attention quant à la détermination de la fréquence et de sa loi en présence d'informations uniformisées. En effet, l'Open DAMIR fournit les quantités par acte mais elle ne renseigne pas le nombre de bénéficiaires par ligne. Or, le nombre de bénéficiaires, représentant l'exposition au risque, est une donnée nécessaire dans le calcul de la fréquence par l'approche fréquence x coût moyen (l'approche priorisée de tarification de ces travaux).

Cette donnée est d'autant plus importante que l'étude de l'évolution temporelle de la consultation et de la téléconsultation se fait à l'aide de séries temporelles modélisant les fréquences de survenance de ces pratiques.

C'est alors à ce moment qu'interviennent les données démographiques de l'INSEE.

2.2.1 Présentation des données brutes

Données démographiques de l'INSEE

Chaque année, l'institut estime la population des régions et des départements (France métropolitaine et DOM) à la date du 1er janvier. Ces estimations annuelles de population sont déclinées par sexe et par âge quinquennal (avec un pas de 5 ans). La base de données retenue a été celle intitulée « Estimation de population par région, sexe et âge quinquennal - Années 1975 à 2022 ». La méthodologie utilisée par l'INSEE pour estimer le nombre d'habitants par région, par sexe et par âge est explicitée dans la section « Documentation » du site d'extraction de la base.

Dans la mesure où l'Open DAMIR détaille l'ensemble des soins prodigués pour la population française, il est possible de joindre cette base avec des données de démographie nationale. Ces données forment ainsi la base « assurés » de la modélisation.

Données externes : variables économiques et médicales

Afin d'apporter des éléments de contexte du paysage médical et économique français aux analyses, il a été décidé de rajouter des données médicales et économiques à la base « assurés ». Les variables ont été choisies pour leur potentielle utilité et l'information qu'elles apportent a priori sur la consommation de consultation et de téléconsultation, mais aussi pour l'accessibilité de ces prestations médicales. Toutes les données externes seront ajoutées à la base « assurés » à la maille régionale, afin de s'aligner avec la granularité des données géographiques de l'Open DAMIR. Les variables retenues sont détaillées dans la figure 2.7.

Contexte médical	<ul style="list-style-type: none"> ❖ Nombre de médecins ❖ Densité médicale ❖ Densité de médecins généralistes ❖ Densité de médecins spécialistes ❖ Nombre de lits d'hôpital ❖ Nombre d'hôpitaux
Contexte économique	<ul style="list-style-type: none"> ❖ Taux de pauvreté ❖ Taux de chômage ❖ Nombre de personnes au RSA

FIGURE 2.8 : Variables médicales et économiques

D'une part, les variables économiques proviennent de la section "Pauvreté - Précarité - Exclusion" du panorama statistique « Cohésion sociale, Travail, Emploi » publié chaque année par la Direction de la recherche, des études, de l'évaluation et des statistiques (DREES). Les descriptifs sont détaillés sur le site officiel du panorama :

- « Le taux de pauvreté correspond à la proportion d'individus dont le niveau de vie est inférieur pour une année donnée à un seuil, dénommé seuil de pauvreté (exprimé en euros). Dans l'approche en termes relatifs, le seuil de pauvreté est déterminé par rapport à la distribution des niveaux de vie de l'ensemble de la population. On privilégie en Europe le seuil de 60% du niveau de vie médian ;
- Le taux de chômage est le pourcentage de chômeurs dans la population active (actifs occupés + chômeurs). Ces taux sont calculés en moyenne trimestrielle pour les régions, les départements et

les zones d'emploi de métropole selon une méthodologie synthétisant les informations de l'enquête emploi pour le niveau national et de Pôle emploi pour la répartition géographique.

- Le taux de RSA est le pourcentage d'individus bénéficiant du revenu de solidarité active (RSA). Celui-ci s'adresse depuis le 1er juin 2009 en France métropolitaine et depuis le 1er janvier 2011 dans les départements d'outre-mer (et le 1er janvier 2012 à Mayotte) à toute personne âgée d'au moins 25 ans ou assumant la charge d'au moins un enfant né ou à naître. Depuis le 1er septembre 2010 en France métropolitaine et le 1er juillet 2011 dans les DOM, il a été étendu aux personnes de moins de 25 ans ayant travaillé au moins deux années au cours des trois dernières années (RSA socle jeune). Le seuil du revenu garanti par le RSA est calculé en additionnant un montant forfaitaire, variable selon la composition du foyer, à une fraction (62%) des revenus professionnels des membres du foyer ».

D'autre part, les données sur les effectifs des médecins et les densités médicales proviennent de l'article intitulé « Atlas de la démographie médicale », publié chaque année par le Conseil National de l'Ordre des Médecins (CNOM). Ce document a pour ambition de rendre compte de la situation annuelle de la démographie médicale en France, en explorant différents sujets tels que la répartition des effectifs par âge, sexe, département, actifs ou retraités, groupe de spécialités, etc.

Concernant les données sur la densité du nombre de lits d'hôpitaux par région, elles proviennent du panorama de la DREES sur les établissements de santé qui contient la densité de places d'hospitalisation partielle par département de 2018 à 2020. Une agrégation par région a donc été réalisée.

Enfin, le nombre d'hôpitaux par région est disponible dans la section « Statistiques » du site officiel de la Fédération Hospitalière de France et inclut dans le décompte les établissements suivants :

- Centres hospitaliers régionaux universitaires (CHRU),
- Centres hospitaliers régionaux (CHR),
- Hôpital dépendant d'un CHRU (CHU),
- Centres hospitaliers (CH),
- Centres hospitaliers spécialisés (CHS),
- Groupements de Coopération spécialisée (GCS).

Ces derniers servent de cadre non seulement à des coopérations entre les établissements de santé publics et privés, mais également entre la ville et l'hôpital (centres de santé, maisons de santé et professionnels médicaux libéraux à titre individuel ou collectif, ou encore acteurs du secteur médico-social, etc.). Les GCS doivent inclure au minimum un établissement de santé, quelle que soit sa nature juridique et peuvent être ouverts à tout acteur concourant à la réalisation de leur objet.

2.2.2 Présentation des retraitements effectués

Comme susmentionné, la construction d'une base "assurés" est une étape fondamentale pour la tarification et pour la modélisation de la fréquence à l'aide de séries temporelles. Il est conjecturé que les profils d'assurés pris en compte par l'Open DAMIR représentent l'ensemble de la population française, supposée couverte par la Sécurité sociale. Ainsi, partant des données démographiques de l'INSEE, la base « assurés » a pour objectif de détailler l'exposition au risque décrit dans l'Open DAMIR.

Plusieurs traitements ont été nécessaires pour y parvenir.

Tout d'abord, il a été remarqué que les modalités des variables âge et région des données de l'INSEE ne correspondent pas à celles des variables correspondantes de l'Open DAMIR. Ainsi, afin de mener à bien la jointure entre les bases « sinistres » et « assurés », quelques retraitements ont été réalisés et ils sont illustrés ci-après :

- Création de nouvelles classes de régions :

Sécurité sociale	INSEE
Régions et Départements d'outre-mer	❖ Guadeloupe ❖ Martinique ❖ Guyane ❖ La Réunion ❖ Mayotte
Ile-de-France	❖ Ile-de-France
Centre-Val de Loire	❖ Centre-Val de Loire
Bourgogne-Franche-Comté	❖ Bourgogne-Franche-Comté
Normandie	❖ Normandie
Hauts-de-France - Nord-Pas-de-Calais-Picardie	❖ Hauts-de-France
Grand Est	❖ Grand Est
Pays de la Loire	❖ Pays de la Loire
Bretagne	❖ Bretagne
Aquitaine-Limousin-Poitou-Charentes	❖ Nouvelle-Aquitaine
Languedoc-Roussillon-Midi-Pyrénées	❖ Occitanie
Auvergne-Rhône-Alpes	❖ Auvergne-Rhône-Alpes
Provence-Alpes-Côte d'Azur et Corse	❖ Provence-Alpes-Côte d'Azur ❖ Corse

FIGURE 2.9 : Regroupement de régions selon la segmentation de la Sécurité sociale

- Création de nouvelles classes d'âges :

Sécurité sociale	INSEE
0-19 ans (Modalité 0)	❖ 0 à 4 ans ❖ 5 à 9 ans ❖ 10 à 14 ans ❖ 15 à 19 ans
20 – 29 ans (Modalité 20)	❖ 20 à 24 ans ❖ 25 à 29 ans
30 – 39 ans (Modalité 30)	❖ 30 à 34 ans ❖ 35 à 39 ans
40 – 49 ans (Modalité 40)	❖ 40 à 44 ans ❖ 45 à 49 ans
50 – 59 ans (Modalité 50)	❖ 50 à 54 ans ❖ 55 à 59 ans
60 – 69 ans (Modalité 60)	❖ 60 à 64 ans ❖ 65 à 69 ans
70 – 79 ans (Modalité 70)	❖ 70 à 74 ans ❖ 75 à 79 ans
80 ans et plus (Modalité 80)	❖ 80 à 84 ans ❖ 85 à 89 ans ❖ 90 à 94 ans ❖ 95 ans et plus

FIGURE 2.10 : Regroupement de classes d'âges selon la segmentation de la Sécurité sociale

Puisque la variable « BEN_TOP_CMU » représentant le critère d'affiliation à la CMU fait partie du profil d'assurés dans la base « sinistres », l'étape suivante a été de rajouter une variable représentant ce critère dans la base « assurés ». A cet effet, la répartition annuelle par région de la population française ayant une complémentaire CMU, provenant de la section « 10 - Pauvreté - Précarité - Exclusion » du panorama statistique « Cohésion sociale, Travail, Emploi » publié chaque année par la DREES, a été utilisée. Selon les données INSEE, le nombre total d'assurés à la CMU s'élevait à 5 176 157 d'habitants en 2019, soit 8% de la population totale. Faute de données trouvées sur les répartitions par âge et sexe des assurés CMU pour les années allant de 2018 à 2020, il a été procédé à une distribution uniforme de ratios régionaux, afin de ventiler la population CMU par tranche d'âge et par sexe.

La dernière étape a consisté à joindre ces données démographiques aux données médicales et économiques par le biais de la variable région. Pour les modalités de la variable région de l'Open DAMIR qui sont composées de plusieurs régions (modalité 93 = Provence-Alpes-Côte d'Azur/ Corse et modalité 5 = Régions et Départements d'outre mer), il a été procédé à des moyennes pondérées sur la population des valeurs économiques et médicales, afin de les regrouper.

Ainsi, la base de modélisation obtenue représente la population française de 2018 à 2020, ventilée selon 416 profils correspondant aux croisements des modalités des variables caractérisant les bénéficiaires de la base « sinistres ». De plus, la couverture de la Sécurité sociale étant permanente, l'exposition annuelle considérée pour un profil d'individus donné correspondra au nombre d'individus faisant partie de ce profil. Ainsi, la base « assurés » finale contient les variables suivantes :

- Sexe,
- Âge,

- Région de résidence,
- Affiliation CMU,
- Exposition ,
- Variables économiques et médicales réparties par région de résidence.

2.3 Création des bases de modélisation finales

Dans le cadre de l'étude, il a été considéré pertinent de créer quatre bases de données :

- Deux bases de données agrégées annuellement par profil d'assurés (en sommant sur les coûts et les quantités), l'une pour la consultation et l'autre pour la téléconsultation, permettant de modéliser la tarification annuelle par profil d'assurés.

Il est alors remarqué que l'agrégation des données introduira une perte d'information sur la volatilité individuelle des coûts des sinistres. Concernant les quantités (et donc de la fréquence) des sinistres, il n'y aura pas de perte d'information car par principe d'anonymisation de l'Open DAMIR, seulement les quantités par profil d'assurés sont connues (et non pas celles individuelles).

Ces deux bases sont obtenues par jointure de la base « sinistres » et la base « assurés » pour les prestations de consultation et de téléconsultation, distinctement, avec comme clés de jointure l'année de remboursement ainsi que les variables constituant les profils d'assurés.

Il est important de noter qu'à chaque étape des jointures (base économique/médicale & base INSEE ; base « assurés » & base « sinistres ») des contrôles ont été effectués, afin d'obtenir un jeu de données cohérent et sans anomalies. Il s'agissait par exemple de s'assurer de la stabilité des coûts et des quantités de sinistres avant et après jointure par profil d'assurés et type de prestation. De plus, il était important de vérifier les totaux des expositions des différents profils d'assurés, avant et après jointure.

Ainsi, pour chaque année, il est possible de calculer les coûts moyens et les fréquences des sinistres (la méthodologie est détaillée dans le chapitre 3).

- Deux bases de données mensuelles incluant les variables : mois et année de remboursement, fréquence de recours à la téléconsultation/consultation.

Celles-ci seront utilisées afin de réaliser la modélisation des évolutions au cours du temps de la fréquence de consommation de consultations et de téléconsultations, à l'aide de séries temporelles. Cette fréquence a ainsi été obtenue en joignant la variable exposition (à partir de la base « assurés », avec comme clés de jointure l'année de remboursement et les variables caractérisant les profils d'assurés) à la base « sinistres » agrégée mensuellement par profil d'assurés, en sommant sur les quantités.

L'agrégation mensuelle a été préférée à celle annuelle car elle permettra de mettre en évidence une éventuelle saisonnalité des données.

2.4 Analyses statistiques des bases de données obtenues

L'analyse de données est une étape importante avant toute modélisation. Elle permet de comprendre les données en mesurant empiriquement l'impact de chaque variable tarifaire sur la consommation de téléconsultations comparée à celle de consultations. De plus, ces analyses serviront de référentiel d'analyse des coefficients tarifaires obtenus à l'aide de modèles linéaires généralisés (GLM).

Dans l'ensemble de la partie, certaines abréviations seront utilisées. Elles sont détaillées ci-après :

- COUT_MOY_C et COUT_MOY_TC représentent respectivement le coût moyen de consultation et le coût moyen de téléconsultation ;
- FREQ_MOY_C et FREQ_MOY_TC caractérisent les fréquences de survenance de consultation et de téléconsultation.

2.4.1 Analyses univariées de coût moyen et de fréquence

Influence du sexe

Bien que réglementairement le sexe ne puisse être retenu comme variable tarifaire (*Gender Directive*), l'analyse de son influence sur la sinistralité permet une meilleure appréhension des risques consultation et téléconsultation. Ainsi, qu'il s'agisse de consultation ou de téléconsultation, les femmes affichent en moyenne une fréquence de consommation 25% plus élevée que celles des hommes.

En revanche, il existe une nette disparité de fréquence entre les prestations de consultation et de téléconsultation, les premières étant plus répandues, quel que soit le sexe pris en compte. A cet effet, la fréquence de consultation est ~ 6 fois plus élevée et ~ 7 fois plus élevée que celle de téléconsultation, respectivement pour les femmes et pour les hommes.

Enfin, les niveaux de dépense restent homogènes pour les deux sexes. Les coûts avoisinent 27,70 euros pour les consultations et 28,25 euros pour les téléconsultations.

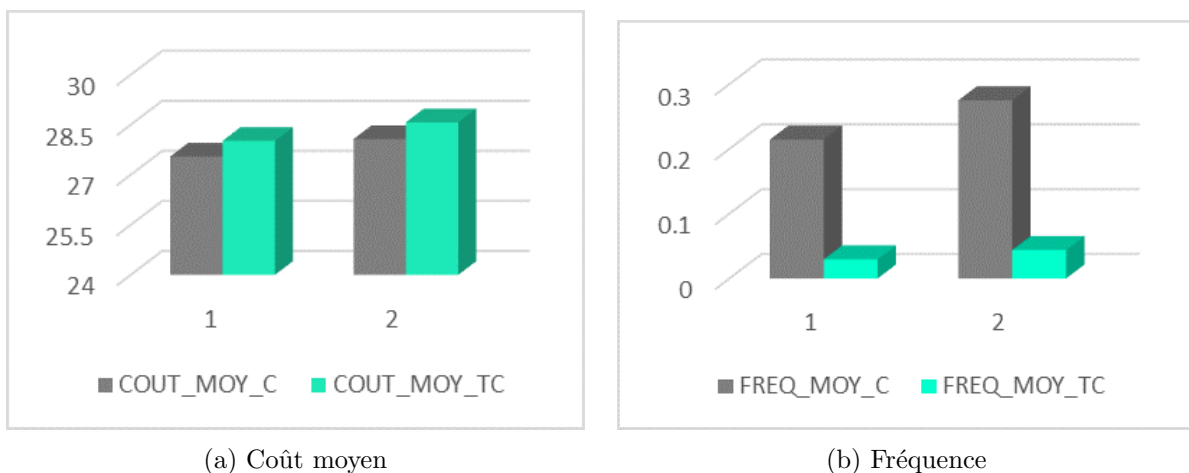


FIGURE 2.11 : Influence de la variable sexe sur la consommation de sinistres

Influence de l'âge des assurés

Au niveau des dépenses, jusqu'à 59 ans, les téléconsultations sont en moyenne 20% plus chères que les consultations. A partir de 60 ans, cette tendance s'inverse.

Concernant la fréquence de téléconsultation, elle est nettement inférieure à la fréquence de consultation à tous âges. Cela est certainement lié à l'arrivée plus récente de la pratique de téléconsultation dans le paysage médical français. De plus, celle-ci se révèle assez homogène même si les personnes âgées de 30 à 39 ans et de 80 ans et plus enregistrent la fréquence la plus élevée ($\sim 0,05$ téléconsultations par assuré), contrairement aux plus jeunes qui enregistrent la fréquence la plus faible ($\sim 0,02$ téléconsultations par assuré).

Ces constats sont cohérents avec la volonté de l'État de développer l'usage de la téléconsultation dans les EHPAD à travers la création d'une tarification spécifique, calibrée selon le type de séjour en EHPAD (tarif global ou partiel) et les différentes spécialités médicales.

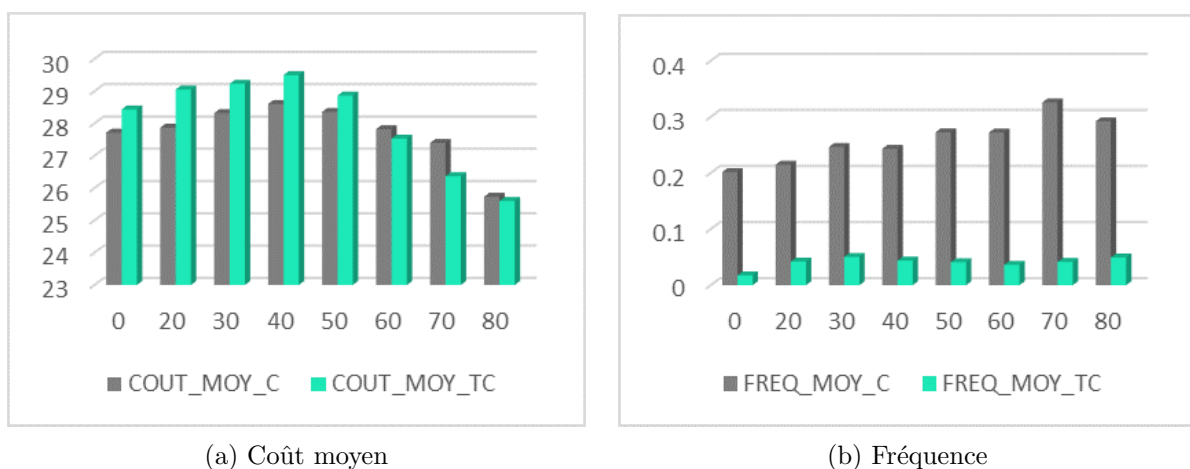


FIGURE 2.12 : Influence de la variable âge sur la consommation de sinistres

Influence de l'affiliation à la CMU

Dans l'ensemble, les coûts moyens de téléconsultation et de consultation sont plus élevés chez les assurés non affiliés à la CMU. Chez les assurés non CMU, l'écart de coût est marginal tandis que pour les affiliés à la CMU les consultations sont 7% moins coûteuses que les téléconsultations ($\sim 25,45$ euros contre $\sim 27,40$ euros).

Chez les assurés affiliés à la CMU comme chez les assurés sans CMU, la fréquence de téléconsultation est bien inférieure à la fréquence de consultation.

Néanmoins l'écart est plus prononcé chez les bénéficiaires de la CMU, qui consomment en moyenne trois fois plus de consultations et deux fois plus de téléconsultations. Une raison plausible qui expliquerait ce phénomène est le fait que les consultations et les téléconsultations (pratiqués par les médecins du secteur 1) sont à 100% prises en charge par la Sécurité sociale pour les assurés CMU. De plus, ils bénéficient du tiers payant, ainsi l'avance des frais sur la part des soins remboursés par l'Assurance maladie ne leur est pas demandée.

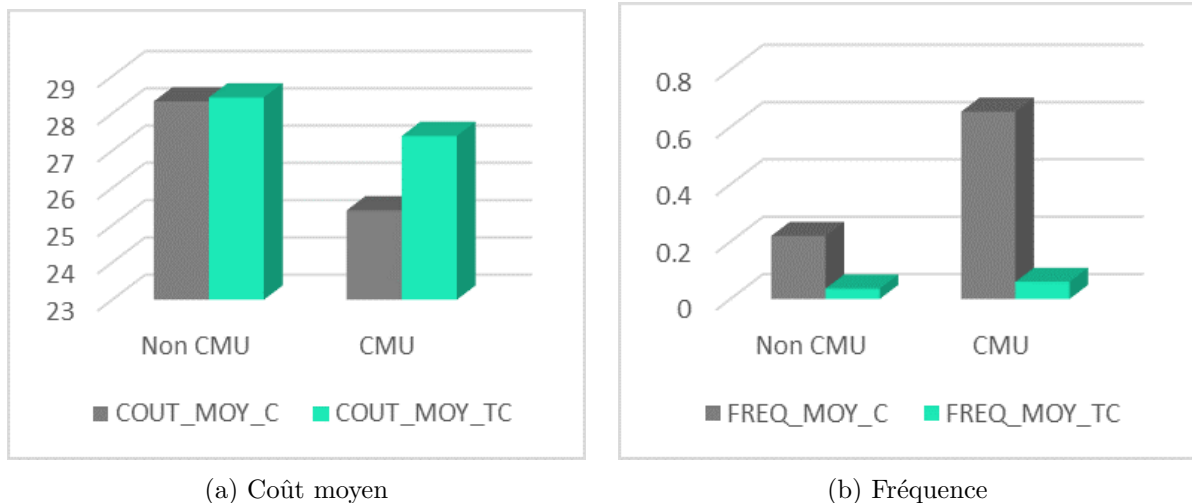


FIGURE 2.13 : Influence de la variable CMU sur la consommation de sinistres

Influence de la région de résidence

Dans l'ensemble, les fréquences de consommation de téléconsultations observées sur le territoire vont de 0,03 à 0,05 actes (de 0,22 à 0,32 pour les consultations) et varient plus ou moins d'une région à l'autre. Les niveaux les plus hauts sont enregistrés dans les régions Île-de-France et PACA/Corse.

Selon les données médicales récoltées lors de l'étude, pour la période entre 2018 et 2020, ces deux zones font partie des régions où la densité médicale est la plus haute. Ainsi les niveaux de fréquence de consommation élevés s'expliquent notamment par un meilleur accès aux soins (et donc une meilleure prise en charge médicale) : environ 340 médecins pour 100 000 habitants contre 276 en moyenne pour les autres régions. Les fréquences de téléconsultation les plus faibles sont enregistrées dans le nord-ouest de la France. Cela touche alors les régions suivantes : Bretagne, Normandie, Pays de la Loire et Centre-Val de Loire. En moyenne ces régions possèdent 257 médecins pour 100 000 habitants contre 298 médecins pour les autres régions.

En terme de coûts, les tendances qui se dégagent des cartes géographiques représentant les coûts moyens de téléconsultation et de consultation, sont assez similaires. Les prestations de téléconsultation sont plus chères en Île-de-France et dans les DOM-TOM. Cette observation est en ligne avec la cherté de la vie enregistrée dans ces régions.

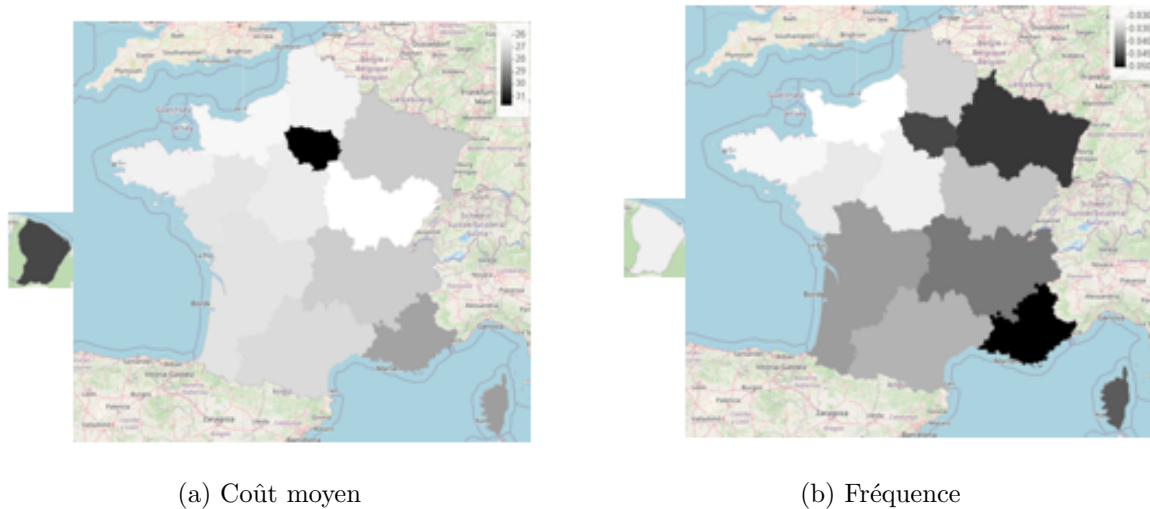


FIGURE 2.14 : Influence de la variable région sur la consommation d'actes de téléconsultation

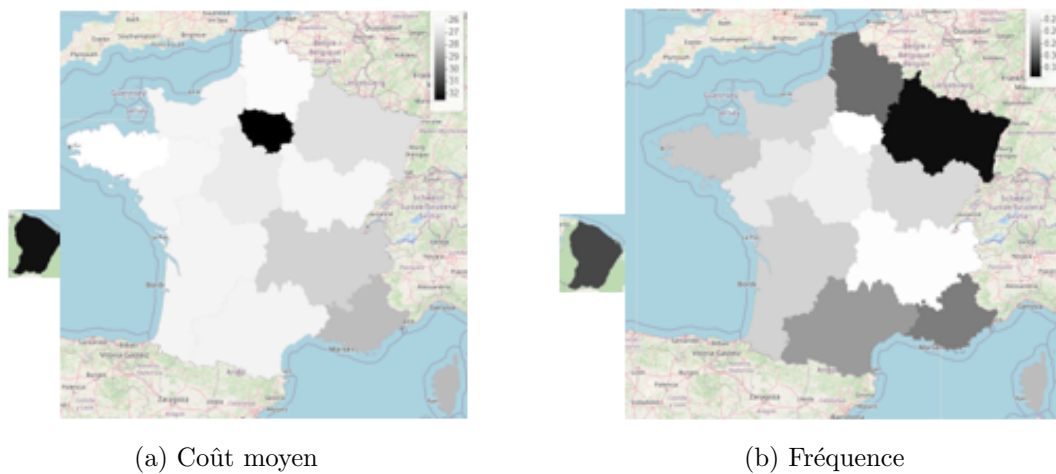


FIGURE 2.15 : Influence de la variable région sur la consommation d'actes de consultation

Lecture : L'Ile-de-France et les régions d'Outre-mer enregistrent les niveaux de dépense les plus haut en matière de téléconsultation. En moyenne, les assurés résidant en Normandie ont consommés 0,26 consultations entre septembre 2018 et décembre 2020.

Influence de l'année de remboursement

L'analyse des coûts moyens a montré qu'au fil des années les coûts de téléconsultation ont diminué, passant de 33,50 euros en 2018 à 28,30 euros en 2020, contrairement aux coûts de consultation qui sont restés réguliers au cours du temps.

Ces observations, mises en perspective avec la décroissance du taux d'inflation (1,8% en 2018, 1,1% en 2019 et 0,5% en 2020), montrent d'une part que ce dernier n'a eu que peu d'influence sur la dépense en consultation entre 2018 et 2020. D'autre part, il est remarqué que, entre 2018 et 2020, les tendances d'évolution du taux d'inflation et du coût de la téléconsultation sont toutes deux décroissantes.

Par ailleurs, la fréquence de consommation d'actes de téléconsultation a augmenté exponentiellement entre 2018 et 2020. En effet, en l'espace de deux ans, cette fréquence s'est multipliée par 3000, jusqu'à atteindre les 0,06 téléconsultations par année et par assuré. Inversement la fréquence de consultation a diminué entre 2019 et 2020, en passant de 0,28 à 0,12 consultations par année et par assuré. Ces observations sont cohérentes avec l'explosion du recours à la téléconsultation en 2020 causé par le confinement, instauré mi-mars pour lutter contre la pandémie de Covid-19, avec plus d'un million de téléconsultations enregistrées par la CNAM pour la seule semaine du 30 mars.

Il est alors possible de conjecturer que la dépense en téléconsultation a diminué au fil du temps car cette pratique a de plus en plus été utilisée pour des soins courants de la vie (une consultation chez un médecin généraliste par exemple), généralement moins chers que des actes médicaux complexes.

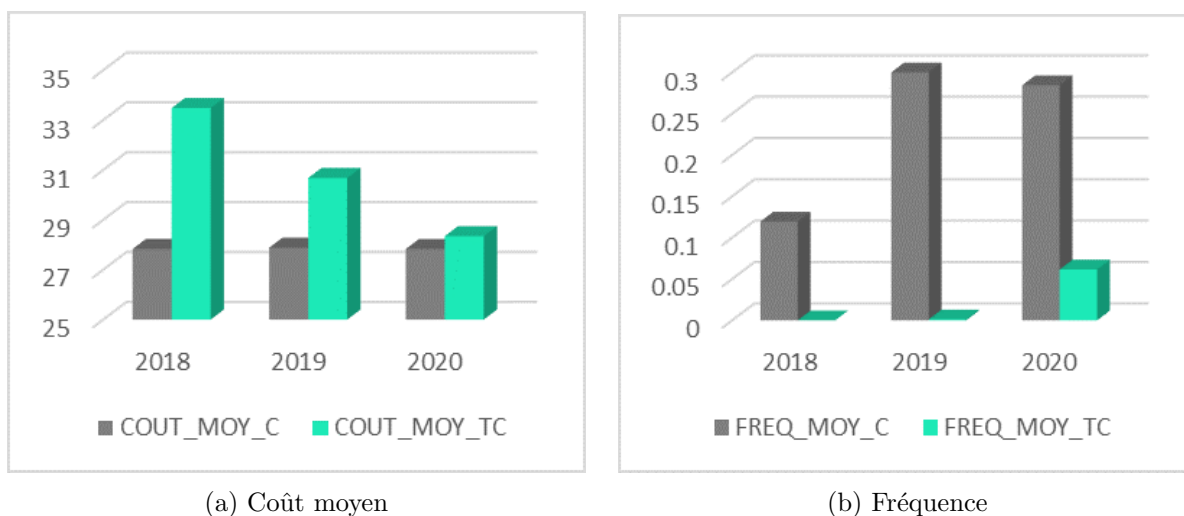


FIGURE 2.16 : Influence de la variable année sur la consommation de sinistres

Influence de la nature des prestations

Grâce au référentiel d'équivalence constitué lors de l'étude, la comparaison entre l'analyse de l'influence de la nature des prestations sur la consommation d'actes de consultation et de téléconsultation a pu être faite.

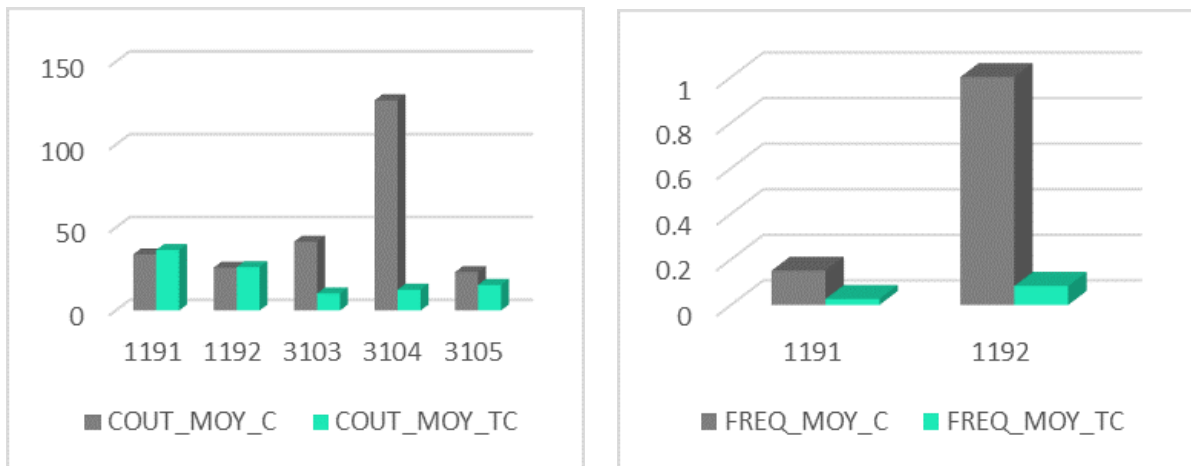
Au niveau des fréquences, les deux pratiques médicales suivent les mêmes tendances : en moyenne les français consomment plus de consultations/téléconsultations chez les médecins généralistes que chez les médecins spécialistes (7,5 fois plus pour la consultation et 3,2 fois plus pour la téléconsultation). Les actes de téléconsultation lors d'un soin (modalité 3103), téléconsultation lieu dédié (modalité 3104) et téléconsultation à domicile (modalité 3105) sont en moyenne 100 fois moins utilisés que les téléconsultations généralistes et spécialistes. Cela s'explique non seulement par le fait que ces

prestations ont été introduites qu'en 2020, mais aussi par le fait qu'elles sont utilisées plus rarement par les assurés (tout comme les prestations de consultation qui leur sont équivalentes).

Au niveau des dépenses, la figure 2.17 révèle une grande disparité de coûts entre téléconsultation et consultation pour la modalité 3104 (126,70 euros pour la consultation contre 12,50 euros pour la téléconsultation). La dépense moyenne pour la consultation est cohérente avec les frais réels correspondant aux consultations dans un centre médico-psycho pédagogique et aux consultations urgentes aux urgences.

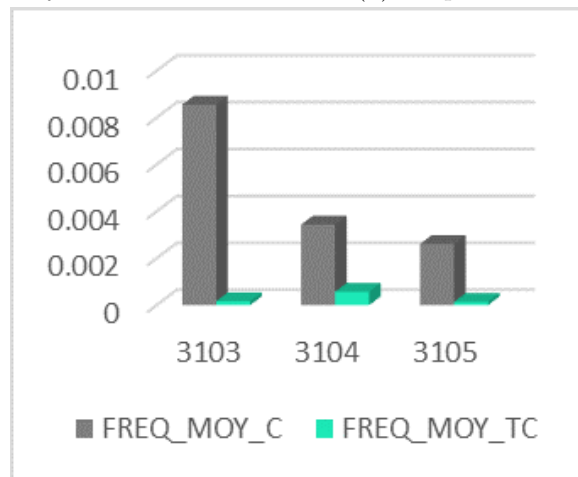
De plus, on remarque que les coûts moyens des prestations 3103, 3104 et 3105 sont en moyenne plus faibles que les autres prestations de consultation/téléconsultation ($\sim 12,70$ euros contre $\sim 40,70$ euros). Ce constat est cohérent avec ce que déclare le GUIDE DGOS (2019) de la Direction Générale de l'Offre de Soins (DGOS) intitulé « Facturation des téléconsultations et des téléexpertises en établissement de santé » : « Au 1er janvier 2020, la prise en charge de l'accompagnement de patients lors d'une téléconsultation dans le cadre des actes et consultations externes sera valorisée de 10 à 15 euros selon les situations suivantes :

- un soin infirmier déjà prévu (acte à 10 euros),
- un soin organisé de manière spécifique à domicile (acte à 15 euros)
- ou encore un soin dans un lieu dédié aux téléconsultations (acte à 12 euros). »



(a) Coût moyen

(b) Fréquence - Généralistes et Spécialistes



(c) Fréquence - Autres prestations

FIGURE 2.17 : Influence de la variable nature de la prestation sur la consommation d'actes de consultation et de téléconsultation

2.4.2 Analyses bivariées de coût moyen et de fréquence

Il est aussi possible d'effectuer des analyses bivariées afin de mettre en relation différentes variables explicatives. Les tendances mises en exergue dans ces analyses sont décrites ci-après.

Influence de l'âge et du sexe

Tant chez les hommes, que chez les femmes, les tendances observées des coûts des deux pratiques médicales par tranche d'âge sont similaires. Toutefois, au niveau des fréquences, il existe une grande disparité entre les deux pratiques médicales : quel que soit l'âge et le sexe la fréquence de consultation est en moyenne 10 fois plus élevée que celle de téléconsultation.

De plus, il est constaté que pour les tranches d'âge 20-29 ans, 30-39 ans et 40-49 ans, la fréquence de consultation des femmes est supérieure à celle des hommes. Cette observation est compréhensible de par le fait qu'il s'agit de la période pendant laquelle la majorité des femmes tombent enceintes, ce qui entraîne alors une plus forte consommation de soins de ville. Néanmoins, cette tendance est inversée pour la téléconsultation. En effet, à tout âge, les hommes consomment plus de téléconsultations que les femmes.

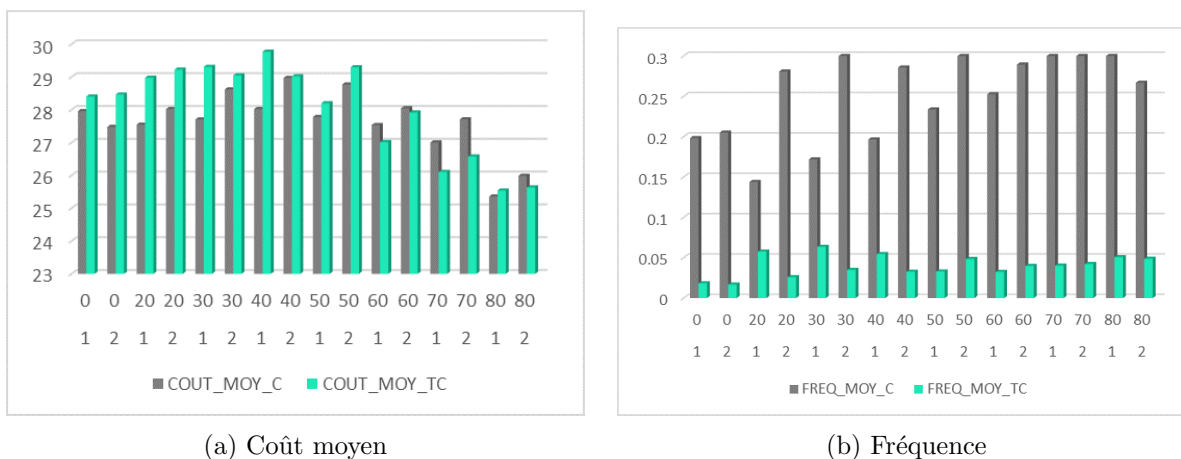


FIGURE 2.18 : Influence des variables âge et sexe sur la consommation d'actes de consultation et de téléconsultation

Influence de la nature de la prestation et de l'année de remboursement

Seules les prestations de consultation/téléconsultation chez les médecins généralistes (modalité 1192) et spécialistes (toute spécialité confondue, modalité 1191) sont incluses dans l'analyse, car les autres actes de téléconsultation n'ont été introduits qu'en 2020 par l'Assurance maladie.

La conjecture faite lors de l'analyse des coûts de téléconsultation selon l'année de remboursement est cohérente avec les résultats illustrés par ces graphes.

En effet, on remarque qu'entre 2018 et 2020, le recours aux téléconsultations généralistes s'est plus développé par rapport aux téléconsultations pratiquées par des médecins spécialistes.

Cependant, il peut être remarqué que la variable année a peu d'effet sur les niveaux de dépense de consultation et de téléconsultation, qui restent homogènes quelle que soit la spécialité du médecin. Les prestations de consultation/téléconsultation chez les médecins spécialistes coûtent en moyenne entre 30 et 35 euros. Tandis que les coûts d'actes de consultation/téléconsultation chez les généralistes sont compris entre 25 et 30 euros.

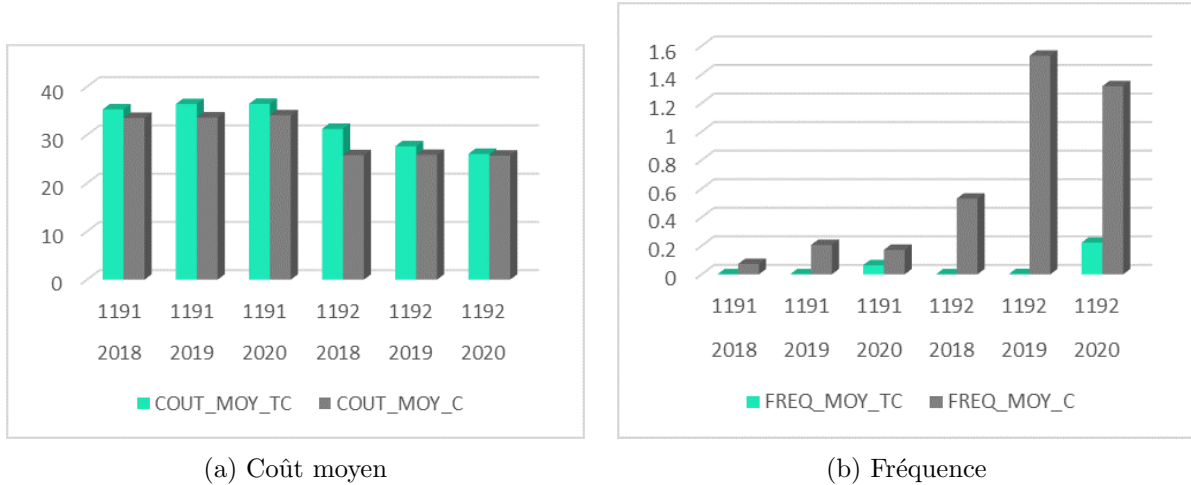


FIGURE 2.19 : Influence des variables nature de la prestation et année sur la consommation d’actes de téléconsultation et de consultation chez les généralistes et les spécialistes

2.4.3 Analyse des dépendances entre variables

Afin d’appliquer les modèles linéaires généralisés (GLM) à notre étude, il est primordial d’analyser la corrélation entre les différentes variables explicatives. En effet, la multicollinéarité rend l’estimation des paramètres du modèle instable et a tendance à fausser les tests statistiques réalisés.

Le V de Cramer est une statistique adaptée à la quantification des liaisons qui existent sur un jeu de données mixtes, c’est-à-dire un jeu de données comportant des variables quantitatives et des variables qualitatives. Parmi les 16 variables considérées dans cette section, seules les variables médicales et économiques sont quantitatives, toutes les autres sont catégorielles. Le V de Cramer est une mesure qui peut être assimilée à une normalisation de la statistique du test d’indépendance du χ^2 de Pearson. Ainsi, avant de la présenter, il convient d’effectuer quelques rappels quant au test d’indépendance du χ^2 .

Test d’indépendance du χ^2

Soient $X = (X_i)_{2 \leq i \leq p}$ et $Y = (Y_j)_{2 \leq j \leq q}$ deux variables auxquelles nous associons le tableau de contingence suivant.

	Y_1	\dots	Y_j	\dots	Y_q	Totaux
X_1	$n_{1,1}$	\dots	$n_{1,j}$	\dots	$n_{1,q}$	$n_{1,\cdot}$
\vdots	\vdots		\vdots		\vdots	\vdots
X_i	$n_{i,1}$	\dots	$n_{i,j}$	\dots	$n_{i,q}$	$n_{i,\cdot}$
\vdots	\vdots		\vdots		\vdots	\vdots
X_p	$n_{p,1}$	\dots	$n_{p,j}$	\dots	$n_{p,q}$	$n_{p,\cdot}$
Totaux	$n_{\cdot,1}$	\dots	$n_{\cdot,j}$	\dots	$n_{\cdot,q}$	n

FIGURE 2.20 : Exemple de tableau de contingence

On définit :

- p : nombre de modalités de la variable X ;
- q : nombre de modalités de la variable Y ;
- $n_{i,j}$: nombre d'individus ayant comme caractéristiques la i -ème modalité de X et la j -ème modalité de Y ;
- $n_{.,j} = \sum_{i=2}^p n_{i,j}$;
- $n_{i,.} = \sum_{j=2}^q n_{i,j}$;
- $n = \sum_{i=2}^p n_{i,.} = \sum_{j=2}^q n_{.,j}$.

Le tableau fournit, pour chaque couple d'indices (i, j) , l'effectif $n_{i,j}$ associé au croisement des modalités (X_i, Y_j) de deux variables considérées. Afin d'étudier la liaison entre ces variables, l'idée du test d'indépendance du χ^2 est de comparer les effectifs observés $n_{i,j}$ aux effectifs théoriques $e_{i,j} = \frac{n_{.,j} \times n_{i,.}}{n}$ que l'on obtiendrait si X et Y étaient indépendants.

Les hypothèses du test sont : (H0) : X et Y sont indépendantes contre (H1) : X et Y ne sont pas indépendantes. La statistique χ^2 quantifie la distance (l'écart) entre les effectifs observés et les effectifs théoriques. Elle est donnée par :

$$\chi^2 = \sum_{i=1}^p \sum_{j=1}^q \frac{(n_{i,j} - e_{i,j})^2}{e_{i,j}}. \quad (2.1)$$

Cette statistique permet d'arbitrer entre la dépendance et l'indépendance. Lorsque $\chi^2 = 0$, X et Y sont indépendantes. Si $\chi^2 > 0$, l'hypothèse d'indépendance entre les deux variables est alors réfutée. Cette interprétation du χ^2 émet des limites, notamment au niveau de la quantification de l'intensité maximale d'une liaison par cette statistique, qui n'est à priori pas possible. En effet, la statistique χ^2 est une somme de fonctions quadratiques des écarts $|n_{i,j} - e_{i,j}|$, elle prend ses valeurs dans $[0, \infty)$. Elle est sensible aux dimensions p et q du tableau de contingence ainsi qu'à l'effectif total n .

V de Cramer

La statistique V de Cramer permet de résoudre cette problématique en proposant une normalisation du χ^2 . Cela permet d'obtenir une mesure comprise entre 0 et 1 dont l'interprétation est la suivante : plus la valeur obtenue est proche de 1 et plus les variables étudiées sont dépendantes. La formule du V de Cramer est donnée par :

$$\sqrt{\frac{\chi^2}{n(\min(p, q) - 1)}}. \quad (2.2)$$

Résultats

Le V de Cramer a ainsi été estimé pour chacune des paires formées par les 16 variables d'intérêts.

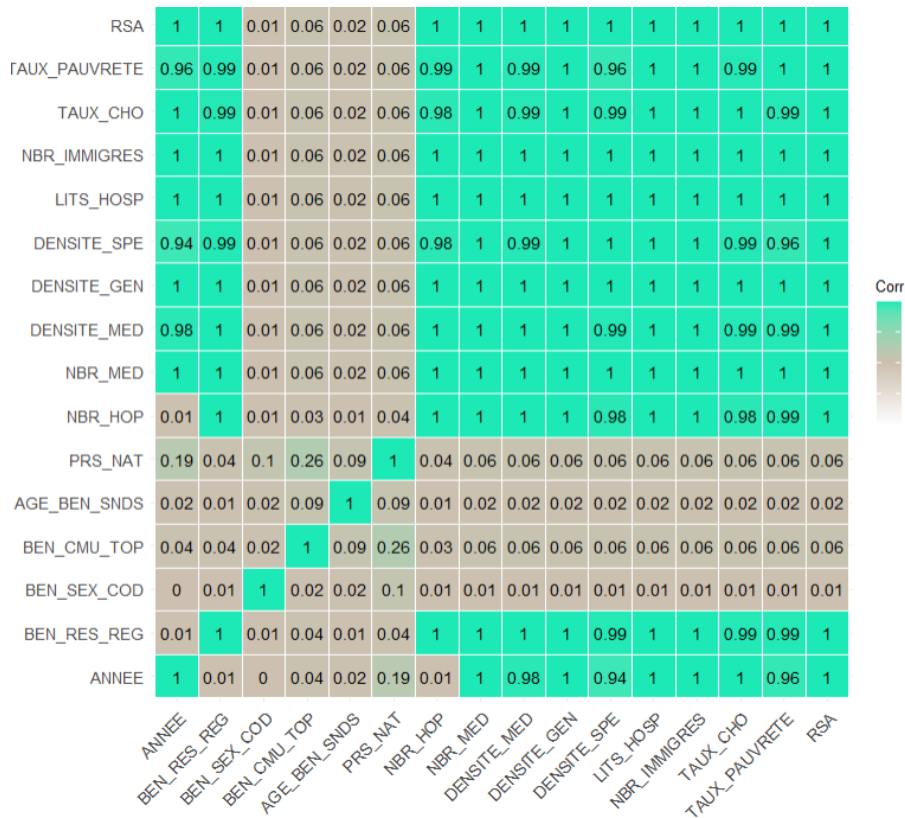


FIGURE 2.21 : V de Cramer

Les résultats donnés par la figure 2.19 montrent que pour chacune des variables qualitatives, aucun coefficient de Cramer associé ne dépasse 0,2. Ces variables peuvent alors toutes être introduites dans les modèles à construire dans le cadre de la tarification des prestations de téléconsultation et de consultation.

En revanche, les variables médicales et économiques seront exclues de la modélisation GLM car, de par leur granularité régionale, elles sont strictement corrélées à la variable région (« BEN_RES_REG »).

Chapitre 3

Analyse de la tarification d'un acte de téléconsultation

Ce chapitre a pour but de comparer l'évolution de la tarification d'un acte de téléconsultation à celle d'un acte de consultation, en portant une attention particulière à la théorie sur lesquels se fondent les modèles utilisés.

3.1 Rappels sur la tarification

3.1.1 La prime d'assurance

La prime assurantielle correspond à une somme d'argent que l'assuré paye en contrepartie de l'engagement de l'assureur de prendre en charge le risque. Cet engagement débutant à la date d'effet des garanties, la prime est payable d'avance et doit être versée à chaque échéance pour la période suivante. On parle aussi de cotisations lorsqu'il s'agit de primes versées aux mutuelles d'assurance et aux instituts de prévoyance. En assurance, le règlement des prestations intervient à une date ultérieure au paiement de la prime, contrairement à l'industrie. C'est l'inversion du cycle de production : le besoin en fonds de roulement est négatif.

Par ailleurs, les assureurs distinguent plusieurs types de primes :

- La prime pure, appelée aussi prime technique, correspond à la tarification du risque. C'est le montant probable que l'assureur va devoir payer au titre des indemnités versées aux bénéficiaires. Elle coïncide donc avec l'espérance des sinistres. Puisqu'en assurance santé il existe différentes sous-familles d'actes pour chaque poste de soins, la prime pure globale est définie comme étant la somme des primes pures de ces différents risques sous hypothèse d'indépendance de ces derniers. En réalité, les risques ne sont pas totalement indépendants (une consultation chez le médecin conduit souvent à des achats de médicaments en pharmacie par exemple) mais dans le cadre de l'étude, une telle hypothèse théorique sera prise.
- La prime commerciale est quant à elle la prime réellement payée par l'assuré. Elle correspond à la prime pure majorée d'une marge de sécurité, de chargements de gestion et d'acquisition et de taxes. Le chargement de sécurité permet de couvrir le risque d'une mauvaise tarification ou encore de la volatilité. Les chargements de gestion et d'acquisition couvrent les coûts de

l'entreprise, y compris les commissions versées aux intermédiaires et aux partenaires (courtiers, agents généraux, ect...).

Ainsi :

$$\begin{aligned} \text{Prime commerciale} = & \text{Prime pure globale} + \text{Chargements de sécurité} \\ & + \text{Chargements de gestion et d'acquisition.} \end{aligned} \quad (3.1)$$

Dans le cadre des travaux, seule la prime pure sera d'intérêt et par abus de langage, elle sera souvent désignée par le terme de « tarif ».

3.1.2 Théorie du modèle coût-fréquence pour la tarification

Traditionnellement, l'approche « coût-fréquence » est utilisée afin de calculer la prime pure d'un contrat d'assurance santé. Cette démarche permet à l'assureur d'estimer le coût que représente les risques en portefeuille et nécessite une modélisation séparée du coût moyen des sinistres et de leur fréquence de survenance, en supposant que ceux-ci sont indépendants.

La nature du test à utiliser pour vérifier l'indépendance entre le coût moyen et la fréquence (c'est-à-dire deux variables quantitatives) diffère lorsque la loi du couple (X, Y) suit une distribution binormale. La binormalité correspond à une distribution normale, en chaque point de X , de la variable Y et vice versa. Une loi binormale est alors une loi normale bi-dimensionnelle.

1. Lorsque la loi du couple est binormale : on considère un test paramétrique sur le coefficient de corrélation linéaire (Pearson).
2. Lorsque la loi du couple n'est pas binormale : on considère un test non paramétrique basé sur le coefficient de corrélation empirique de Spearman.

Vérification de la binormalité

Pour évaluer la binormalité de la loi du couple (Coût moyen, Fréquence), la fonction `R` appelée `mqqnorm` du package `RVAideMemoire` a été utilisée. Elle permet d'obtenir le qqplot d'une distribution multivariée.

La figure 3.1 illustre les qqplots des distributions bivariées du coût moyen et de la fréquence de consultation et de téléconsultation.

Il est observé que quelle que soit la pratique médicale considérée, les points ne semblent pas s'aligner le long de la droite et forment une courbe. Or cette droite représente les quantiles d'une distribution binormale centrée et réduite.

Il est alors conclu, par approche graphique, que les distributions bivariées du coût moyen et de la fréquence de consultation et de téléconsultation ne sont pas binormales.

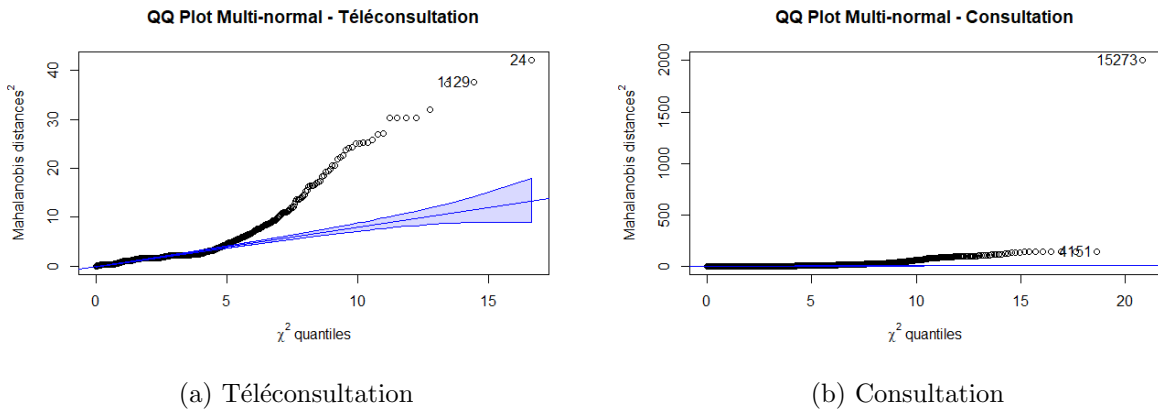


FIGURE 3.1 : QQplot des distributions bivariées du couple de variables (Coût moyen, Fréquence) pour la consultation et la téléconsultation

Vérification de l'indépendance entre coût moyen et fréquence de survenance

Suite aux résultats précédents, le test de corrélation de Spearman a été utilisé pour estimer une éventuelle association entre le coût moyen et la fréquence de consultation et de téléconsultation.

La fonction `R cor.test(x,y)` calcule le coefficient de corrélation entre deux variables ainsi que le niveau de significativité de la corrélation (ie la p-valeur).

Voici les résultats du test :

- Pour la téléconsultation : la p-valeur vaut 0,052 (supérieure à la valeur critique 0,05) et coefficient est égal à -0.15 . Il est déduit que le coût moyen et la fréquence ne sont pas corrélés. Ainsi, le recours à l'approche coût-fréquence est justifié.
- Pour la consultation : la p-valeur vaut 0,071 et coefficient est égal à 0.28 . Ainsi, l'approche coût-fréquence peut aussi être employée pour tarifer ce poste de soins.

Fondements du modèle collectif

Le modèle collectif donne la somme agrégée des montants des sinistres sur une période d'exercice donnée (ici 1 an) à la maille portefeuille plutôt qu'à la maille contrat ou groupe homogène de contrat.

Il est supposé que :

- le nombre de sinistres pour une année donnée est représenté par une variable de comptage $N \in \mathbb{N}$, de loi notée

$$p_N(n) = \mathbb{P}(N = n), \quad n = 0, 1, \dots$$

- les montants de sinistres forment une suite de variables aléatoires positives i.i.d. $(X_i, i = 1, 2, \dots, N)$ telle que pour tout i , X_i représente la charge du sinistre i . On considère que pour tout i strictement positif, les variables aléatoires N et X_i sont supposées indépendantes.

La charge totale de sinistres est définie par

$$S = \sum_{i=1}^N X_i. \quad (3.2)$$

L'objectif est d'estimer le montant moyen de sinistres attendu par l'assureur à horizon 1 an i.e. la prime pure annuelle. Mathématiquement, cette prime pure est égale, sur l'ensemble du portefeuille, à l'espérance des pertes $\mathbb{E}[S]$.

Grâce aux hypothèses précédentes et par utilisation de la formule des espérances conditionnelles totales, on obtient :

$$\mathbb{E}[S] = \mathbb{E}[N \times X_1] = \underbrace{\mathbb{E}[N]}_{\text{fréquence}} \times \underbrace{\mathbb{E}[X_1]}_{\text{Coût moyen}} . \quad (3.3)$$

Ainsi pour chaque risque :

$$\text{Prime pure} = \text{Fréquence} \times \text{Coût moyen}. \quad (3.4)$$

Les définitions suivantes sont considérées :

$$\text{Fréquence} = \frac{\text{Nombre total de sinistres}}{\text{Exposition totale}} = \frac{\text{Quantité totale d'actes de soins}}{\text{Somme des expositions annuelle des bénéficiaires du contrat}^*}, \quad (3.5)$$

et :

$$\text{Coût moyen} = \frac{\text{Coût total des actes de soins}}{\text{Quantité totale d'actes de soins}}. \quad (3.6)$$

La décomposition de cette prime pure en un produit coût \times fréquence permet de cerner les divers éléments susceptibles d'influer la loi de probabilité adossée à la fréquence de survenance des sinistres ainsi que ceux pouvant influencer la loi de probabilité adossée au coût des sinistres.

3.2 Théorie des Modèles linéaires généralisés

Classiquement, la modélisation du coût moyen des sinistres et de leur fréquence de survenance se fait à l'aide des modèles linéaires généralisés (GLM). Cette section, consacrée à la description théorique de ces modèles, est inspirée du cours sur le Modèle linéaire dispensé par Sophie Donnet à l'Université Paris Dauphine, DONNET (2019).

3.2.1 Un rappel sur le modèle linéaire gaussien

L'objectif général d'un modèle linéaire est d'expliquer l'évolution d'une variable Y , dite réponse, variable exogène ou variable à expliquer, en fonction de p variables X_1, \dots, X_p , dites variables explicatives ou endogènes. Cela se traduit par la relation suivante :

$$\forall i = 1, \dots, n, \quad y_i = \beta_0 + \beta_1 x_1^i + \dots + \beta_p x_p^i + e_i, \quad (3.7)$$

où, pour n individus :

- $Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{R}^n$ est le vecteur aléatoire à expliquer. Pour tout i , la donnée y_i correspond au $i^{\text{ème}}$ individu ;

*L'exposition annuelle d'un bénéficiaire est calculée au prorata de la durée du contrat souscrit par celui-ci.

- $\forall j = 1, \dots, p, X_j = \begin{pmatrix} x_1^j \\ x_2^j \\ \vdots \\ x_n^j \end{pmatrix} \in \mathbb{R}^n$ inclut les observations de la $j^{\text{ième}}$ variable explicative;
- $Y = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} \in \mathbb{R}^p$ est un vecteur inconnu à p lignes. C'est un paramètre du modèle à estimer.

Cette estimation se fait généralement par la méthode des moindres carrés tel que l'estimateur $\hat{\beta}$ en découlant vérifie :

$$\hat{\beta} = (X'X)^{-1} X'Y; \quad (3.8)$$

- $\epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$ est le terme d'erreur aléatoire du modèle qui représente l'écart entre la valeur observée et la valeur estimée par combinaison des variables explicatives.

Le caractère gaussien du modèle de régression linéaire se distingue par les termes d'erreur ϵ qui sont distribués selon la loi d'un vecteur gaussien de taille n , d'espérance nulle et de matrice de variance-covariance diagonale :

$$\epsilon \sim \mathcal{N}(0_{\mathbb{R}^n}, \sigma^2 I_n).$$

La formule 3.7 peut se réécrire sous forme matricielle comme suit :

$$Y = X\beta + \epsilon, \quad (3.9)$$

avec X , matrice $n * p$ non-aléatoire, complètement observée, qui se définit par

$$X = \begin{pmatrix} 1 & x_1^1 & \dots & x_1^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_i^1 & \dots & x_i^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n^1 & \dots & x_n^p \end{pmatrix} = (\mathbf{1} \ X_1 \ \dots \ X_p).$$

Dans le cadre de ce mémoire, Y représentera soit le coût moyen, soit la fréquence de survenance associée à la téléconsultation et à la consultation traditionnelle.

La loi sur les termes d'erreur implique quatre postulats concernant les observations, vérifiable par des outils graphiques et non par des tests statistiques :

1. Les erreurs sont centrées : $\mathbb{E}[\epsilon] = 0_{\mathbb{R}^n}$. En pratique, cela retranscrit la pertinence du modèle et montre qu'aucun terme déterminant a été oublié.
2. Les erreurs sont de variance constante : $\text{Var}[\epsilon_i] = \sigma^2, \forall i \in \{1, \dots, n\}$. On parle de modèle homoscédastique, par opposition à un modèle hétéroscédastique où le terme d'erreur n'aurait pas la même variance pour toutes les observations.

3. Les termes d'erreur sont supposés indépendants. Les observations sont ainsi supposées indépendantes, résultant soit d'un échantillonnage indépendant soit d'une expérience menée dans des conditions indépendantes.
4. Les erreurs sont supposées gaussiennes. Toutefois, ce postulat pourra être négligé si le nombre d'observations est grand (au delà de 20 ou 30 observations).

Ainsi, ces postulats supposent que la variable à expliquer Y suit une loi normale d'espérance $X\beta$ et de variance $\sigma^2\mathbb{I}_n$. L'usage du modèle linéaire gaussien est alors considéré inadéquat dans de nombreuses situations, notamment si la variable à modéliser Y est de nombre fini ou dénombrable, allant à l'encontre du 4ième postulat ou encore si Y ne prend pas des valeurs négatives (fréquence/montants de sinistres...), contrairement aux variables gaussiennes qui sont à valeurs dans \mathbb{R} (conséquence du 1er postulat : $\mathbb{E}[y_i] = X'_i\beta$).

Le modèle linéaire généralisé permet de supprimer ces contraintes concernant la variable à expliquer.

3.2.2 Modèles linéaires généralisés

Le modèle linéaire généralisé est une extension de la régression linéaire, permettant de se libérer des postulats 1, 2 et 3 et de traiter des observations dont la loi de probabilité appartient à une famille de lois plus large. Il s'agit de la donnée d'une loi de probabilité pour les y_i et d'une fonction g , appelée fonction de lien, telle que $g(\mathbb{E}(y_i)) = X'_i\beta$. La fonction de lien permet ainsi d'obtenir une relation non linéaire entre l'espérance de la variable réponse et les variables qualitatives.

La loi de distribution associée à la variable à expliquer doit appartenir à la famille exponentielle naturelle. C'est une famille de lois de probabilité qui ont en commun une écriture sous forme exponentielle, permettant d'unifier la présentation des résultats. Soit f_Y la densité de probabilité de la variable Y , f_Y appartient à la famille exponentielle naturelle si elle s'écrit sous la forme suivante :

$$f_Y(y) = \exp \left[\frac{1}{a(\phi)}(y\theta - b(\theta)) + c(y, \phi) \right],$$

où c représente une fonction dérivable, b désigne une fonction trois fois dérivable de dérivée première inversible, θ est le paramètre naturel de la loi et ϕ est le paramètre de dispersion. Cette famille de lois se distingue aussi par les résultats suivants :

$$\begin{aligned} \mathbb{E}[Y] &= b'(\theta), \\ \text{Var}[Y] &= b''(\theta) \times a(\phi). \end{aligned} \tag{3.10}$$

Implémenter un modèle linéaire généralisé requiert le choix de deux éléments :

- Le choix d'une loi de probabilité pour les variables aléatoires y_i , ce choix étant guidé par la nature du problème ;
- Le choix de la fonction de lien g permettant de modéliser le lien entre l'espérance des y_i et les variables explicatives.

Le choix de la fonction de lien :

Toute bijection de l'espace de $\mathbb{E}[Y]$ dans \mathbb{R} peut être choisie comme fonction de lien. Toutefois, souvent la fonction de lien considérée est la fonction qui transforme l'espérance $\mathbb{E}(Y)$ en $g = (b')^{-1}$, le paramètre naturel. On appelle alors g fonction de lien canonique.

Le tableau suivant présente quelques modèles linéaires usuels. A chaque loi $Y = x$ correspond une fonction de lien canonique $g(\cdot)$, qui donne son nom à la régression.

Choix de la loi de $Y \mid x$	Bernoulli/ Binomial	Poisson	Gamma	Gaussienne
Fonction de lien canonique	$g(\mu) = \text{logit}(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$	$g(\mu) = \log(\mu)$	$g(\mu) = -\frac{1}{\mu}$	$g(\mu) = \mu$
Nom du lien	logit	log	réciroque	identité

FIGURE 3.2 : MLG usuels

D'autres fonctions de lien non canoniques peuvent être usitées :

- Le lien *probit*
- Le lien *log-log*

Par ailleurs, le choix de la fonction de lien a des conséquences sur l'estimation des paramètres β du modèle, qui se fait par la méthode du maximum de vraisemblance. Pour Y issue d'une famille exponentielle, l'expression de la vraisemblance s'écrit :

$$L(y_1, \dots, y_n; \theta, \Phi) = \exp\left(\sum_{k=1}^n \frac{y_k \theta_k - b(\theta_k)}{a_k(\Phi)} + c(y_k, \Phi)\right). \quad (3.11)$$

Il découle ainsi l'expression de la log-vraisemblance :

$$\log(L) = \sum_{k=1}^n \frac{y_k \theta_k - b(\theta_k)}{a_k(\Phi)} + c(y_k, \Phi). \quad (3.12)$$

Les paramètres $\hat{\beta}_i$ sont alors obtenus suite à la résolution du système d'équation suivant :

$$\frac{\partial}{\partial \beta_i} \log(L) = \sum_{k=1}^n \frac{\partial}{\partial \beta_i} \left(\frac{y_k \theta_k - b(\theta_k)}{a_k(\Phi)} + c(y_k, \Phi) \right) = 0. \quad (3.13)$$

Les équations de vraisemblance n'ont pas de solution explicite en général, sauf pour le cas du modèle linéaire gaussien. Il faut recourir à des algorithmes d'optimisation itératifs pour approcher la solution. Un algorithme souvent utilisé est celui de Newton-Raphson. Il est possible de démontrer que dans le cas de la fonction de lien canonique, cet algorithme converge vers le maximum de vraisemblance. Dans le cas particulier de la fonction de lien canonique, les formules suivantes sont donc appliquées :

$$g = (b')^{-1} \text{ et } \mathbb{E}(Y) = b'(\theta_i) = b'(x_i \beta). \quad (3.14)$$

Modélisation du nombre de sinistres

Tel que mentionné auparavant, afin de calculer la prime pure, il est nécessaire de déterminer la quantité probable de sinistres qui devront être indemnisés durant un temps de couverture donné (1 an dans ce mémoire). Cette variable s'apparente à un processus de comptage classiquement modélisé par soit une loi de Poisson, supposant une équi-dispersion des données, soit une loi binomiale négative.

Loi de Poisson

Une variable X suit une loi de Poisson de paramètre $\lambda > 0$ ($X \sim P(\lambda)$) si pour tout $k \in \mathbb{N}$ sa loi de probabilité s'écrit :

$$\mathbb{P}(X = k; \lambda) = \frac{\lambda^k \times \exp(-\lambda)}{k!}. \quad (3.15)$$

On définit $\mathbb{P}(X = k; \lambda)$ comme la somme des expositions individuelles des assurés. A titre d'illustration, un assuré bénéficiant d'un contrat prenant effet le 01/09/2018 avec comme date de fin de couverture le 30/12/2018, comptera pour $\sim 0,33$ personne pour l'année 2018. Cette loi assure des nombres de sinistres entiers et positifs et vérifie :

$$\mathbb{E}[X] = \text{Var}[X] = \lambda. \quad (3.16)$$

Ainsi, l'estimation de la quantité de sinistres d'un individu i en fonction de diverses variables explicatives sera réalisé à travers un modèle linéaire généralisé avec la fonction logarithmique $g(x) = \ln(x)$ pour fonction de lien canonique, come rappelé dans le tableau 3.1. L'équation de la régression s'écrira alors :

$$g(\mathbb{E}[y_i]) = \ln(\lambda(x_{1i}, x_{2i} \dots x_{ni})) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}, \quad (3.17)$$

où y_i est la réalisation d'une loi de poisson de paramètre λ correspondant au nombre de sinistres

de l'individu i et $X_i = \begin{pmatrix} x_{1i} \\ x_{2i} \\ \vdots \\ x_{ni} \end{pmatrix}$ est le vecteur des variables explicatives influençant la valeur de y_i .

Cela revient à supposer que $y_i \sim P(\exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}))$ et à estimer les coefficients par la méthode du maximum de vraisemblance présentée supra.

Toutefois, la loi de Poisson repose sur une hypothèse forte : l'équi-dispersion des données via l'égalité de l'espérance et de la variance. Or cette hypothèse est rarement vérifiée en pratique. Lorsqu'il y a une trop forte dispersion, une autre loi plus adaptée est utilisée. Il s'agit de la loi Binomiale négative. En effet, cette loi de fréquence est caractérisée par une variance plus forte que l'espérance et peut permettre une meilleure adéquation des données au modèle.

Loi binomiale négative

La loi Binomiale négative est une combinaison d'une loi de Poisson et d'une loi Gamma. Ce modèle s'adapte bien aux données qui suivent une loi de Poisson de paramètre aléatoire. A titre d'exemple, les données peuvent être assimilées à différentes populations qui sont caractérisées par des risques moyens différents. Cette loi appartient également à la famille exponentielle et a la même fonction de lien g que la loi de Poisson :

$$\forall k \in \mathbb{N}, f(k) = \binom{k+r-1}{k} (1-p)^r p^k = e^{k \log(p) + r \log(1-p) + \log\left(\binom{k+r-1}{k}\right)}.$$

On considère $r > 0$ et $0 < p < 1$ comme étant les paramètres de la loi.

Classiquement, cette loi est choisie pour inclure une sur-dispersion. Celle-ci est introduite via un paramètre τ_i . Pour chaque individu i , les expressions de l'espérance conditionnelle et de la densité conditionnelle en fonction des variables explicatives sont les suivantes :

$$\begin{aligned}\mathbb{E}[y_i | X_i, \tau_i] &= \exp(X_i^t \beta + \epsilon_i) = \mu_i \tau_i, \\ f(y_i | X_i, \tau_i) &= \exp(-\mu_i \tau_i) \frac{(\mu_i \tau_i)^{y_i}}{y_i!}.\end{aligned}\tag{3.18}$$

Le terme τ_i suit une loi Gamma d'espérance 1 et de variance $\frac{1}{\theta}$ permettant de déterminer la densité de la variable y_i conditionnellement à X_i :

$$f(y_i | X_i) = \frac{\Gamma(y_i + \theta)}{\Gamma(y_i + 1) \Gamma(\theta)} \left(\frac{\theta}{\theta + \mu_i} \right)^\theta \frac{\mu_i^{y_i}}{\mu_i + \theta}.\tag{3.19}$$

Ce qui permet d'obtenir :

$$\mathbb{E}[y_i | X_i] = \mu_i \text{ et } \text{Var}[y_i | X_i] = \left(1 + \frac{\mu_i}{\theta}\right) \mu_i = \Phi \mathbb{E}[y_i | X_i],$$

$\Phi = \left(1 + \frac{\mu_i}{\theta}\right) > 1$ correspondant au coefficient de dispersion à estimer.

Ainsi, la loi binomiale négative prend bien en compte le phénomène de sur-dispersion.

Modélisation du coût moyen des sinistres

Le coût moyen des sinistres est une variable aléatoire continue et positive. Ainsi une modélisation par la loi normale classique ne serait pas pertinente. Les distributions les plus utilisées dans ce cas sont les lois Gamma et log-normale. Le choix final portera sur la loi s'ajustant au mieux aux données traitées.

Loi Gamma

La loi Gamma possède une distribution strictement positive et asymétrique, modulée par deux paramètres, un paramètre de forme $\alpha > 0$ et un paramètre d'échelle $\beta > 0$. Elle appartient également à la famille des lois exponentielles et a pour fonction de lien $g(u) = \log(u)$.

Mathématiquement, une variable $X \geq 0$ suit une loi Gamma de paramètres $\alpha > 0$ et $\beta > 0$ si elle vérifie :

- Fonction de densité : $\forall x \in \mathbb{R}_+, f(x) = \frac{1}{\Gamma(\alpha)} \beta^\alpha x^{\alpha-1} \exp^{-\beta x}$, avec $\Gamma(\cdot)$ la fonction Gamma ;
- Espérance : $\mathbb{E}[X] = \frac{\alpha}{\beta}$;
- Variance : $\text{Var}[X] = \frac{\alpha}{\beta^2} = \Phi \mathbb{E}[Y]$ avec $\Phi = \frac{1}{\beta}$.

On remarque que pour $\alpha = 1$ la loi gamma correspond à la loi exponentielle. De plus, étant donné le lien entre la moyenne et la variance, elle tient compte d'une éventuelle hétéroscédasticité dans les données. Concernant la modélisation, bien que le lien canonique de la loi Gamma soit la fonction inverse, il est fréquent d'utiliser un lien logarithmique pour privilégier une forme multiplicative permettant des interprétations plus simples.

Loi log-normale

Une variable aléatoire $X \geq 0$ suit une loi log-normale de paramètres μ et σ ($\mu, \sigma \in \mathbb{R}$) si elle vérifie :

- Fonction de densité : $f(x) = \frac{1}{x} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\ln(Y)-\mu}{\sigma}\right)^2\right)$;
- Espérance : $\mathbb{E}[X] = \exp\left(\mu + \frac{\sigma^2}{2}\right)$;
- Variance : $\text{Var}[X] = \exp(2\mu + \sigma^2) (\exp(\sigma^2) - 1)$.

La loi Log-normale n'appartient pas à la famille exponentielle. Cependant, une régression log-normale peut être obtenue en considérant une régression linéaire gaussienne appliquée sur le logarithme du coût. Pour passer des estimations du modèle log-normal aux prédictions du coût, un terme doit être rajouté. Mathématiquement, cela est obtenu :

$$\begin{aligned} Y \sim \mathcal{LN}(\mu, \sigma^2) &\Leftrightarrow \log(Y) \sim \mathcal{N}(\mu, \sigma^2), \\ \forall i = 1, \dots, n, \mathbb{E}[\log(Y_i)] &= X_i' \beta. \end{aligned} \tag{3.20}$$

La différence entre la loi Gamma et la loi Log-normale réside dans la sensibilité aux valeurs extrêmes. La loi Gamma est plus sensible à ces valeurs, ainsi l'application de la fonction logarithme permet de diminuer l'importance des sinistres extrêmes.

Choix du modèle :

L'objectif du processus de modélisation est d'aboutir à un modèle s'adaptant au mieux la variable réponse. Il est donc important de sélectionner soigneusement les variables explicatives. Pour ce faire, plusieurs méthodes et critères de sélection existent. Les méthodes les plus utilisées sont présentées ci-après.

La sélection de variables

La sélection de variables permet de déterminer la meilleure combinaison de variables explicatives à retenir. Certaines variables n'apportent pas d'explications supplémentaires sur la variable réponse, il est donc préférable de les éliminer du modèle afin de ne pas augmenter la variabilité des estimateurs et de minimiser la complexité du modèle. Pour ce faire, il existe quatre procédures de sélection de variables connues permettant de limiter le nombre de modèles à tester :

- Sélection exhaustive : le principe est de tester toutes les combinaisons possibles de variables explicatives dans le but d'identifier le meilleur modèle selon un critère donné. Cette méthode n'est pas la plus effective en raison du temps de calcul qui peut être considérable, particulièrement si le nombre de variables explicatives est important.
- Sélection Forward : l'initialisation se fait à partir du modèle sans variable auquel on rajoute une par une les variables explicatives. Si l'ajout d'une variable explicative donnée augmente la qualité du modèle selon un critère donné, cette dernière est retenue. Et ainsi de suite, jusqu'à ce que toutes les variables explicatives soient testées. Toutefois, cette méthode a un inconvénient : lorsqu'une variable est retenue, elle ne peut être écartée ensuite au fil de l'ajout des autres variables, alors que la significativité d'une variable donnée peut être impactée lorsqu'elle est combinée à une autre.
- Sélection backward : il s'agit du même procédé que précédemment mais dans le sens contraire. L'initialisation se fait à partir du modèle complet et une variable est ôtée à chaque itération. Si le modèle sans une variable donnée s'avère meilleur que le modèle l'incluant, la variable n'est pas retenue pour la modélisation.

- Sélection stepwise : c'est une combinaison des méthodes forward et backward. A chaque étape, une variable est enlevée et une variable est ajoutée. Ainsi, une variable jugée intéressante à une étape est réévaluée par la suite afin de s'assurer qu'elle l'est toujours. L'algorithme s'arrête lorsqu'aucune variable n'est ajoutée ou supprimée.

L'implémentation des méthodes de sélection de variables, vues plus haut, nécessite des critères d'évaluation afin de définir le meilleur sous-ensemble de variables. Les mesures les plus souvent considérées sont les critères AIC et BIC.

Les critères AIC et BIC sont des mesures qui apportent une information sur le niveau des maximums de vraisemblance des modèles potentiels tout en pénalisant ceux comportant un nombre important de variables pouvant conduire à du sur-apprentissage. Ces critères d'information permettent un arbitrage entre l'information apportée par des variables supplémentaires et la complexité du modèle (nombre de paramètres à estimer). Le meilleur modèle sera celui minimisant ces critères.

On définit :

$$\begin{aligned} AIC &= -2 \log(LL) + 2k, \\ BIC &= -2 \log(LL) + k \log(n), \end{aligned} \tag{3.21}$$

avec $\log(LL)$ la log vraisemblance, k le nombre de paramètres estimés et n le nombre d'observations.

Les critères de comparaison de modèles

Afin de sélectionner le modèle qui s'adapte au mieux aux données, en plus de la sélection des variables explicatives, il est possible d'étudier la qualité de prédiction des modèles à comparer. Pour cela, il est d'usage de recourir aux critères de sélection suivants :

- La déviance : cette mesure est basée sur la vraisemblance du modèle. Elle compare la vraisemblance du modèle calibré à celle du modèle parfait en matière d'adéquation aux données, dit modèle saturé. Plus la vraisemblance est proche de 1, mieux le modèle s'ajuste aux données. L'expression de la déviance est la suivante :

$$D = 2 \left(\underbrace{\log L(y, y, \Phi)}_{\text{modèle saturé}} - \underbrace{\log L(y, \mu, \Phi)}_{\text{modèle calibré}} \right). \tag{3.22}$$

Ainsi, le meilleur modèle sera celui qui aura la déviance la plus faible.

- L'erreur quadratique moyenne (RMSE) : elle mesure la qualité de prédiction d'un modèle en s'intéressant aux résidus résultant par cette modélisation. Ainsi, la racine carrée de la moyenne du carré des écarts entre chaque valeur prédite et sa valeur réelle constitue la RMSE :

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2},$$

avec :

- y_i la valeur observée de la variable à expliquer pour l'individu i
- \hat{y}_i la valeur prédite de la variable à expliquer pour l'individu i

- n le nombre d'observations dans la base de test

Le modèle à privilégier sera alors celui qui présente la RMSE la plus faible.

3.3 Modélisation de la sinistralité par une approche coût-fréquence

Cette partie est consacrée à l'application des fondements théoriques de la tarification rappelés dans la section précédente, afin de construire des modèles tarifaires pour la consultation et la téléconsultation. Une comparaison des résultats des deux modèles sera ensuite réalisée. Par soucis de concision et de clarté, seule la modélisation du coût moyen et de la fréquence des prestations de téléconsultation sera détaillée par la suite. Le résumé de la modélisation tarifaire d'actes de consultation (effectuée en appliquant exactement les mêmes méthodes et tests que pour la téléconsultation) pourra être consulté dans l'annexe 2.

3.3.1 Modélisation du coût moyen de téléconsultation

Dans cette partie, les mêmes méthodes de modélisation utilisées pour le coût moyen de téléconsultation seront appliquées à la distribution de la fréquence de survenance de téléconsultation.

Choix d'une loi de probabilité

Comme mentionné plus haut, pour mesurer l'adéquation des lois Gamma et Lognormale à la base du coût de téléconsultation, la première approche (qualitative) a été de comparer graphiquement les distributions théoriques des lois à la distribution empirique des données. La figure 3.3 présente les résultats de l'estimation par la méthode du maximum de vraisemblance des paramètres des lois Gamma et Lognormale.

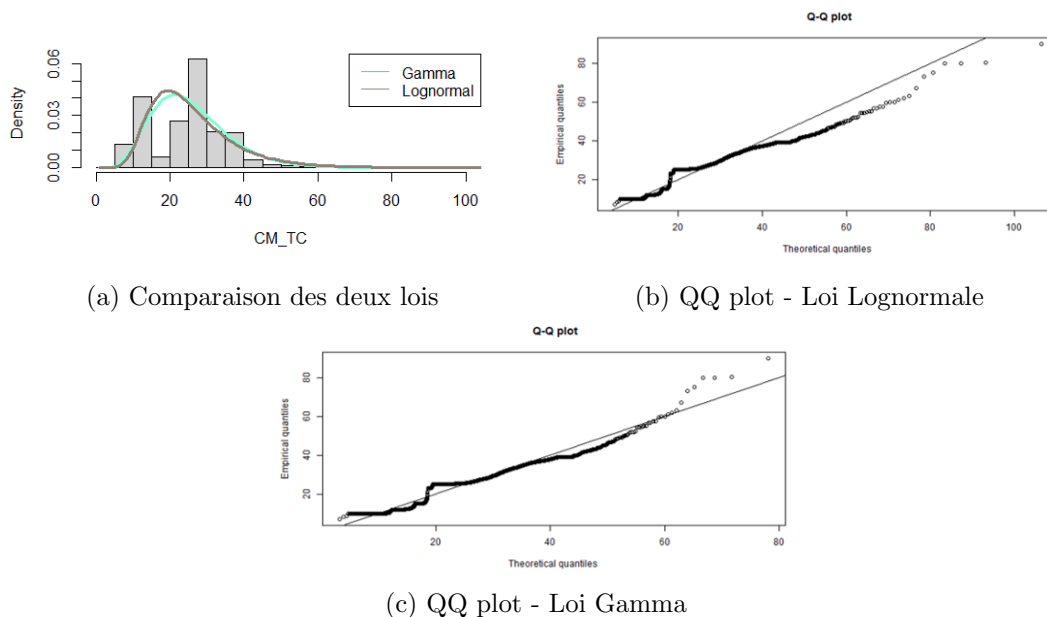


FIGURE 3.3 : Adéquation graphique des modèles de coût moyen pour les prestations de téléconsultation

L'analyse des graphiques révèle qu'aucune des lois n'est adéquate pour la distribution des coûts moyens observés. En effet, on remarque la présence de deux « classes ou zones » de la variable coût, la première incluant les coûts allant de 0 à 20 euros et la deuxième incluant les coûts restants.

Afin de prendre en compte ces observations, il a été considéré d'ajuster au jeu de données une somme de deux lois Gamma, une pour chacune des zones identifiées, afin de parvenir à un meilleur modèle. Ainsi, le package *flexmix* de R a été utilisé. Il permet d'implémenter les modèles de mélange finis à des fins de régression, dans le cadre des modèles linéaires généralisés. En statistiques, un modèle de mélange est un modèle probabiliste permettant de modéliser différentes sous-populations dans la population globale, sans que ces dernières soient identifiées dans les données par une variable observée. Une fois décomposées, il sera possible de travailler sur chacune des composantes séparément.

La définition formelle qui va suivre d'un modèle de mélange fini est reprise de l'article « *FlexMix : An R Package for Finite Mixture Modelling* », publié par GRUN et LEISCH (2007).

Un modèle de mélange fini est donné par la combinaison convexe de K composantes différentes, c'est-à-dire que les poids qui leur sont associés sont positifs ou nuls et leur somme est égale à 1. On suppose que chacune des sous-populations suit une distribution paramétrique ou qu'elle est donnée par un modèle plus complexe, tel qu'un modèle linéaire généralisé. Dans ce qui suit, on considère des densités de mélange finies $h(\cdot | \cdot)$, avec K composantes, une variable réponse (dépendante) \mathbf{y} et des variables explicatives (indépendantes) \mathbf{x} :

$$h(\mathbf{y} | \mathbf{x}, \mathbf{w}, \Theta) = \sum_{k=1}^K \pi_k(\mathbf{w}, \alpha) f(\mathbf{y} | \mathbf{x}, \vartheta_k), \quad (3.23)$$

où $\forall \mathbf{w}, \alpha$:

$$\pi_k(\mathbf{w}, \alpha) \geq 0 \forall k \quad \wedge \quad \sum_{k=1}^K \pi_k(\mathbf{w}, \alpha) = 1, \quad (3.24)$$

et :

$$\vartheta_k \neq \vartheta_l \quad \forall k \neq l. \quad (3.25)$$

Il est supposé que les distributions des composants $f(\cdot | \cdot)$ appartiennent à la même famille de distributions (exponentielle dans le cadre de ce mémoire) avec des paramètres spécifiques aux composantes ϑ_k . Les poids des composantes ou les probabilités de classe antérieures π_k peuvent éventuellement dépendre de variables concomitantes w et des paramètres α et seront alors modélisés par des modèles logit multinomiaux, comme suggéré par Dayton et Macready (1988).

Il existe plusieurs algorithmes d'estimation de modèles de mélange. Classiquement, le plus utilisé est l'algorithme Espérance-Maximisation (souvent abrégé EM), proposé par Dempster et al. (1977)¹, qui est un algorithme itératif permettant de trouver les paramètres du maximum de vraisemblance d'un modèle probabiliste lorsque ce dernier dépend de variables latentes non observables. C'est ainsi la méthode qui a été retenue lors de l'utilisation du package *flexmix*. Elle se divise en deux étapes : une première étape d'évaluation de l'espérance (E), où l'on calcule l'espérance de la vraisemblance en tenant compte des dernières variables observées ; une deuxième étape de maximisation (M), où l'on estime le maximum de vraisemblance des paramètres en maximisant la vraisemblance trouvée à l'étape E, à partir des modèles individuels représentant chacune des composantes, fournis par l'utilisateur.

Dans le cadre du mémoire, la loi individuelle Gamma a été préféré à la loi individuelle Lognormale car sur la figure 3.3 les QQ plots montrent que les quantiles de la première coïncident le mieux avec les quantiles empiriques (correspondant à la courbe de la fonction identité). La figure 3.4 représente

l'ajustement de la somme de lois Gamma à la distribution empirique des données. Elle permet de constater graphiquement que ce dernier choix de loi de probabilité est le plus adapté aux coûts moyens observés.

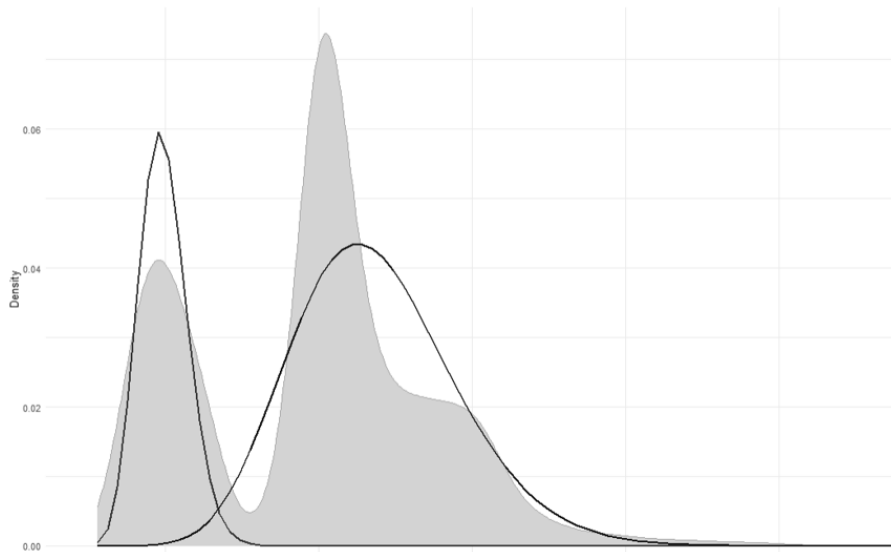


FIGURE 3.4 : Adéquation d'une loi mélange aux données de coût de la téléconsultation

C'est ainsi le modèle mélange de deux lois Gamma qui a été choisi.

Selection de variables

La sélection des variables a été réalisée par la méthode *stepwise* à l'aide de la fonction *stepAIC* du package *MASS* sur R CORE TEAM (2022), où la mesure de sélection retenue est le critère d'information AIC. Étant donné que l'étude porte sur un nombre restreint de variables explicatives, le critère BIC n'a pas été considéré car il pénalise davantage le surparamétrage des modèles.

Il en résulte que le modèle suivant est celui qui minimise le plus l'AIC :

$$\text{Année} + \text{Age} + \text{Affiliation CMU} + \text{Nature de la prestation.} \quad (3.26)$$

C'est ainsi ce modèle qui a été retenu.

Analyse des coefficients

Le tableau 3.5 présente en détail les coefficients obtenus, les p-valeurs du test de Student associées ainsi que leurs niveaux de significativité. Le test de Student permet de vérifier la non-nullité des coefficients obtenus : l'hypothèse n'est pas rejetée si la p-valeur est supérieure à 0,05. Le niveau de significativité est un indicateur permettant d'exposer les variables ayant une forte importance sur le coût de téléconsultation. Plus l'importance est grande et plus le nombre d'étoiles (*) associé sera grand (le maximum de significativité étant représenté par 3 étoiles).

VARIABLE	MODALITE	COEFFICIENT	P-VALEUR	SIGNIFICATIVITE
Intercept	-	3,62	< 2e-16	***
Année	2019	0,01	0,97	
	2020	- 0,02	0,95	
Age	[20, 29]	0,05	5,51e-08	***
	[30, 39]	0,06	6,03e-10	***
	[40, 49]	0,04	2,17e-05	***
	[50, 59]	0,05	5,99e-07	***
	[60, 69]	0,01	0,68	
	[70, 79]	- 0,01	0,27	
	80 +	- 0,02	0,17	
Affiliation CMU	1	- 0,01	0,34	
Nature de prestation	1192	-0,37	< 2e-16	***
	3103	- 1,27	< 2e-16	***
	3104	- 1,1	< 2e-16	***
	3105	- 0,86	2,64e-07	***

FIGURE 3.5 : Coefficients du modèle retenu pour le coût moyen de téléconsultation

L'individu fictif de référence (représenté par l'intercept) possède le profil suivant :

- En 2018, il a eu recours à au moins une prestation de téléconsultation avec un médecin spécialiste (toutes spécialités confondues) ;
- Il a moins de 20 ans ;
- Il ne bénéficie pas de la complémentaire CMU.

Globalement, il est remarqué que la majorité des modalités des variables contenues dans le modèle sont très significatives. Il est notable que les variables nature de la prestation et âge influencent le plus le modèle.

Par ailleurs, lorsqu'un individu possède une caractéristique différente par rapport à l'individu de référence, le coefficient associé à la modalité correspondante est additionné. Les coefficients finaux permettant de prédire le coût moyen sont ainsi obtenus. Cet ajout peut avoir un effet négatif ou positif selon les cas. A titre d'illustration, si un individu possédant les mêmes caractéristiques que l'individu de référence - sauf pour l'âge pour lequel il a entre 30 et 39 ans - est considéré, alors son coût prédit sera supérieur au coût de ce dernier car le coefficient associé à la modalité vaut 0.06.

Les figures suivantes illustrent l'étude des coefficients tarifaires en comparaison des tendances observées lors de l'étape d'analyse descriptive des données, afin d'évaluer leur cohérence.

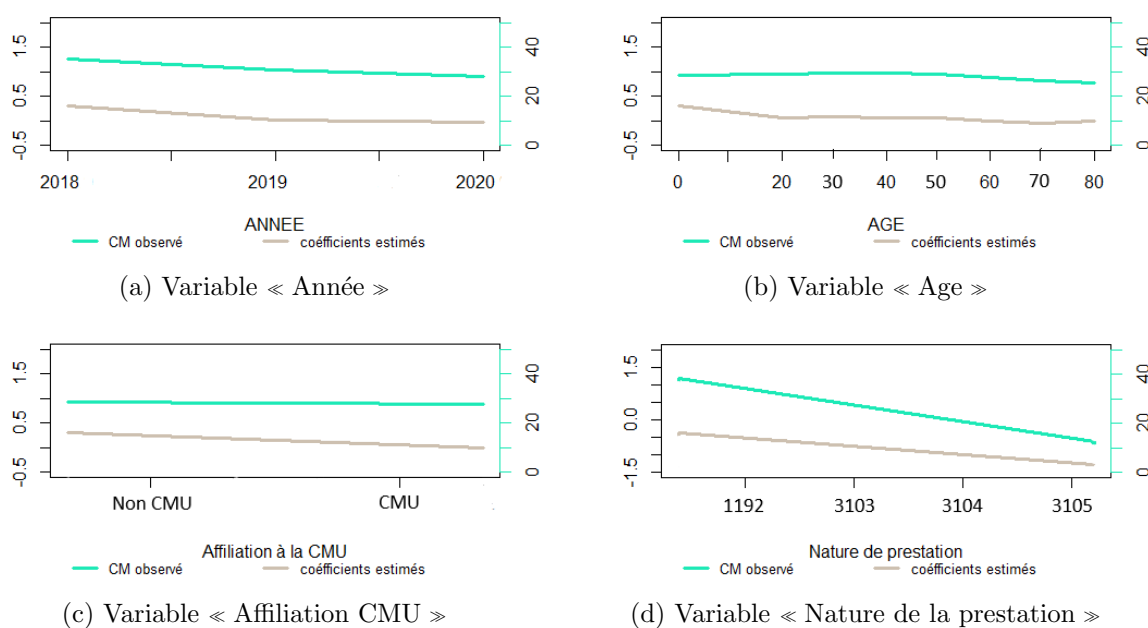


FIGURE 3.6 : Analyse des coefficients tarifaires en fonction des variables explicatives

L'analyse des coefficients associés à ces variables révèle que, globalement, le modèle retenu retranscrit bien les tendances empiriques analysées précédemment.

De plus, quelle que soit la variable explicative considérée, les coûts de téléconsultation restent relativement stables et homogènes. Une tendance décroissante est remarquée pour les coûts de téléconsultation réellement observés selon la nature de la prestation, les téléconsultations généralistes étant les plus

chères et les téléconsultations à domicile les moins chères. Cette tendance a bien été reproduite par les coefficients estimés par le modèle.

Il en est déduit que le modèle est cohérent avec les données de la base du coût de téléconsultation.

Validation du modèle

Afin de valider le modèle, il est possible de vérifier les hypothèses faites sur les résidus d'un GLM, par approche graphique mais aussi par l'usage de tests statistiques.

Postulat 1 : Résidus centrés

Il est observé dans la figure suivante que les résidus, représentés par le nuage de points, sont le plus souvent centrés et alignés autour de la ligne horizontale, sans structure particulière, ainsi le postulat est satisfait.

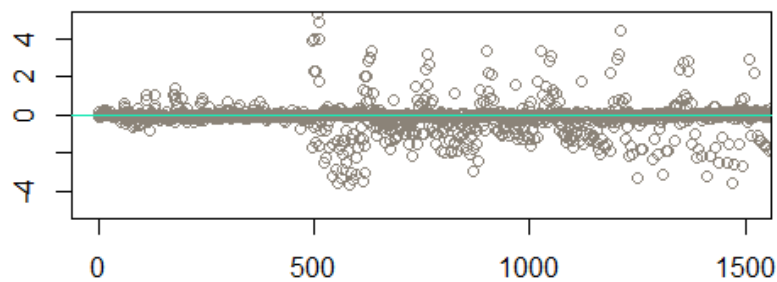
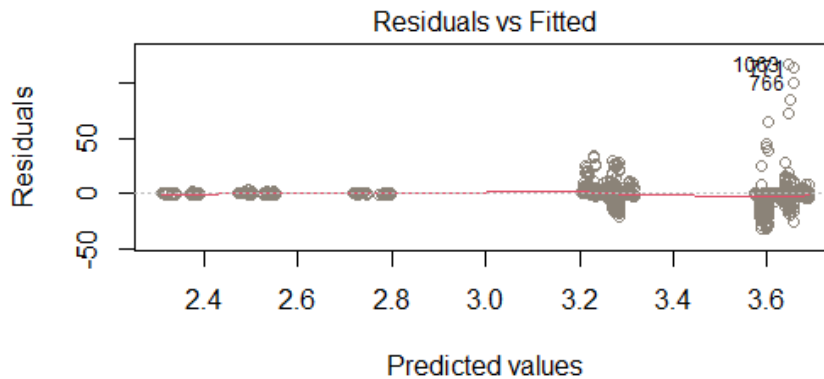


FIGURE 3.7 : Résidus de Pearson - Coût moyen de téléconsultation

Postulat 2 : Résidus homoscedastiques

Ce graphique illustre l'évolution des résidus standardisés $\hat{r}_i = \frac{\hat{e}_i}{se_i}$ (où $se_i = \hat{\sigma} \sqrt{1 - (P_{[X]})_{ii}}$ est l'écart-type estimé des \hat{e}_i) en fonction des valeurs prédites. Le nuage de points est bien centré autour de l'axe des abscisses, néanmoins il existe une grande hétérogénéité parmi les résidus standardisés car il est possible d'identifier quatre nuages de points distincts. Ainsi, le postulat n'est pas vérifié (graphiquement).

FIGURE 3.8 : *Residuals vs Fitted* - Coût moyen de téléconsultation

Postulat 3 : Résidus indépendants

L'autocorrélogramme des résidus du modèle de coût de téléconsultation présente un seul pic dépassant de peu l'intervalle de confiance (représenté par les deux lignes pointillées autour de l'axe des abscisses). Ainsi, il est considéré que graphiquement l'hypothèse d'indépendance est vérifiée.

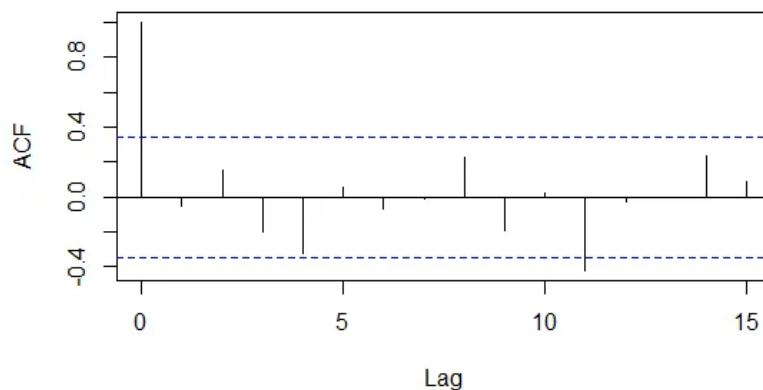


FIGURE 3.9 : Autocorrélogramme des résidus - Coût moyen de téléconsultation

Étude de la qualité de prédiction

Afin d'évaluer la qualité du modèle, il est possible d'analyser la qualité de prédiction du modèle. Un exercice de prédiction a été réalisé sur la base de test, extraite de la base de coût de téléconsultation. La construction de cette base s'inscrit dans le processus de validation croisée des modèles consistant à diviser l'échantillon de données disponibles en deux parties pour calibrer le modèle sur une première partie, dite base d'apprentissage, puis réaliser la prédiction sur la seconde partie appelée base de test.

Dans le cadre de l'étude, un point d'attention a été soulevé quant au fait que la base de données est une base en *model points* et que donc le mode classique de séparation de l'échantillon initial ne peut s'appliquer, au risque d'écarter certains profils d'individus. Une autre méthode, qui réduit ce risque, a alors été retenue : pour chaque mois de modélisation, il a été réalisé une sélection aléatoire de 80% des données pour constituer la base d'apprentissage et une sélection aléatoire de 20% des données, qui

seront affectées à la base de test.

La qualité de cette prédiction a été évaluée selon plusieurs axes :

1. Approche graphique : la représentation graphique des coûts moyens de téléconsultation prédits en fonction des coûts moyens observés illustre deux nuages de points relativement bien centrés autour de la droite identité, sauf pour une petite minorité de points se trouvant plus loin. Il en découle graphiquement que le modèle est de bonne qualité.

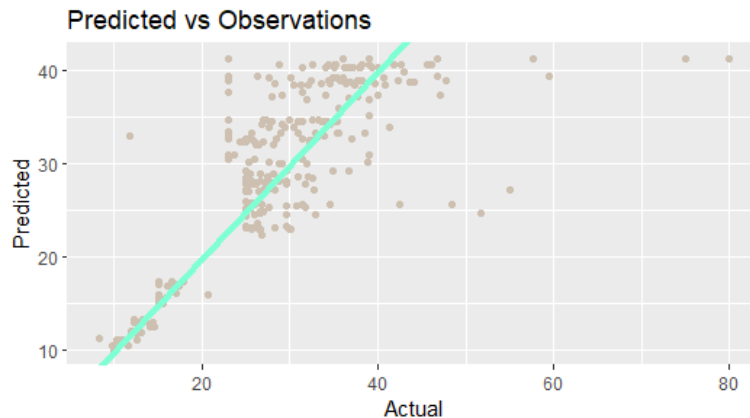


FIGURE 3.10 : Prédications vs Observations - Coût moyen de téléconsultation

2. Approche quantitative : calcul du RMSE.

Comme défini dans les rappels théoriques, le RMSE est l'erreur quadratique des résidus. Elle s'élève à 5,19, ce qui représente 20,87% du coût moyen de téléconsultation, tous profils d'assurés et nature des prestations confondus. Le modèle admet ainsi un biais de prédiction de 20,87%. Il en résulte quantitativement que le modèle est de moyenne qualité.

3.3.2 Modélisation de la fréquence de téléconsultation

Choix d'une loi de probabilité

Comme mentionné auparavant, classiquement la fréquence des sinistres se modélise à l'aide des lois Poisson ou Binomiale négative. Ainsi, de manière analogue aux analyses effectuées sur le modèle de dépense engagée, ces lois ont été testées afin de choisir celle qui s'adapte le mieux à la base de la fréquence de survenance de la téléconsultation. La figure 3.11 présente les résultats de l'estimation par la méthode du maximum de vraisemblance des paramètres des lois Poisson et Binomiale négative s'adaptant davantage aux données.

L'analyse des graphiques révèle que la loi Binomiale négative est la plus adéquate pour la distribution des fréquences moyennes observées. En effet, il est observé que le graphe de sa distribution se rapproche davantage à celui de la distribution empirique étudiées.

C'est ainsi le modèle Binomial négatif qui a été retenu pour la suite des travaux.

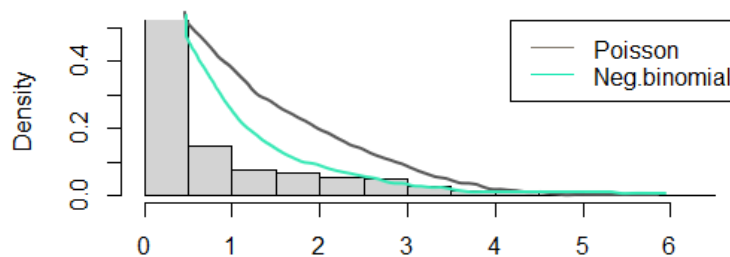


FIGURE 3.11 : Adéquation graphique des modèles de fréquence pour les prestations de téléconsultation

Sélection de variables

Dans le cas présent, la sortie de la fonction *stepAIC* présente le modèle suivant comme étant celui qui minimise le plus l'AIC :

$$\text{Année} + \hat{\text{Âge}} + \text{Affiliation CMU} + \text{Sexe} + \text{Nature de la prestation.} \quad (3.27)$$

C'est ainsi ce modèle qui a été retenu.

Analyse des coefficients

La figure 3.12 présente en détail les coefficients obtenus, les p-valeurs du test de Student associées ainsi que leurs niveaux de significativité. Comme mentionné plus haut, plus l'importance est grande et plus le nombre d'étoiles (*) associé à la modalité de la variable explicative sera grand (le maximum de significativité étant représenté par 3 étoiles).

L'individu fictif de référence (représenté par l'intercept) possède le profil suivant :

- En 2018, il a eu recours à au moins une prestation de téléconsultation avec un médecin spécialiste (toutes spécialités confondues) ;
- Il est de sexe masculin ;
- Il a moins de 20 ans ;
- Il ne bénéficie pas de la complémentaire CMU.

Ce tableau montre que la majorité des modalités ont une grande importance dans l'estimation du modèle. Cela met en évidence la pertinence du modèle et des variables choisies.

VARIABLE	MODALITE	COEFFICIENT	P-VALEUR	SIGNIFICATIVITE
Intercept	-	-10,4	< 2e-16	***
Année	2019	2,18	< 2e-16	***
	2020	7,18	< 2e-16	***
Age	[20, 29]	0,13	0,2	
	[30, 39]	0,18	0,08	*
	[40, 49]	0,04e-1	0,97	
	[50, 59]	- 0,27	0,01	**
	[60, 69]	- 0,41	0,01e-2	***
	[70, 79]	- 0,35	0,01e-1	**
	80 +	0,49	5,29e-06	***
Affiliation CMU	1	0,08	0,18	
Sexe	2	0,37	5,87e-13	***
Nature de prestation	1192	1,17	< 2e-16	***
	3103	- 5,6	< 2e-16	***
	3104	- 4,33	< 2e-16	***
	3105	- 5,72	< 2e-17	***

FIGURE 3.12 : Coefficients du modèle retenu pour la fréquence de téléconsultation

Les figures suivantes illustrent l'étude des coefficients tarifaires en comparaison des tendances observées lors de l'étape d'analyse descriptive des données, afin d'évaluer leur cohérence.

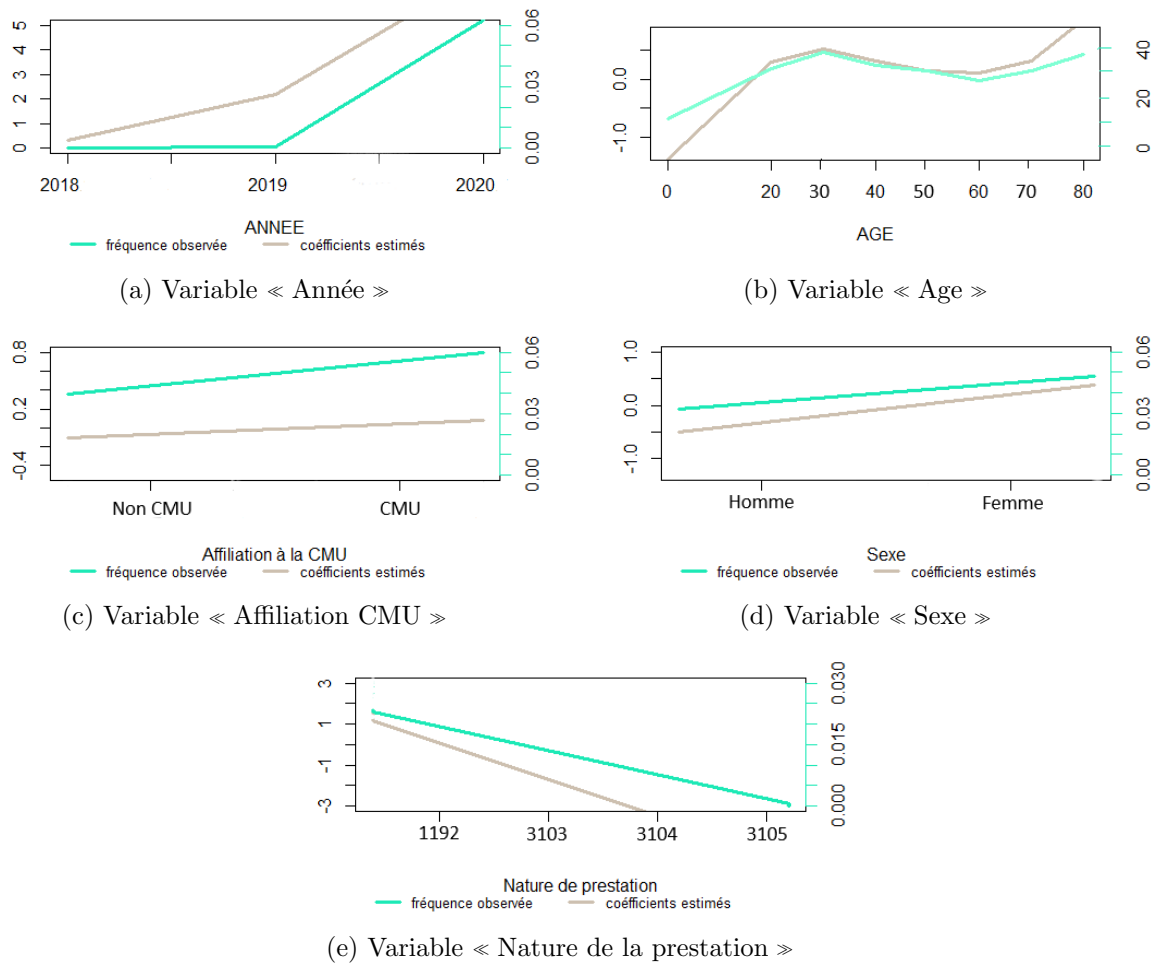


FIGURE 3.13 : Illustration des coefficients tarifaires en fonction des variables explicatives

L'analyse de ces figures confirme la pertinence et la cohérence du modèle. En effet, il est observé que globalement le modèle retenu retranscrit de manière satisfaisante les tendances empiriques analysées lors du Chapitre 2. Il est remarqué que l'âge, l'année et la nature des prestations sont les variables qui influencent le plus la fréquence de survenance de la pratique de téléconsultation. A titre d'illustration, le fait que les assurés sont davantage susceptibles de recourir à la téléconsultation entre 30 et 39 ans et à 80 ans ou plus est bien retranscrit par les coefficients tarifaires du modèle, étant donné qu'à ces âges il est observé des pics de fréquence.

Validation du modèle

Afin de valider le modèle, il est possible de vérifier les hypothèses faites sur les résidus d'un GLM, par un approche graphique mais aussi par l'usage de tests statistiques.

Postulat 1 : Résidus centrés

La figure 3.14 montre un nuages de points aligné et centré (pour la majorité des points) autour de l'axe des abscisses, sans structure particulière. Il est alors considéré que le postulat est satisfait.

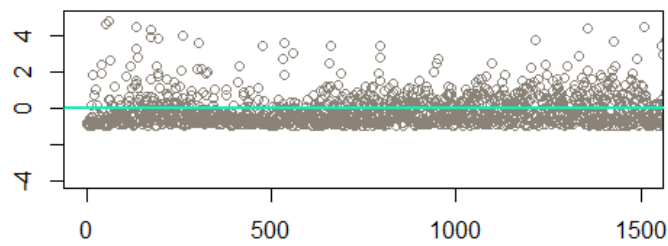


FIGURE 3.14 : Résidus de Pearson - Fréquence de téléconsultation

Postulat 2 : Résidus homoscédastiques

Le graphique suivant illustre également un nuage de points bien centré autour de l'axe des abscisses, sans structure particulière. Ainsi, le postulat est bien vérifié.

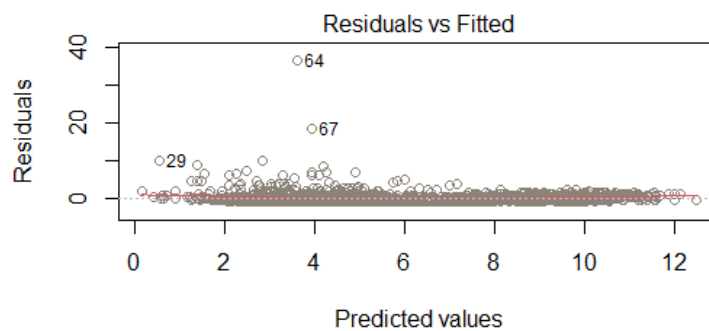


FIGURE 3.15 : *Residuals vs Fitted* - Fréquence de téléconsultation

Postulat 3 : Résidus indépendants

Enfin, l'autocorréogramme des résidus du modèle de coût de téléconsultation présente plusieurs pics dépassant l'intervalle de confiance (représenté par les deux lignes pointillées autour de la droite $y = 0$). Ainsi, il est considéré que graphiquement l'hypothèse d'indépendance est rejetée.

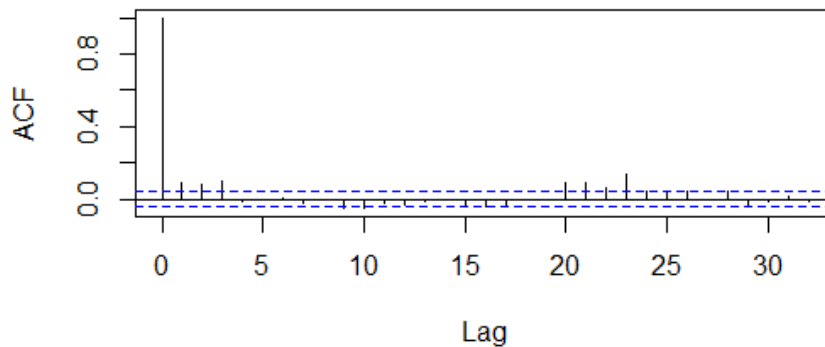


FIGURE 3.16 : Autocorrélogramme des résidus - Fréquence de téléconsultation

Étude de la qualité de prédiction

Afin de valider le modèle, les méthodes utilisées pour le modèle de coût de téléconsultation (validation croisée, approche graphique/ quantitative) ont été exploitées pour le modèle de fréquence. Les résultats sont détaillés ci-dessous.

La qualité de cette prédiction a été évaluée selon plusieurs axes :

1. Approche graphique : la représentation graphique des coûts moyens de téléconsultation prédits en fonction des coûts moyens observés illustre un nuage de points relativement bien centré autour de la droite identité, sauf pour une petite minorité de points se trouvant plus loin. Il en découle graphiquement que le modèle est de bonne qualité.

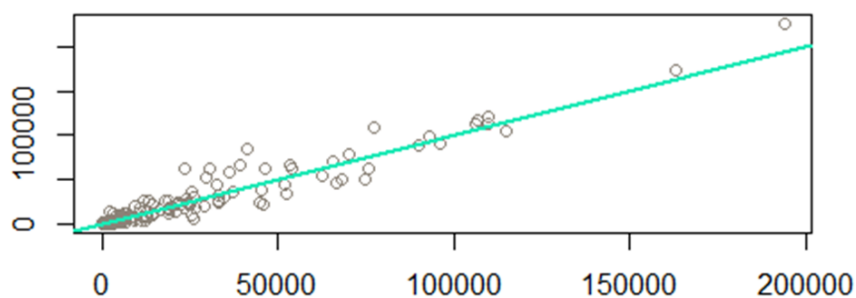


FIGURE 3.17 : Prédiction vs Observations - Fréquence de téléconsultation

2. Approche quantitative : calcul du RMSE.

Il s'élève à 1666,34, ce qui représente 19,60% de la quantité moyenne de téléconsultation, tous profils d'assuré et nature de la prestation confondus. Le modèle admet ainsi un biais de prédiction de 19,60%. Il en résulte quantitativement que le modèle est de moyenne qualité.

3.3.3 Résultats obtenus des modèles tarifaires de consultation et de téléconsultation

Ainsi, suite aux modélisations GLM du coût moyen et de la fréquence de téléconsultation et de consultation (en suivant le principe de tarification « Coût-Fréquence », rappelé lors la première section du chapitre), il a été possible de déterminer les primes pures des prestations de téléconsultation. Celles-ci sont présentées par année de remboursement et par nature de la prestation considérée.

Année	Poste de soins	Type de prestation	Fréquence moyenne	Coût moyen (en euros)	Sinistralité moyenne par prestation (en euros)	Sinistralité moyenne globale (en euros)
2018	Consultation traditionnelle	Consultation toutes spécialités	0,08	36,36	2,77	17,47
		Consultation généraliste	0,57	26,01	14,70	
	Téléconsultation	Téléconsultation toutes spécialités	0,53e-4	37,29	0,19e-2	0,25e-2
		Téléconsultation généraliste	0,2e-4	26,81	0,54e-3	
2019	Consultation traditionnelle	Consultation toutes spécialités	0,24	36,36	8,61	48,99
		Consultation généraliste	1,55	26,08	40,37	
	Téléconsultation	Téléconsultation toutes spécialités	0,94e-3	36,41	0,03	0,08
		Téléconsultation généraliste	0,18e-2	26,18	0,05	
2020	Consultation traditionnelle	Consultation toutes spécialités	0,19	36,36	7,22	43,17
		Consultation généraliste	1,33	26,08	34,47	
		Consultation lieu dédié et urgences	0,02	92,07	1,41	
		Consultation à domicile	0,32e-2	23,52	0,07	
	Téléconsultation	Téléconsultation toutes spécialités	0,08	35,12	2,64	9,34
		Téléconsultation généraliste	0,26	25,25	6,67	
		Téléconsultation lieu dédié et urgences	0,99e-3	23,83	0,02	
		Téléconsultation à domicile	0,18e-3	16,53	0,29e-2	

FIGURE 3.18 : Primes pures par année de remboursement et nature de la prestation

Il en résulte qu'au fil des années l'écart de sinistralité entre la sinistralité des pratiques médicales de consultation et de téléconsultation se comble (car la différence entre les fréquences moyennes de consultation et de téléconsultation s'est amoindrie). Ce phénomène est principalement dû à l'accélération du développement de la fréquence de recours aux prestations de téléconsultation après apparition de la Covid-19.

A titre d'illustration, il est observé que la fréquence moyenne de recours à une téléconsultation spécialiste (toutes spécialités confondues) a été multipliée par 10 entre 2018 et 2019 et par 100 entre 2019 et 2020 (approximativement), en atteignant sensiblement les mêmes ordres de grandeur des fréquences moyennes des prestations de consultation spécialiste (0,08 pour les téléconsultations toutes spécialités contre 0,19 pour les consultations toutes spécialités).

Ces évolutions représentent toutefois pour l'assureur de l'incertitude quant aux engagements qu'il porte et soulèvent ainsi le besoin d'en étudier l'étendue et les conséquences futures.

Chapitre 4

Analyse de l'évolution temporelle de la sinistralité en téléconsultation

Ce chapitre sera consacré à l'étude de l'évolution temporelle de la sinistralité en consultation et en téléconsultation, ainsi qu'à l'analyse de sa projection dans le futur.

Pour ce faire, dans un premier temps, les fréquences du recours à la consultation classique et à la téléconsultation seront modélisées par séries temporelles multivariées. Dans un deuxième temps des scénarios d'évolution de la consommation seront calibrés à partir des modèles GLM de tarification et les modèles de séries temporelles retenus lors des travaux.

4.1 Modélisation à l'aide des séries temporelles

L'objectif de cette section est de montrer l'existence d'un phénomène de « vases communicants » entre les évolutions de consommation des deux pratiques médicales en France, actuelles et futures, à l'aide des séries temporelles. L'étude se focalise sur les fréquences d'utilisation car, comme observé dans les parties précédentes, les écarts entre les coûts moyens de l'ensemble des prestations de consultation et de téléconsultation sont négligeables comparés aux écarts de fréquence.

Une série temporelle, ou chronologique, $(X_t, t \in \Theta)$ correspond à une suite d'observations réalisées au cours du temps Θ représentant l'évolution d'un phénomène spécifique. Historiquement, elles ont été utilisées en astronomie, en économétrie, en démographie ou encore en médecine. L'étude d'une série chronologique permet d'analyser, de décrire et d'expliquer un phénomène au cours du temps et d'en tirer des conséquences. Néanmoins, l'objectif principal de l'analyse d'une série temporelle est la prévision de l'évolution de ce phénomène dans le futur, qui consiste à prévoir les valeurs futures X_{T+h} ($h = 1, 2, 3, \dots$) à partir des données historiques observées jusqu'au temps T : X_1, X_2, \dots, X_T . On peut alors définir l'erreur de prédiction par la différence « en moyenne » entre la valeur prédite au temps $T + h$, notée $\hat{X}_T(h)$ et la valeur réelle X_{T+h} .

Une autre application intéressante des séries temporelles est la détection de ruptures résultantes au cours du temps. Ces ruptures peuvent être de deux ordres : une rupture de pente ou une rupture de niveau. Dans le cadre de notre étude, un point d'attention se soulève concernant l'éventuelle (et espérée) apparition d'une rupture résultante de l'épidémie du Coronavirus.

Enfin, l'utilisation des séries temporelles multivariées en particulier (qui correspondent à l'ob-

servation simultanée de plusieurs séries temporelles univariées à valeurs réelles) est pertinente dans le contexte de l'étude car elle permet de mettre en évidence les effets de corrélation et de causalité entre différentes variables. Ici, la modélisation simultanée des séries univariées des fréquences de téléconsultation et de consultation a été réalisée à l'aide du modèle vectoriel autorégressif (VAR), qui sera défini et développé dans la suite. Il sera alors intéressant de savoir si les valeurs prises par une variable (X_t) (fréquence de téléconsultation) sont la conséquence des valeurs prises par la variable (Y_t) (fréquence de consultation) ou inversement et de regarder les phénomènes d'anticipation entre ces deux variables.

Afin de résoudre ces problématiques, une étude exhaustive des deux séries temporelles univariées a été exécutée.

Comme mentionné ci-dessus, l'ambition principale de l'analyse d'une série temporelle est de projeter la série dans le futur. Pour y parvenir, il est nécessaire de modéliser le mécanisme de production de la série chronologique. Étant donné que souvent les variables X_t ne sont ni indépendantes ni identiquement distribuées alors que les méthodes statistiques classiques d'étude de séries temporelles reposent sur ces hypothèses, il devient nécessaire de traiter les données au préalable. Les étapes classiques du traitement statistique d'une série chronologique sont les suivantes :

- Correction des données : cette étape comporte notamment l'évaluation des données manquantes et le remplacement de données accidentelles. Cela a déjà détaillé au Chapitre 2 ;
- Analyse qualitative de la série et analyse des caractéristiques fondamentales de la série ;
- Choix et ajustement des modèles ;
- Projections du modèle.

4.1.1 Analyse des caractéristiques fondamentales de la série

Une série temporelle ($X_t, t \in \Theta$) est souvent décomposée soit de manière additive soit multiplicative :

$$X_t = M_t + S_t + Z_t \text{ ou } X_t = M_t \cdot S_t \cdot (1 + Z_t), \quad (4.1)$$

avec :

- $t \mapsto M_t$: la tendance ou *trend* représente l'évolution à long terme de la série étudiée (croissance ou décroissance linéaire, quadratique...). Elle traduit le comportement « moyen » de la série. Cette composante peut aussi avoir une expression différente pour des périodes distinctes (affine par morceaux par exemple). Plus généralement, il est possible d'apparenter cette composante à une fonction lisse du temps t ;
- $t \mapsto S_t$: la composante saisonnière ou saisonnalité correspond à un phénomène qui se répète à intervalles de temps réguliers (périodiques, par exemple de période 12 de la consommation d'électricité, de période 4 pour les séries trimestrielles, etc.) ;
- $t \mapsto Z_t$: la composante d'erreur ou résidu, aléatoire et idéalement de faible variabilité par rapport aux deux autres. Elle représente une composante irrégulière et constitue un point d'attention car en pratique ce terme d'erreur sera le plus souvent autocorrélé (c'est à dire que la covariance entre Z_t et Z_{t+h} sera non nulle).

Analyse descriptive

En règle générale, l'étape initiale en statistique descriptive est d'analyser graphiquement la série temporelle avant d'effectuer le moindre calcul. Les graphiques des séries corrigées et prétraitées permettent de visualiser de nombreuses caractéristiques des données, notamment les tendances, les composantes saisonnières, ou encore des observations anormales et les phénomènes accidentels des séries. Les attributs mis en exergue par l'approche qualitative devront ensuite être pris en compte, autant que possible, dans le choix et l'exécution des méthodes de prévision utilisées.

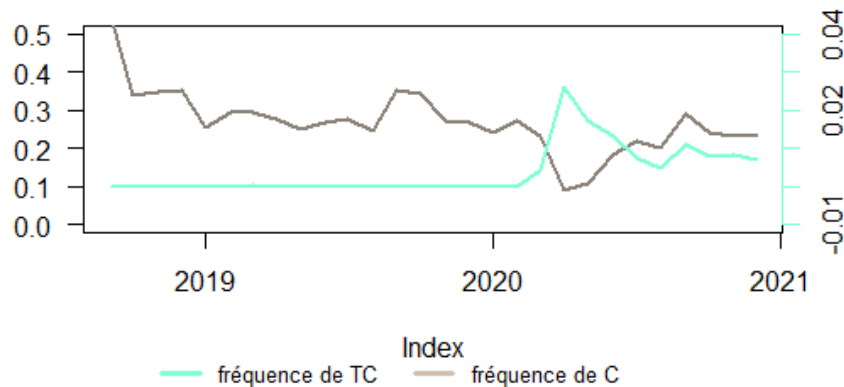


FIGURE 4.1 : Représentation graphique des fréquences de téléconsultation (TC) et de consultation (C) en fonction du temps

- Tout d'abord, il est retrouvé clairement le phénomène de « vases communicants » entre la fréquence d'utilisation de téléconsultation et celle de consultation en début d'année 2020, correspondant à la période d'apparition de la Covid-19. Bien que ce soit à des échelles différentes, entre février et mars 2020 la fréquence de téléconsultation a doublé de volume (en passant de 10^{-4} à 10^{-2} en ordre de grandeur), tandis que la fréquence de consultation a chuté (en passant de 10^{-1} à 10^{-2} en ordre de grandeur). Cette observation nous indique de l'existence éventuelle d'une corrélation et d'une causalité entre les deux séries temporelles.
- Ensuite le graphique montre que la deuxième moitié de 2020 a connu une faible diminution suivie par une augmentation des deux séries, simultanément. Il peut être supposé que ces fluctuations ont été entraînées par la phase de déconfinement et confinement qui ont eu lieu respectivement à la mi-mai 2020 et en début novembre 2020. Ces changements soulèvent un point d'attention quant à l'instabilité du développement futur de la pratique de téléconsultation, sachant que depuis le 1er août 2022 la France n'est plus en état d'urgence sanitaire.
- Enfin, il est constaté que les deux séries ne semblent pas inclure de composante saisonnière. En effet, on n'observe pas plus que 2 périodes, ce qui est d'ailleurs confirmé par la fonction *decompose()* sur R.

Estimation de la tendance

Il existe différents procédés permettant d'analyser la tendance d'une série chronologique. Dans ce projet, la méthode du lissage par moyennes mobiles a été employée.

Une moyenne mobile est une transformation de X_t s'écrivant comme combinaison linéaire finie des valeurs de la série correspondant à des dates entourant t . Le calcul de la moyenne mobile dépend d'un paramètre l appelé la largeur de fenêtre, qui correspond au nombre d'observations prises en compte dans le calcul. Plus l est grand plus le lissage est important, jusqu'à atteindre la fonction constante.

Le but de ce type de lissage est de faire apparaître l'allure de la tendance, et ainsi de faire disparaître la saisonnalité et de réduire au maximum l'erreur. Les figures suivantes représentent les courbes lissées des séries des fréquences de recours à la téléconsultation et à la consultation obtenues au moyen de cette méthode.

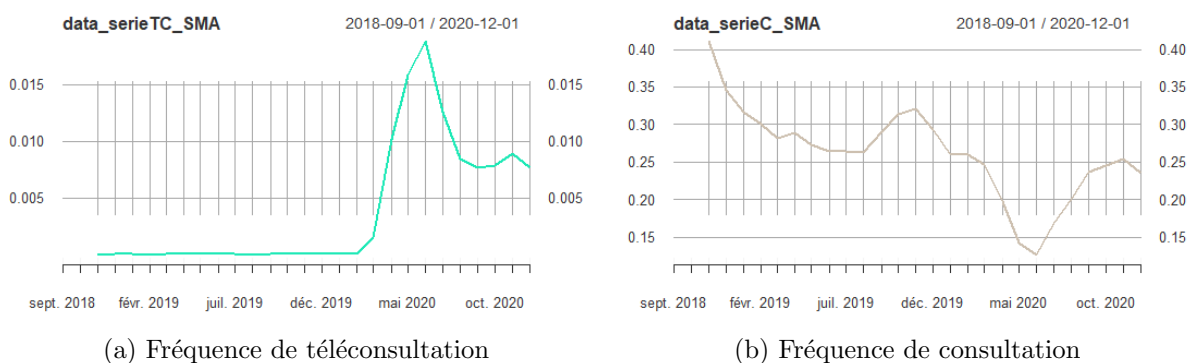


FIGURE 4.2 : Graphiques des séries lissées

Elles donnent une image plus claire de la tendance. En effet, elles mettent en exergue les courbes de tendance polynomiale caractérisant l'évolution des fréquences de téléconsultation (d'ordre 3) et de consultation (d'ordre 5). Il est également noté que la baisse de consommation des prestations de consultation ne date pas seulement de la Covid-19. En réalité, la fréquence de consultation diminue (globalement) depuis septembre 2018 et en particulier la fréquence de consultations généralistes par habitant baisse depuis 2000, d'après la SNIR (tableau suivant).

1980	1985	1990	1995	2000	2005	2009	2013
3,3	3,9	4,1	4,5	4,8	4,6	4,3	4,1

FIGURE 4.3 : Nombre d'actes de généralistes par habitant (France métropolitaine)

La quantité d'actes pourrait être considéré comme une mesure inadaptée de l'intensité de la prise en charge par les généralistes : des consultations/visites plus longues pourraient expliquer que leur nombre se réduise. A cet égard, l'étude réalisée par BREUIL-GENIER et GOFFETTE (2006) aboutit à un temps de consultation de 15 minutes tandis que l'étude réalisée par JAKOUBOVITCH et al. (2012) conclut à une durée d'environ 18 minutes. Néanmoins ces durées n'ont pas été mesurées dans des conditions homogènes, ainsi il est difficile de tirer de ces données des conclusions robustes.

Stationnarité

La modélisation du comportement aléatoire d'une série temporelle est plus aisée lorsque certains aspects de la loi de $(X_t, t \in \Theta)$ ne changent pas lorsque le temps t varie. On parle alors de séries ayant certaines propriétés de stationnarité. Une série stationnaire est une série dont les propriétés statistiques ne dépendent pas du temps. Ainsi, les séries chronologiques avec des tendances, ou avec une saisonnalité, ne sont pas stationnaires.

Plus formellement, il existe deux types de stationnarité : celle au sens strict, celle au sens faible.

- Un processus $(X_t, t \in \Theta)$ est dit stationnaire au sens strict si les lois des vecteurs finis dimensionnels $(X_{t_1}, X_{t_2}, \dots, X_{t_k})$ et $(X_{t_1+h}, X_{t_2+h}, \dots, X_{t_k+h})$ coïncident $\forall k \in \mathbb{N}, \forall t_1, \dots, t_k \in \Theta$ et $\forall h \in \mathbb{N}$.
- Un processus $(X_t, t \in \Theta)$ est dit stationnaire au sens faible si et seulement si :
 - Son espérance est constante au cours du temps, c'est-à-dire, $\mathbb{E}[X_t] = \mu, \forall t \in \Theta$;
 - Sa fonction d'autocovariance dépend seulement de l'écart de temps $t_2 - t_1$, c'est-à-dire, $Cov(X_{t_1}, X_{t_2}) = Cov(X_{t_1+c}, X_{t_2+c}) = \gamma(t_2 - t_1)$ avec c , une constante quelconque.
 Dans ce cas, $\mathbb{E}[X_t^2] = \mu^2 + \gamma(0)$ est une constante, $\forall t \in \Theta$.

Pour déterminer si les séries étudiées sont stationnaires ou non, le test augmenté de Dickey-Fuller (ADF) a été implémenté. Ce test est fondé sur l'estimation par les Moindres Carrés Ordinaires (MCO) de trois modèles autorégressifs du premier ordre dont les erreurs sont identiquement et indépendamment distribuées : le modèle sans constante, le modèle avec constante et le modèle avec constante et tendance. Le test ADF teste l'hypothèse nulle de la présence d'une racine unité dans un modèle de série chronologique autorégressive ce qui se traduit par la non stationnarité de la série.

La fonction `adf.test()` sur R permet de déterminer les p-valeurs du test pour les deux séries temporelle : p-valeur = 0.2805 pour la série de la fréquence de consultation et p-valeur = 0.4416 pour la série de la fréquence de téléconsultation. Or il est considéré que la valeur critique du test est 5%. Ainsi, puisque les p-valeurs sont supérieures à 5%, l'hypothèse nulle est acceptée i.e. les deux séries ne sont pas stationnaires.

Autocovariance

La fonction d'autocovariance d'une série temporelle permet de calculer et de représenter graphiquement la corrélation existant entre les observations d'une série chronologique. En effet, cette notion fait référence au fait que dans une série temporelle la mesure d'un phénomène à un instant t peut être corrélée aux mesures précédentes (aux temps $t - 1, t - 2$, etc.) ou aux mesures suivantes (à $t + 1, t + 2, \dots$). Une série autocorrélée est ainsi corrélée à elle-même, avec un décalage (ou *lag*) donné. Les deux droites horizontales pointillées sur les autocorrélogrammes (ACF) indiquent le seuil critique au-delà duquel l'autocorrélation est considérée significative.

Mathématiquement, on définit l'autocovariance pour une variable X_t de moyenne μ et de variance σ par :

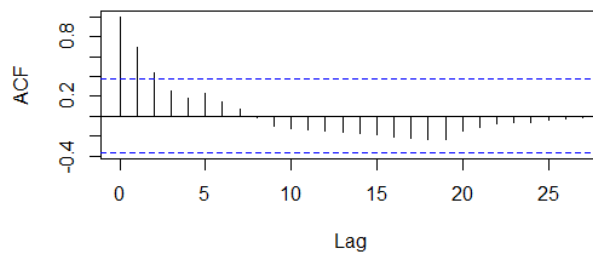
$$\gamma_k = \mathbb{E}((X_t - \mu)(X_{t+k} - \mu)). \quad (4.2)$$

L'autocorrélation se définit à partir de l'autocovariance :

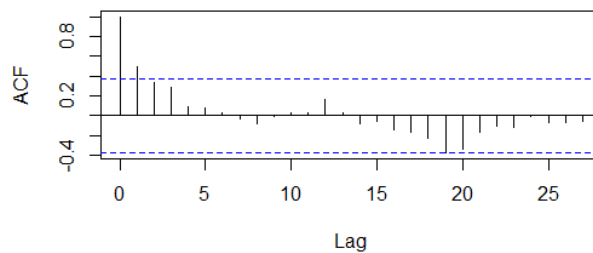
$$\rho = \frac{\gamma_k}{\sigma^2}. \quad (4.3)$$

L'autocorrélogramme d'une série temporelle peut aider à détecter une tendance et/ou une saisonnalité de celle-ci. En effet, l'ACF d'une série avec tendance est enclin à avoir des valeurs positives qui diminuent lentement à mesure que les décalages augmentent. De plus, lorsque les données sont saisonnières, les autocorrélations seront plus importantes pour les décalages saisonniers, c'est-à-dire pour des multiples de la fréquence périodique, que pour les autres décalages.

Grâce à la figure 4.4, on observe aisément que les valeurs de fréquence des prestations de téléconsultation et de consultation connaissent globalement une diminution en même temps que les décalages augmentent. Toutefois, on ne constate pas de composante saisonnière. On confirme ainsi les conclusions tirées des estimations précédentes, ainsi que la non stationnarité des séries analysées.



(a) Fréquence de téléconsultation



(b) Fréquence de consultation

FIGURE 4.4 : Autocorrélogrammes des séries étudiées

Analyse spectrale

Toute série temporelle peut être exprimée comme une combinaison d'ondes cosinusoïdales et sinusoidales avec des périodes (temps nécessaire pour compléter un cycle complet) et des amplitudes (valeur maximale/minimale pendant le cycle) différentes.

Ainsi un périodogramme peut être utilisé pour examiner le comportement saisonnier d'une série chronologique. L'analyse spectrale est l'analyse des fréquences dominantes d'une série chronologique. En pratique, l'analyse spectrale impose des techniques de lissage au périodogramme.

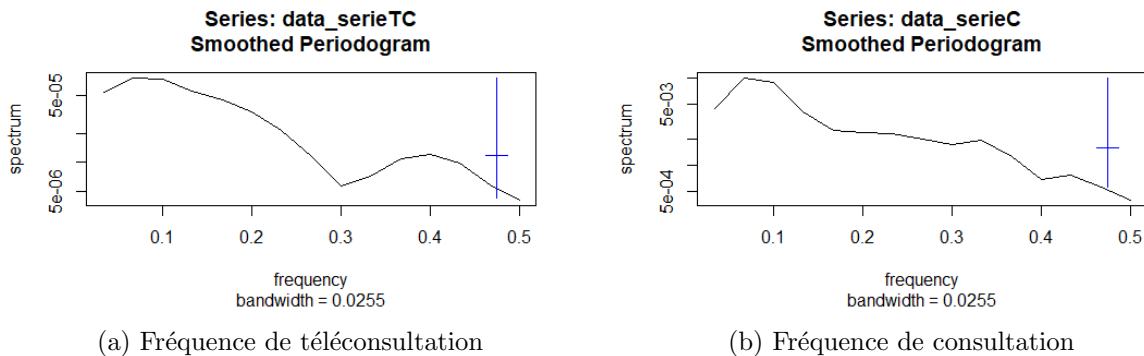


FIGURE 4.5 : Périodogrammes lissés des séries étudiées

A cet effet, la figure 4.5 représente les périodogrammes lissés des séries des fréquences de téléconsultation et de consultation. Ces derniers permettent de constater plus facilement la présence d'un cycle saisonnier incomplet pour la fréquence de consultation, ainsi que la présence d'une période d'un pic qui se répète seulement 2 fois. Les données ne sont pas suffisamment longues pour couvrir au moins deux cycles saisonniers, ce qui ne permet pas de décomposer et donc d'estimer la saisonnalité des deux séries temporelles. Il est considéré alors que les séries ne possèdent pas de composantes saisonnières.

4.1.2 Choix et ajustement des modèles

Comme mentionné lors du préambule de cette section, les modèles retenus sont les modèles VAR.

Le modèle vectoriel autorégressif est l'un des modèles les plus efficaces, flexibles et simples d'utilisation pour l'analyse des séries chronologiques multivariées. Il s'agit d'une extension naturelle du modèle autorégressif univarié aux séries chronologiques dynamiques multivariées. Le modèle VAR s'est avéré particulièrement utile pour décrire le comportement dynamique des séries chronologiques économiques et financières et pour la prévision. Il fournit souvent des prévisions supérieures à celles des modèles de séries temporelles univariées et des modèles élaborés d'équations simultanées basés sur la théorie d'équations simultanées, car elles sont conditionnelles aux trajectoires futures potentielles des variables spécifiées dans le modèle.

Outre la description des données et la prévision, le modèle VAR est également utilisé pour l'inférence structurelle, notamment pour l'étude de la structure causale des variables retenues dans la modèle, pertinente dans le cadre des problématiques soulevées précédemment.

Cette section se concentre sur l'analyse des séries temporelles multivariées stationnaires par covariance à l'aide de modèles VAR.

Représentation VAR stationnaire

On note $\mathbf{Y}_t = (y_{1t}, y_{2t}, \dots, y_{nt})'$ un vecteur de séries temporelles de dimension $(n \times 1)$. Classiquement, le modèle d'un vecteur autorégressif à p décalages (VAR(p)) a pour forme :

$$\mathbf{Y}_t = \mathbf{c} + \Pi_1 \mathbf{Y}_{t-1} + \Pi_2 \mathbf{Y}_{t-2} + \dots + \Pi_p \mathbf{Y}_{t-p} + \varepsilon_t, \quad t = 1, \dots, T. \quad (4.4)$$

Il est considéré que :

- $\Pi_i, i = 1, \dots, p$ est la matrice $(n \times n)$ des coefficients du modèle,
- $\forall t, \varepsilon_t$ est un bruit blanc aléatoire, centré, de variance constante et de matrice de variance-covariance Σ invariante dans le temps.

A titre d'illustration, le système d'équations suivant correspond à un modèle VAR(2) bivarié :

$$\begin{aligned} y_{1t} &= c_1 + \pi_{11}^1 y_{1t-1} + \pi_{12}^1 y_{2t-1} + \pi_{11}^2 y_{1t-2} + \pi_{12}^2 y_{2t-2} + \varepsilon_{1t}; \\ y_{2t} &= c_2 + \pi_{21}^1 y_{1t-1} + \pi_{22}^1 y_{2t-1} + \pi_{21}^2 y_{1t-2} + \pi_{22}^2 y_{2t-2} + \varepsilon_{2t}; \end{aligned} \quad (4.5)$$

avec :

$$\begin{aligned} \text{Cov}(\varepsilon_{1t}, \varepsilon_{2s}) &= \sigma_{12} \quad \text{si } t = s \\ \text{Cov}(\varepsilon_{1t}, \varepsilon_{2s}) &= 0 \quad \text{sinon.} \end{aligned} \quad (4.6)$$

À la lecture de ce modèle, il apparaît qu'il n'est pas sous forme réduite : en effet, y_{1t} a un effet immédiat sur y_{2t} et réciproquement y_{2t} a un effet immédiat sur y_{1t} . Ce système initial est appelé la forme structurelle de la représentation VAR.

Différenciation

La modélisation VAR repose sur la stationnarité des variables constituant le modèle. Ainsi, les paramètres du processus VAR ne pourront pas être estimés sur les séries des fréquences de téléconsultation et de consultation puisqu'il a été montré qu'elles ne sont pas stationnaires.

Toutefois, il est possible de rendre stationnaires ces séries par différenciation. La différenciation aide à stabiliser la moyenne d'une série temporelle en éliminant les changements de niveau d'une série temporelle, et donc en éliminant (ou en réduisant) la tendance et la saisonnalité. C'est un passage obligatoire pour pouvoir ajuster le modèle VAR(p) à des séries non stationnaires. Soit une série temporelle (X_t) , on définit l'opérateur de différence par $\Delta X_t = X_t - X_{t-1}$.

Chacune des séries étudiées a dû être différenciée $d = 2$ fois afin que l'hypothèse nulle de présence d'une racine unité puisse être rejetée dans le test augmenté de Dickey-Fuller (p-valeur = 0,019 pour la fréquence de consultation et p-valeur = 0,022 pour la fréquence de téléconsultation, les deux étant inférieures à la valeur critique 0,05). A cet effet, les fonctions `diff()` et `adf.test()` ont été utilisées sur R. Les figures suivantes présentent les séries étudiées et leur version différenciée.

La figure 4.6 met en exergue la transformation des séries par processus de différenciation. Il est remarqué que les deux séries différenciées sont centrées autour de l'axe des abscisses, caractéristique classique des processus stationnaires.

La stationnarité par différenciation est donc vérifiée, il est alors possible de procéder à l'estimation des paramètres.

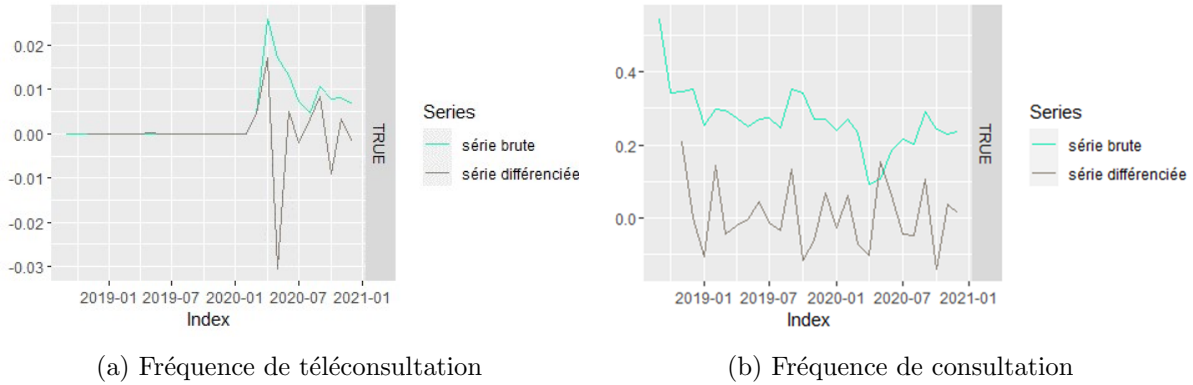


FIGURE 4.6 : Les évolutions comparées des séries temporelles brutes et leurs formes différenciées

Estimation des paramètres du modèle

Dans le cas d'un processus VAR, chacune des équations du modèle peut être estimée par les MCO, indépendamment les unes des autres (ou par une méthode de maximum de vraisemblance).

Alors le modèle VAR(p) estimé est de la forme :

$$Y_t = \hat{c} + \hat{\Pi}_1 Y_{t-1} + \hat{\Pi}_2 Y_{t-2} + \dots + \hat{\Pi}_p Y_{t-p} + e, \quad (4.7)$$

e étant le vecteur de dimension $(k, 1)$ des résidus d'estimation $e_{1t}, e_{2t}, \dots, e_{kt}$. De plus, on note : \sum_e la matrice des variances covariances estimées des résidus du modèle.

Dans le cas de la représentation VAR, les critères d'Akaike et de Schwarz peuvent être utilisés pour déterminer l'ordre p des décalages du modèle. La procédure de sélection de l'ordre de la représentation consiste à estimer tous les modèles VAR pour un ordre allant de 0 à h (h étant le retard maximum admissible admis par les données étudiées). Les fonctions $AIC(p)$ et $SC(p)$ sont calculées de la manière suivante :

$$\begin{aligned} AIC(p) &= \text{Ln} \left[\det \left| \sum_e \right| \right] + \frac{2k^2 p}{n}, \\ SC(p) &= \text{Ln} \left[\det \left| \sum_e \right| \right] + \frac{k^2 p \text{Ln}(n)}{n}, \end{aligned} \quad (4.8)$$

avec : k = nombre de variables du système ; n = nombre d'observations ; p = nombre de retards ; \sum_e = matrice des variances covariances des résidus du modèle.

La fonction `VARselect()` sur R retourne le retard p qui minimise les critères AIC et SC directement pour la série temporelle multivariée Y_t correspondant au vecteur des séries de fréquence des deux pratiques médicales étudiées. La valeur optimale de l'ordre p est alors égale à 7.

Ainsi, la modélisation VAR a bien pu être ajustée aux données à l'aide de la fonction `VAR`, avec comme instructions en entrée : l'ordre $p = 7$ et la valeur initiale de type constante.

4.1.3 Projection des séries temporelles

Les coefficients du modèle étant estimés, la prévision peut être calculée en n à l'horizon d'une période, par exemple pour un VAR(1), de la manière suivante :

$$\hat{Y}_n(1) = \hat{\Pi}_0 + \hat{\Pi}_1 Y_n. \quad (4.9)$$

À l'horizon de 2 périodes, la prévision est :

$$\hat{Y}_n(2) = \hat{\Pi}_0 + \hat{\Pi}_1 \hat{Y}_n(1) = \hat{\Pi}_0 + \hat{\Pi}_1 \hat{\Pi}_0 + \hat{\Pi}_1^2 Y_n. \quad (4.10)$$

À l'horizon de 3 périodes, la prévision s'écrit :

$$\hat{Y}_n(3) = \hat{\Pi}_0 + \hat{\Pi}_1 \hat{Y}_n(2) = \left(I + \hat{\Pi}_1 + \hat{\Pi}_1^2 \right) \hat{\Pi}_0 + \hat{\Pi}_1^3 Y_n. \quad (4.11)$$

et ainsi de suite...

Lorsque h tend vers l'infini, il est constaté que la projection tend vers une valeur constante (état stationnaire) puisque $\hat{\Pi}_1^i \rightarrow 0$ si $i \rightarrow \infty$.

L'espérance de la matrice de l'erreur de prévision est nulle, sa variance est donnée par :

$$\sum_e(h) = M_0 \sum_e M'_0 + M_1 \sum_e M'_1 + \dots + M_{h-1} \sum_e M'_{h-1}, \quad (4.12)$$

où M_i est calculé par la formule de récurrence suivante :

$$M_i = \sum_{j=1}^{\min(p,i)} \hat{\Pi}_j M_{i-j} \quad i = 1, 2, \dots \text{ et } M_0 = I. \quad (4.13)$$

La variance de l'erreur de prévision pour chacune des prévisions des k variables ($\hat{\sigma}_n^2(h)$) se lit sur la première diagonale de la matrice $\sum_e(h)$. L'intervalle de prévision au seuil de (1- (2)) est donné par : $\hat{Y}_n(h) \pm t^{\alpha/2} \times \hat{\sigma}_n(h)$ avec $t^{\alpha/2}$ valeur de la loi normale.

La figure suivante représente les projections réalisées à l'aide du modèle $VAR(7)$ sur les séries des fréquences du recours aux prestations de téléconsultation et de consultation.

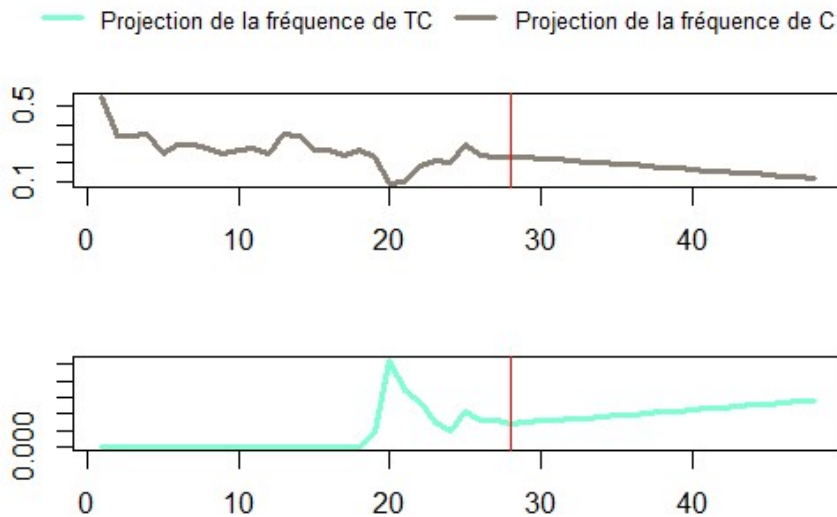


FIGURE 4.7 : Projections dans le futur des fréquences de survenance de la consultation et de la téléconsultation

Ces figures mettent en évidence la relation inverse de proportionnalité qui existe entre les évolutions futures des deux pratiques médicales. A partir des données historiques, le modèle prévoit une augmentation continue de la fréquence de consommation de téléconsultation, tandis que celle de consultation

est vouée à diminuer dans le futur (selon le modèle). De plus, le modèle mesure une corrélation très négative (-0.933) entre les résidus des prédictions de la série de fréquence de téléconsultation et ceux de la série de fréquence de consultation.

Toutefois, il faut tenir en compte que ces résultats sont en grande partie dus à l'apparition de la Covid-19 en 2020, permettant à la téléconsultation de connaître un incroyable succès en France, de part le contexte exceptionnel de l'épidémie. Or, ce type d'événements n'advient que très rarement, ainsi il serait intéressant de poursuivre l'étude de l'évolution de la pratique de téléconsultation, une fois les effets de la Covid-19 atténués.

Qualité du modèle

Afin d'évaluer la qualité du modèle, il est courant d'analyser les résidus de prédiction, pour vérifier les hypothèses théoriques du modèle faites sur le terme d'erreur. Comme pour la modélisation GLM, il est important de procéder à la validation croisée des modèles de séries temporelles. Ainsi, l'ajustement du modèle se fait sur 80% de la base de données considérée tandis que la prédiction est effectuée sur les 20% restants de cette même base.

Les hypothèses à vérifier sont les suivantes :

1. Les erreurs sont indépendantes : le test de Portemanteau évalue l'absence d'autocorrélation des résidus de prédiction (hypothèse nulle). Or le test a produit une p-valeur égale à 0.356 (sup 0,05), elle se situe ainsi dans la région d'acceptation de l'hypothèse nulle. Les résidus ne sont donc pas autocorrélés.
2. Les erreurs sont des bruits blancs : le test de Jacques-Béra permet de vérifier l'hypothèse de normalité des résidus de modélisation. Or la p-valeur obtenue est inférieure à 0,05 (0.008), ainsi l'hypothèse de normalité est rejetée. Néanmoins, il est important de noter que le test est susceptible d'être biaisé lorsque la base de données est restreinte (c'est le cas ici).
3. Les erreurs sont de variance constante (homoscédastiques) : Le test ARCH-LM permet de juger de l'absence d'hétéroscédasticité (en opposition à l'homoscédasticité). La p-valeur obtenue est supérieure à la valeur critique (0.393), ainsi il est confirmé que les erreurs sont de variance constante.

Il est également possible de vérifier l'hypothèse d'homoscédasticité en étudiant les autocorrélogrammes des erreurs du modèle au carré.

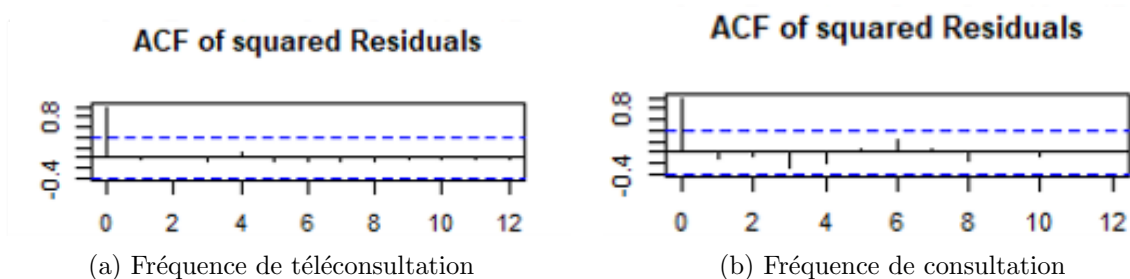


FIGURE 4.8 : Autocorrélogrammes des résidus au carré

Il est remarqué qu'aucun des retards ne sort de l'intervalle de confiance, ainsi il est possible de confirmer l'homoscédasticité des résidus par approche graphique.

Analyse structurelle

Comme mentionné auparavant, l'une des principales applications des modèles VAR est la prévision. La structure du modèle VAR fournit des informations sur la capacité de prévision d'une variable ou d'un groupe de variables pour d'autres variables.

Selon Granger (1969), si une variable, ou un groupe de variables, Y_1 s'avère utile pour prédire une autre variable, ou groupe de variables, Y_2 on dit que Y_1 est la cause de Granger de y_2 ; sinon, on dit qu'elle n'est pas la cause de Granger de Y_2 .

Formellement, Y_1 ne cause pas au sens de Granger Y_2 si, pour tous les $s > 0$, l'EQM d'une prévision de $Y_{2,t+s}$ basée sur $(Y_{2,t}, Y_{2,t-1}, \dots)$ est la même que la MSE d'une prévision de $y_{2,t+s}$ basée sur $(Y_{2,t}, Y_{2,t-1}, \dots)$ et $(Y_{1,t}, Y_{1,t-1}, \dots)$. Il est clair que la notion de causalité de Granger n'implique pas une véritable causalité. Elle implique seulement une capacité de prévision.

A la suite de l'application aux données du test de causalité au sens de Granger, il résulte que la fréquence de consultation influence, au sens de Granger, la fréquence de téléconsultation. Néanmoins le contraire n'est pas vérifié.

4.2 Calibration de scénarios d'évolution de la consommation

Afin d'approfondir l'étude de l'évolution future de la sinistralité en téléconsultation, divers scénarios d'évolution de la consommation de consultations et de téléconsultations entre 2021 et 2070 ont été analysés. Ces scénarios ont été calibrés à partir des modèles de tarification et de séries temporelles précédemment étudiés afin de les challenger. Il sera ainsi possible de tester la sensibilité des modèles GLM de coût moyen et de fréquence des sinistres en les stressant à l'aide des scénarios. Cela est également valable pour les prédictions des fréquences de consultation et de téléconsultation fournies par le modèle VAR de séries temporelles utilisé plus tôt. L'objectif final est de tirer des éventuelles conjectures quant à la place qu'occupera la téléconsultation dans le paysage assurantiel en santé à l'horizon 2070.

Un scénario est une représentation plausible de différents futurs envisageables. La construction de scénarios repose sur les variables pivots préalablement identifiées. Les variables pivots sont les facteurs susceptibles d'affecter significativement le phénomène étudié. Dans le cadre du mémoire, des hypothèses de projection sur les paramètres suivants seront prises en compte :

- les expositions futures aux sinistres,
- le coût moyen des sinistres futurs,
- la fréquence de survenance des sinistres futurs.

La constitution des scénarios repose alors sur le croisement de ces hypothèses appliquées aux modèles de tarification et de séries temporelles précédemment utilisés.

4.2.1 Présentation des scénarios de déformation de la population

La première étape a été d'émettre des hypothèses sur la déformation future de la population française. A cet effet, les projections de population, pour la période du 1er janvier 2021 au 1er janvier 2070,

élaborées par l'INSEE (2019) ont été exploitées pour interroger les modèles prospectifs (GLM) de coût moyen et de fréquence détaillés auparavant.

Ces projections portent sur la France dans son ensemble (France métropolitaine + 5 DOM), tout comme les bases de modélisation utilisées antérieurement. Elles s'appuient sur les estimations de population par sexe et âge au 1er janvier 2021, publiées en mars 2021. Les projections ont été réalisées à partir d'hypothèses sur les trois éléments conditionnant l'évolution d'une population : fécondité, mortalité et migration. Chaque facteur admet une hypothèse centrale, haute et basse. La figure suivante détaille les hypothèses prises en compte lors de l'exercice de projection de 2021.

	Situation en 2020	Hypothèse centrale	Hypothèse basse	Hypothèse haute
Fécondité				
Indice conjoncturel de fécondité	1,83 enfant par femme	1,80 enfant à partir de 2022	1,60 enfant à partir de 2030	2,00 enfant par femme à partir de 2030
Age moyen à la maternité	30,8 ans	33,0 ans à partir de 2052		
Espérance de vie				
Espérance de vie à la naissance des femmes	85,1 ans	90,0 ans en 2070	86,5 ans en 2070	93,5 ans à partir de 2021
Espérance de vie à la naissance des hommes	79,1 ans	87,5 ans en 2070	84,0 ans en 2070	91,0 ans à partir de 2021
Migrations				
Valeur du solde migratoire annuel	+ 87 000 par an	+ 70 000 par an	+ 20 000 par an	+120 000 par an

FIGURE 4.9 : Les hypothèses de l'exercice de projection de 2021 (INSEE)

La combinaison de ces hypothèses entraîne ainsi la création de 27 scénarios de déformation de la population française, de 2021 jusqu'à 2070.

Pour chaque scénario, les données suivantes sont fournies dans un fichier au format *.csv* :

- La population par âge détaillé et sexe de 1962 à 2070 ;
- Le nombre de décès par âge détaillé et sexe de 1962 à 2070 ;
- Les quotients de mortalité par âge détaillé et sexe de 1962 à 2070 ;
- Les espérances de vie à la naissance, à 60 ans et à 65 ans par sexe de 1962 à 2070 ;
- Les naissances par âge de la mère de 1975 à 2070 ;
- Les taux de fécondité par âge de la mère de 1962 à 2070 ;
- L'indicateur conjoncturel de fécondité et l'âge moyen à la maternité de 1962 à 2070 ;
- La descendance finale et l'âge moyen à la maternité des générations 1947 à 2020 ;
- Le solde migratoire par âge détaillé et sexe de 2021 à 2070, le solde migratoire global seulement de 1962 à 2020 ;

- L'ajustement par âge détaillé et sexe de 2021 à 2022, l'ajustement global seulement de 1962 à 2020.

4.2.2 Présentation des conjectures effectuées sur le coût moyen

Dans un deuxième temps, des hypothèses sur le taux annuel d'inflation ont été réalisées dans le cadre de l'étude.

L'inflation se caractérise par une augmentation durable de la moyenne des prix des biens et des services, indépendamment des variations sectorielles, sur l'ensemble du spectre économique et sur plusieurs trimestres consécutifs. Elle se traduit donc par une diminution du pouvoir d'achat de la monnaie.

Pour une zone économique et une période données, le taux d'inflation s'évalue via un relevé systématique des prix à la consommation et l'établissement d'un indice (IPC en France, évalué par l'INSEE). Dans la zone euro, la hausse des prix à la consommation est mesurée par « l'indice des prix à la consommation harmonisé » (IPCH). Le terme « harmonisé » signifie que l'ensemble des pays de l'Union européenne ont adopté la même méthodologie, de sorte que les données peuvent être comparées entre les pays. Cette mesure globale permet, notamment à la Banque Centrale Européenne (BCE), de suivre de près l'évolution des prix dans l'économie.

L'intégration de l'inflation dans le processus de construction des scénarios devient ainsi une nécessité pour assurer leur pertinence et leur cohérence. Pour ce faire, les projections réalisées par la Banque de France sur le taux d'inflation pour la période 2022-2024 ont été mises à profit lors des travaux (dernière mise à jour : 15 septembre 2022). Pour les années $2024 + k$, $k = 1, \dots, 46$, il est supposé que les taux annuels d'inflation pour la période 2024-2070 restent constants.

Le tableau suivant récapitule les taux d'inflation et les facteurs de multiplication associés. Ces derniers seront appliqués aux valeurs fournies par les modèles GLM de coût moyen de la consultation et de la téléconsultation.

Année	Taux prédits par la Banque de France (Mise à jour : 15 septembre 2022)	Formule du coût moyen (CM) annuel des sinistres, en prenant compte de l'inflation
2021	2,1%	$CM_{2020} \times (1 + 2,1\%)$
2022	5,8%	$CM_{2021} \times (1 + 5,8\%)$
2023	4,7%	$CM_{2022} \times (1 + 4,7\%)$
2024	2,7%	$CM_{2023} \times (1 + 2,7\%)$
2024 + k	-	$CM_{2023} \times (1 + 2,7\%)^k, k = 1, \dots, 46$

FIGURE 4.10 : Tableau récapitulatif des hypothèses de projection prises sur le taux d'inflation

4.2.3 Présentation des hypothèses de projections de la fréquence

Contrairement au coût moyen, la fréquence de survenance des sinistres n'admet pas d'indicateur d'évolution incontestable tel que le taux d'inflation. En effet, l'inflation permet de rendre plus réalistes et pertinentes les prédictions des niveaux futurs de dépenses en consultation et en téléconsultation.

Pour pallier à ce constat, en plus des valeurs fournies par les modèles GLM de fréquence des deux pratiques médicales étudiées, les conjectures suivantes ont été considérées :

- L'évolution des fréquences de survenance suit les tendances résultant des prédictions du modèle VAR de séries temporelles. Ainsi il est supposé que la fréquence de téléconsultation évolue linéairement entre deux janvier 2021 et janvier 2070 : elle augmente de 0,0075 à 0,2 téléconsultations par mois, tous profils d'assuré confondus.

Concernant la fréquence de consultation, il est prévu théoriquement qu'elle décroisse linéairement de 0,23 jusqu'à -3 consultations par mois. Néanmoins une fréquence ne peut pas être négative. Ainsi, suite à l'avis d'experts sur le sujet, il a été jugé pertinent de supposer que la fréquence de consultation ne pourra jamais atteindre la fréquence de consultation critique . Cette dernière est égale à la moitié de la fréquence de téléconsultation maximale.

Ainsi l'évolution de la fréquence de consultation serait représentée par une droite linéaire décroissante, passant par les points 0,23 en janvier 2021 et 0,1 en janvier 2070.

- Uniquement pour la fréquence de téléconsultation : le scénario d'évolution dit « pessimiste » prévoit le retour en 2070 à la fréquence de survenance enregistrée en 2019. Par conséquence, en suivant une approche linéaire, cette variable pivot devrait chuter à 0,0012 téléconsultations en moyenne par mois, contre 0,0075 actes de téléconsultations enregistrés en janvier 2021.
- Uniquement pour la fréquence de consultation : ce dernier scénario a pour but de challenger la relation existante entre la fréquence de consultation et la fréquence de téléconsultation. Pour cela, l'évaluation de leurs corrélation et covariance empirique sera suivie de l'application de la formule de corrélation linéaire (Pearson) à la table des valeurs résultant du modèle GLM de fréquence de téléconsultation, pour prédire la fréquence de consultation associée.

Focus sur le scénario d'évolution basé sur la corrélation entre les fréquences de consultation et de téléconsultation

On considère deux variables réelles finies $\mathbf{x} = (x_i)$ et $\mathbf{y} = (y_i)$, pour $i = 1, \dots, n$. Le coefficient de corrélation de Pearson ρ est calculé en utilisant l'expression :

$$\rho = \frac{\sum [(x_i - \bar{x})(y_i - \bar{y})]}{\sqrt{\sum (x_i - \bar{x})^2 * \sum (y_i - \bar{y})^2}}. \quad (4.14)$$

Par ailleurs, le coefficient de corrélation théorique peut aussi se définir à partir de la covariance :

$$\rho = \frac{\text{Cov}(\mathbf{x}, \mathbf{y})}{\sqrt{\text{Var}(\mathbf{x}) \text{Var}(\mathbf{y})}}, \quad (4.15)$$

avec :

$$\text{Cov}(\mathbf{x}, \mathbf{y}) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n - 1}. \quad (4.16)$$

Dans le cadre de l'étude, trois corrélations distinctes ont été considérées :

- Corrélation « mensuelle » : calculée à partir des séries mensuelles des fréquences de consultation et de téléconsultation, résultant de l'agrégation par mois de remboursement et profil d'assuré. Cette dernière agrégation permet d'harmoniser le nombre d'observations entre les deux pratiques médicales.
- Corrélation « annuelle » : évaluée en agrégeant les données par profil d'assuré et par année de remboursement. Cela donne lieu à deux séries de fréquences mensuelles, une pour la consultation et l'autre pour la téléconsultation.
- Corrélation « globale » : calculée à partir d'agrégation des observations par profil d'assuré. Il en découle alors deux séries de données, une représentant la fréquence moyenne de consultation par profil d'assuré et l'autre celle de téléconsultation par profil d'assuré.

Ces corrélations ont été établies à l'aide du test de corrélation utilisant la méthode de Pearson (fonction *cor.test* sur R). Le tableau suivant présente les coefficients, les p-valeurs et les niveaux de significativité produit par le test.

	Coefficient	P-valeur	Significativité
Corrélation globale	0,05	4,91E-06	***
Corrélation mensuelle	-0,73	8,85E-06	***
Corrélation annuelle	-0,99	0,06	.

FIGURE 4.11 : Tableau récapitulatif des hypothèses de projection prises sur le taux d'inflation

Il est important de rappeler que la vérification des hypothèses de normalité et de covariance linéaire entre x et y est nécessaire au calcul de leur corrélation linéaire de Pearson (méthode paramétrique). Ainsi, afin de déterminer la corrélation linéaire à inclure dans ce dernier scénario d'évolution de la fréquence de consultation, le choix portera sur les séries de données vérifiant davantage les hypothèses théoriques.

Vérification de la normalité

Pour tester la normalité des séries globales, mensuelles et annuelles, il a été opté pour une approche graphique. Cela a ainsi consisté en l'analyse de leurs qqplots, présentés ci-après.

Il est clair que les séries globales de fréquences de consultation et de téléconsultation ne sont pas normales. En effet, il est observé que la majorité des points reste à l'écart de l'intervalle de confiance illustré sur les graphiques. Au contraire, les nuages de points des séries annuelles sont entièrement compris entre les limites de normalité. Il est déduit, par approche graphique, que ces dernières suivent une distribution normale. Enfin, les QQ plot des séries mensuelles sont moins explicites que ceux précédemment analysés. Toutefois, il est remarqué que la plupart des points se situent dans l'intervalle de confiance mis en évidence sur les figures. Il est alors considéré que les séries mensuelles sont aussi normales.

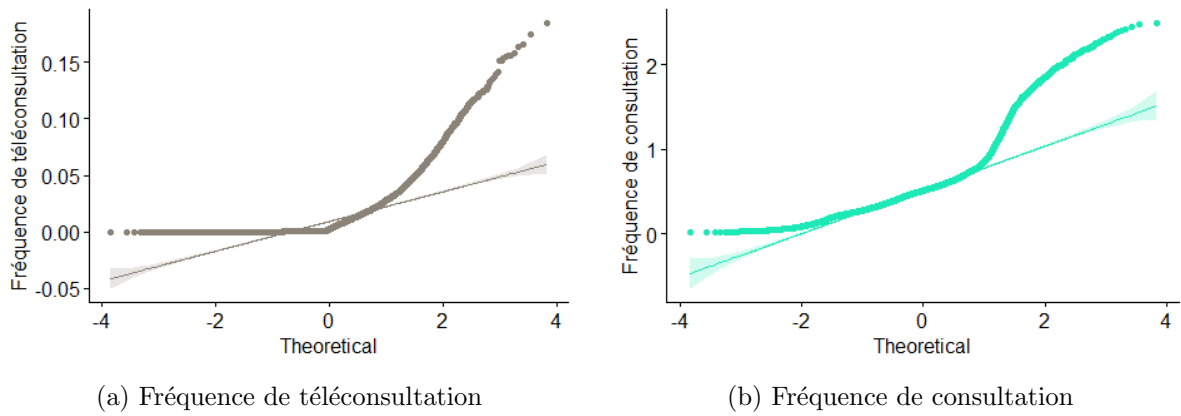


FIGURE 4.12 : Q-Q Plot des séries globales de fréquence de consultation et de téléconsultation

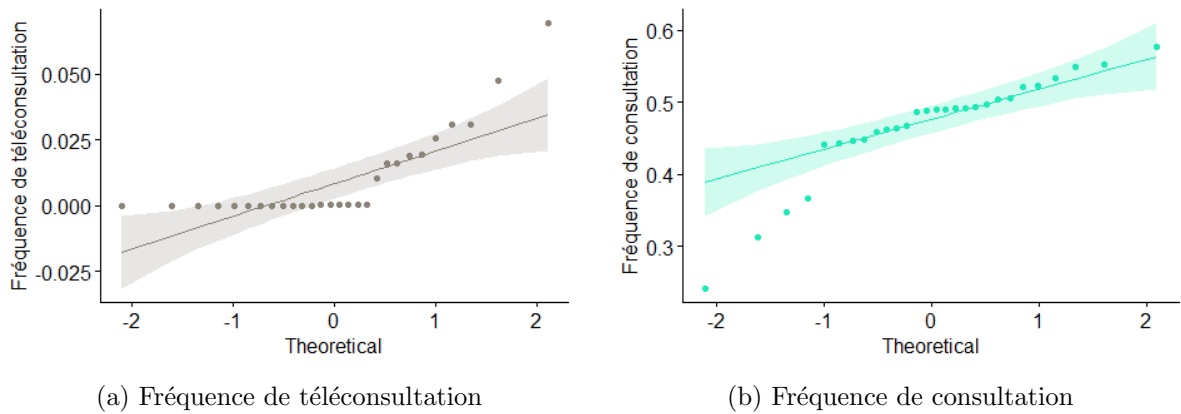


FIGURE 4.13 : Q-Q Plot des séries mensuelles de fréquence de consultation et de téléconsultation

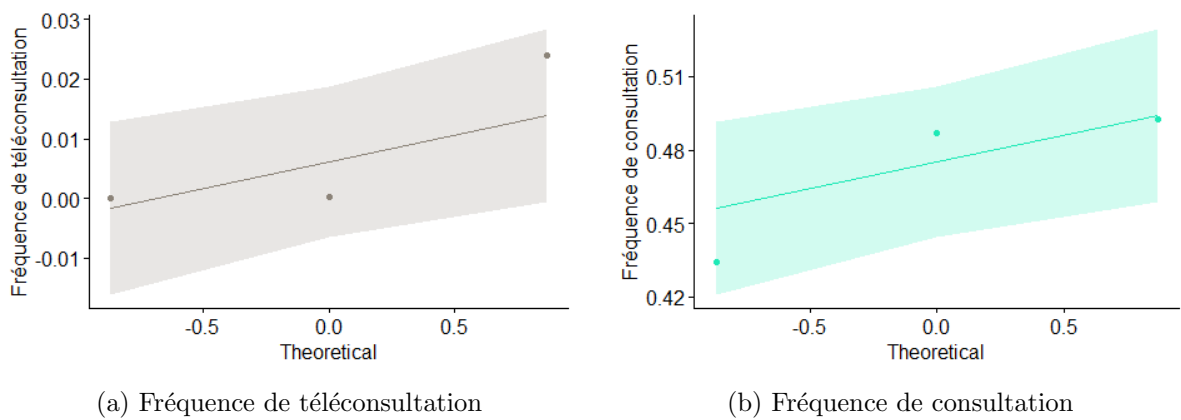


FIGURE 4.14 : Q-Q Plot des séries annuelles de fréquence de consultation et de téléconsultation

Vérification de la covariance linéaire

Concernant la covariance linéaire, les figures suivantes - représentant l'évolution de la fréquence de téléconsultation en fonction de la fréquence de consultation - permettront de trancher sur le sujet, par approche graphique.

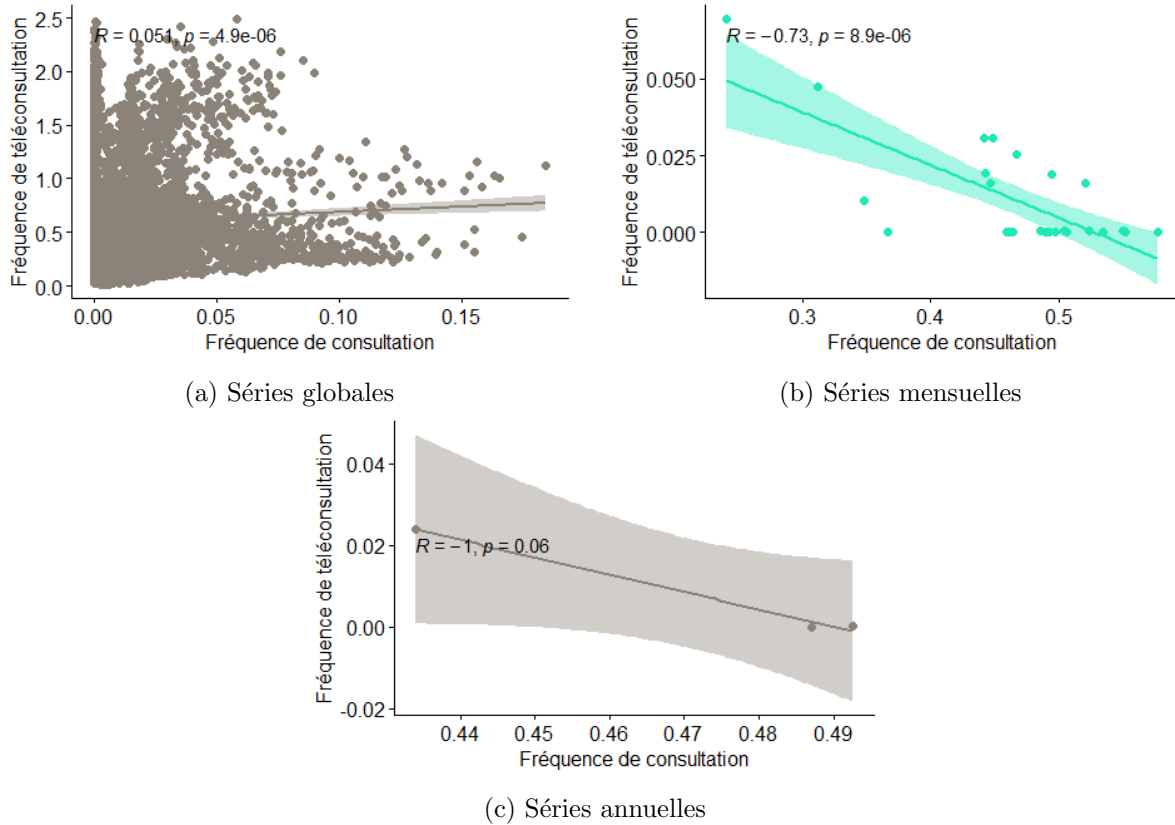


FIGURE 4.15 : L'évolution de la fréquence de téléconsultation en fonction de la fréquence de consultation, pour les séries globales, mensuelles et annuelles

Graphiquement, il peut être déduit que le couple de séries annuelles et le couple de séries mensuelles sont linéairement liés. En effet, les nuages de points ne présentent pas de tendances particulières et sont relativement proches de l'intervalle de confiance. Cependant, le nuage de points représentant la fréquence globale de téléconsultation en fonction de la fréquence globale de consultation n'est pas centré autour de la droite linéaire, ainsi les séries globales ne sont pas linéairement liées.

Compte tenu des résultats trouvés lors de la vérification des hypothèses de la corrélation linéaire et des niveaux de significativité associées aux différentes corrélations (figure 4.11), il a été considéré judicieux d'utiliser la corrélation mensuelle. D'une part, cela s'explique par le fait que son niveau de significativité est plus élevé que celui associé à la corrélation annuelle. D'autre part, les séries mensuelles de consultation et de téléconsultation s'avèrent plus pertinentes que les séries globales au regard des hypothèses du test de corrélation.

Comme susmentionné, la corrélation empirique choisie sera appliquée aux données de fréquence de téléconsultation fournies par le modèle GLM associé. Cette application se fera à l'aide de la formule du coefficient de corrélation linéaire. Cependant, il existe une contrainte non négligeable : la relation entre le coefficient de corrélation et les variables considérées n'est pas linéaire. En réalité, il s'agit

d'une équation à n inconnues ($n = 28$ pour les séries mensuelles). Ainsi l'équation, dans laquelle les valeurs des fréquences mensuelles de consultation sont les inconnues, ne peut pas être résolue.

Pour répondre à cette problématique, l'approximation suivante a été retenue : il est supposé que les fréquences mensuelles de consultation (y_i) et de téléconsultation (x_i) soient linéairement reliées par l'équation $y_i = a \times x_i$. Par conséquent, en partant de l'équation 3.42 et sachant que la covariance empirique des séries mensuelles est connue, il est déduit que :

$$\rho = \frac{\text{Cov}(\mathbf{x}, \mathbf{y})}{a^2 \text{Var}(\mathbf{x})^2} \rightarrow a = \frac{1}{\text{Var}(\mathbf{x})} \times \sqrt{\frac{\text{Cov}(\mathbf{x}, \mathbf{y})}{\rho}}$$

Ainsi, il est possible de projeter la fréquence de consultation dans le futur à partir des valeurs de fréquence de téléconsultation prédites par le modèle GLM correspondant. Suite à l'évaluation des facteurs empiriques suivants : variance, covariance et corrélation empiriques, il résulte que le coefficient linéaire a vaut 28,8. Il peut ainsi être conclu que le scénario tenant compte de la corrélation entre fréquence de téléconsultation et fréquence de consultation conjecture que cette dernière sera constamment proportionnelle à la première avec un coefficient de proportionnalité égal à 28,8.

4.2.4 Implémentation de ces hypothèses

Il est rappelé que les modèles GLM retenus pour simuler la fréquence de survenance et le coût moyen d'actes de téléconsultation sont respectivement les suivants :

- Le modèle binomial négatif d'équation :

Année + Age + Affiliation CMU + Sexe + Nature de la prestation

- Le modèle mélange de deux lois gamma d'équation :

Année + Age + Affiliation CMU + Nature de la prestation

Les modèles GLM choisis pour la consultation sont présentés ci-après :

- Le modèle Binomial négatif de la fréquence d'équation :

Année + Age + Affiliation CMU + Sexe + Nature de la prestation

- Le modèle gamma du coût moyen d'équation :

Sexe + Age + Affiliation CMU + Nature de la prestation

Une première problématique a été relevée : les seules données de projection connues et pouvant être exploitées (i.e. apparaissant dans les modèles ci-dessus), sont les variables année, âge et sexe. D'ultérieures hypothèses sur les variables affiliation à la CMU et nature de prestation devront ainsi être élaborées afin de projeter la consommation d'actes de téléconsultation dans le futur.

Faute de ressources trouvées non seulement sur des scénarios de prévision de la répartition par âge et sexe des assurés CMU à l'horizon 2070, mais aussi sur des scénarios d'évolution de la sinistralité en téléconsultation en fonction de la nature de prestations, certaines approximations ont été effectuées.

Concernant l'affiliation à la CMU, des moyennes pondérées ont été calculées à l'aide des nouvelles répartitions de la population par âge et par sexe, en partant des ratios régionaux d'assurés CMU moyens enregistrés sur la période de modélisation historique 2018-2020. Il a été observé que d'une année à une autre ces ratios sont relativement stables (écart relatif = 0.4% entre 2018 et 2020). Ainsi, l'utilisation de la répartition régionale historique des affiliés à la CMU est justifiée.

A propos de la nature des prestations de téléconsultation, il a été relevé qu'il s'agit d'une donnée renseignant sur la sinistralité à une maille plus fine. De ce fait, en l'absence de ressources documentaires et d'avis d'experts fiables sur le sujet, il a été préféré de limiter l'exercice de projection à une approche globale afin d'analyser la consommation en téléconsultation. Dans cette intention, les modèles GLM de coût moyen et de fréquence de téléconsultation (et leurs coefficients) ont été agrégés selon la nature des actes de la pratique médicale.

Une deuxième problématique s'est posée lors de la calibration des modélisations GLM de coût moyen et de fréquence de consultation et de téléconsultation en partant des nouvelles expositions fournies par les scénarios de déformation de la population. Il a été constaté que quel que soit le scénario de déformation considéré, les prédictions de coût (en l'absence de prise en compte de l'inflation) et de fréquence fournies par les modèles restent uniformes. Entre 2021 et 2070, il est prévu que les habitudes de consommation restent homogènes, autour de 0,062 téléconsultations par an et de 0,35 consultations par an. Ces valeurs caractérisent les fréquences prises en compte dans le scénario dit « central ». Il a donc été décidé de ne plus inclure les scénarios d'évolution de la population parmi les hypothèses de projection.

4.2.5 Synthèse des scénarios finaux et analyse des résultats

L'imbrication des hypothèses de projections précédemment explicitées, donne lieu aux scénarios d'évolution présentés par les tableaux suivants.

Stress Téléconsultation (TC)			
	Source de fréquence	Source de coût moyen	Hypothèses de projection
Scénario central	Résultats des modèles GLM	Résultats des modèles GLM	Prise en compte de l'inflation
Scénario séries temporelles	Prédictions du modèle VAR de séries temporelles	Résultats des modèles GLM	Prise en compte de l'inflation
Scénario pessimiste	Résultats des modèles GLM	Résultats des modèles GLM	<ul style="list-style-type: none"> Prise en compte de l'inflation Décroissance linéaire de la fréquence jusqu'à atteindre en 2070 la valeur enregistrée en 2019
Stress Consultation (C)			
	Source de fréquence	Source de coût moyen	Hypothèse de projection
Scénario central	Résultats des modèles GLM	Résultats des modèles GLM	Prise en compte de l'inflation
Scénario séries temporelles	Prédictions du modèle VAR de séries temporelles	Résultats des modèles GLM	<ul style="list-style-type: none"> Prise en compte de l'inflation Etablissement d'une valeur critique pour la fréquence de consultation par avis d'experts
Scénario tenant compte de la corrélation	Résultats des modèles GLM	Résultats des modèles GLM	<ul style="list-style-type: none"> Prise en compte de l'inflation Etude de l'influence de la corrélation empirique existant entre les fréquences de C et de TC

FIGURE 4.16 : Tableaux récapitulatifs des scénarios d'évolution étudiés lors des travaux

Le croisement de ces scénarios d'évolution permet ainsi de déterminer les primes pures (PP) annuelles de consultation (C) et de téléconsultation (TC) entre 2021 et 2070.

Afin d'exploiter ces résultats, il a été considéré que les agréger ou les multiplier ne serait pas cohérent. Finalement, il a été préféré d'analyser les divers couples de primes pures ayant comme expression générale $[PP_{\text{Année}}(C); PP_{\text{Année}}(TC)]$ et établis par croisement des scénarios de stress de la consultation et de la téléconsultation.

Les figures se trouvant en annexe 3 présentent sur un même graphique :

- d'une part, l'évolution des primes pures de consultation et de téléconsultation entre 2018 et 2070, selon les différents scénarios élaborés,
- d'autre part, les tendances suivies par les sommes des couples de primes pures de consultation et de téléconsultation entre 2018 et 2070, selon les croisements des scénarios calibrés. Celles-ci seront qualifiées de « primes totales » du panier de soins de consultation et de téléconsultation dans la suite du mémoire.

L'analyse de ces figures montrent que, dans la vaste majorité des cas, les tendances d'évolution de la consommation de consultations et de téléconsultation restent constantes dans le temps. En effet, il est observé que la sinistralité en consultation conserve sa prépondérance par rapport à celle de téléconsultation, tandis que la consommation globale d'actes des deux pratiques médicales croît quelque soit le scénario de projection pris en compte. Cependant, cela ne s'applique pas au cas du croisement de scénarios prenant compte les prédictions du modèle de série temporelle VAR(7) appliqué au données. En effet, à partir de 2054, l'influence de la téléconsultation, en termes de sinistralité, surpasse celle de consultation. Cela sera aussi confirmé par les analyses quantitatives présentées dans la suite du mémoire.

L'objectif principal de la représentation $[PP_{2070}(C); PP_{2070}(TC)]$ est de prédire l'influence qu'aura la téléconsultation dans le paysage de l'assurance Santé à l'horizon 2070. A cet effet, un focus est fait sur la part de sinistralité portée par la téléconsultation par rapport à la prime totale du panier d'actes de consultation et de téléconsultation en 2070, en fonction des différentes hypothèses de projection.

Le tableau suivant répertorie les couples de primes pures des deux pratiques médicales selon les croisements de scénarios, ainsi que leur somme.

Couple (PP_C ; PP_TC) prédit pour 2070 et leur somme, en €	Scénario central	Scénario séries temporelles	Scénario pessimiste
Scénario central	(26,94 ; 4,93) 31,87	(26,94 ; 15,98) 42,92	(26,94 ; 0,94) 27,88
Scénario séries temporelles	(7,89 ; 4,93) 12,82	(7,89 ; 15,98) 23,87	(7,89 ; 0,94) 8,83
Scénario tenant compte de la corrélation	(142,05 ; 4,93) 146,98	(142,05 ; 15,98) 158,03	(142,05 ; 0,94) 142,99

FIGURE 4.17 : Tableau récapitulatif des primes pures résultant du croisement des scénarios d'évolution

Tout d'abord, il est rappelé que le couple de primes pures de consultation et de téléconsultation pour l'année de modélisation 2020 est (43, 17; 9, 34) (obtenu à la suite d'une analyse tarifaire détaillée dans le chapitre 2). Il est alors remarqué que les sinistralités de consultation, téléconsultation et du panier global prévues par les scénarios centraux rencontreront, en moyenne, une réduction de 40% à l'horizon 2070 par rapport à celles datant de 2020.

De plus, il est observé que le niveau de volatilité par rapport à la prime totale résultant des scénarios centraux de consultation et de téléconsultation est relativement mesuré, selon les scénarios considérés. En effet, six croisements de scénarios sur neuf enregistrent un écart relatif moyen de 35%. Les couples de primes pures s'éloignant le plus du scénario central de la sinistralité globale (écart relatif $\sim 400\%$) sont ceux résultant des hypothèses de projection de la relation existante entre la fréquence des deux pratiques médicales à travers la notion de corrélation. La principale explication réside dans le fait que la corrélation retenue est négativement très élevée.

Par ailleurs, il est constaté que dans le scénario le plus défavorable pour les assureurs, la sinistralité globale est multipliée par 5 par rapport au couple de scénarios centraux. Ce scénario correspond à l'association de l'effet de l'inflation aux hypothèses de :

- relation linéaire entre les fréquences de consultation et de téléconsultation pour la prime pure de consultation ;
- projection de la fréquence de téléconsultation résultant du modèle de séries temporelles VAR(7) appliqué plus tôt dans le mémoire.

Le scénario le plus avantageux est, quant à lui, caractérisé par une sinistralité globale qui se divise en 4, comparée à celle provenant des scénarios centraux. Ce scénario se compose des hypothèses suivantes (en plus de la prise en considération de l'inflation) :

- le retour à la fréquence de téléconsultation enregistrée en 2019 pour la prime de téléconsultation ;
- la prise en compte de la prévision de fréquence de consultation engendrée par la modélisation VAR(7).

En outre, il est observé que quelque soit la pratique médicale, la volatilité entre les sinistralités individuelles prédites par les scénarios de projection est haute. En effet, la téléconsultation enregistre un écart relatif moyen de 150% par rapport à la prime pure dérivant du croisement des scénarios centraux pour l'année 2070. Cet indicateur vaut 250% pour la consultation.

Le tableau suivant présente les parts (en termes de sinistralité), du panier de soins étudié, prises par la téléconsultation en fonction des scénarios croisés. Ces ratios sont calculés selon la formule suivante :

$$\frac{PP_{2070}(TC)}{PP_{Totale2070}}$$

En pourcentage (%)	Scénario central	Scénario séries temporelles	Scénario pessimiste
Scénario central	15,5	37,2	3,4
Scénario séries temporelles	38,5	66,9	10,6
Scénario tenant compte de la corrélation	3,4	10,1	0,7

FIGURE 4.18 : Ratios de sinistralité portés par la téléconsultation en fonction des scénarios d'évolution

Celui-ci montre qu'il existe quatre niveaux caractérisant les projections de la proportion d'actes de téléconsultation parmi la consommation globale de consultations et de téléconsultations :

- Le premier niveau inclut les projections les plus défavorables à l'essor de la téléconsultation à l'horizon 2070. Elle comprend les ratios allant de 0,7 à 3,4 %, touchant trois croisements de scénarios d'évolution (sur neuf). L'hypothèse pénalisant le plus la croissance de la pratique de téléconsultation à l'horizon 2070 est alors celle de corrélation linéaire entre les fréquences de consultation et de téléconsultation.
- Le deuxième niveau comprend les ratios allant de 10,1 à 15,5 % . Les projections de trois croisements de scénarios d'évolution (sur neuf) font partie de ce niveau. Étant donné que la part du panier de soins prise par la téléconsultation valait 17,8 % en 2020, ces croisements de scénarios sont ceux qui se rapprochent le plus de la part de sinistralité actuelle de téléconsultation.
- Le troisième niveau est partagé par deux croisements et considère qu'en 2070, la sinistralité de téléconsultation occupera approximativement le tiers de la sinistralité globale de consultations et de téléconsultations (entre 37,2 et 38,5 %). Il s'agit du croisement du scénario central de consultation et du scénario intégrant les projections des modèles de séries temporelles pour la téléconsultation et vice versa.
- Enfin, la projection fournie par le croisement des scénarios tenant compte des prévisions des modèles de séries temporelles de la fréquence de consultation et de téléconsultation (ainsi que l'application de l'inflation) constitue le dernier niveau. Elle prévoit que la sinistralité en téléconsultation aura deux fois plus de poids et d'influence (66,9%) dans le panier d'assurance santé constitués de consultations et de téléconsultations que la pratique de consultation, à l'horizon 2070.

Conclusion

Après des débuts timides, la crise de la Covid a agit en tant qu'accélérateur d'adoption de la téléconsultation, qui, en l'espace de quelques semaines, a connu un véritable essor en France en 2020. Le regard à la fois des professionnels mais aussi des patients a changé vis à vis du numérique à la faveur de la téléconsultation. Cette pratique médicale représente un enjeu essentiel dans l'amélioration du système de santé actuel car pour autant que l'on dispose d'un smartphone, elle offre une réponse immédiate à des problématiques telles que la difficulté d'accès aux soins, dans les déserts médicaux par exemple, ou encore l'amélioration de la prise en charge et du suivi des patients.

Ces évolutions de consommation représentent une incertitude pour les assureurs et soulèvent donc le besoin d'en étudier l'essor et l'amplitude actuelle, dans un contexte de maîtrise des risques futurs.

Ce mémoire a voulu s'inscrire dans ce cadre en proposant tout d'abord la mise en place de modèles tarifaires prédictifs de la sinistralité en consultation et en téléconsultation, afin d'analyser leur relation dans le temps. Ensuite, un deuxième objectif s'est ajouté : projeter la sinistralité dans le futur afin de prédire l'influence qu'aura la téléconsultation en assurance Santé, notamment en comparaison du poste de soins "Soins de ville" qui lui est équivalent.

Pour ce faire, une analyse tarifaire de la consommation des deux pratiques médicales a d'abord été réalisée, à travers la mise en place de modèles GLM. Cela a permis de déterminer les primes pures annuelles de 2018 à 2020. Ensuite, l'évolution temporelle de la sinistralité de consultation et de téléconsultation future a été modélisée, à l'aide de modèles multivariés de séries temporelles des fréquences associées. Les projections du modèle retenu mettent en exergue la relation inverse de proportionnalité qu'il existe entre les évolutions futures des deux pratiques médicales. A partir des données historiques, le modèle prévoit une augmentation continue de la fréquence de consommation de téléconsultation, tandis que celle de consultation est vouée à diminuer dans le futur. Enfin, des scénarios de projection de la consommation en téléconsultation et en consultation classique ont été produits, sur la base des modèles précédemment utilisés. Cela a permis de tirer des conjectures quant aux possibles évolutions de la sinistralité future de téléconsultation.

Ces études ont montré que la sinistralité du panier de soins constitué d'actes de consultation et de téléconsultation a de grandes chances de doubler voir tripler dans le futur et ce, avec ou sans la prise en compte du taux d'inflation annuel. De plus, il est constaté que les écarts de sinistralité entre les deux pratiques médicales par rapport à la consommation globale de soins se creuseront probablement au cours du temps. Cependant, il est relevé que la place de la téléconsultation dans ce panier de soins (en termes de sinistralité) restera vraisemblablement inférieure à celle de consultation à l'horizon 2070, bien que les tendances projetées par les modèles soient en majorité linéairement croissantes entre 2020 et 2070.

Il est toutefois important de soulever un point d'attention sur les limites des travaux réalisés et les ouvertures potentielles dans le cadre de ce mémoire.

Tout d'abord, la profondeur de l'historique des données est limitée, ce qui introduit un biais dans les études. Tout au long du mémoire, il a été constaté que la majorité des résultats de modélisation sont en grande partie dus à l'apparition de la Covid-19 en 2020. Comme susmentionné, cet événement a favorisé l'incroyable succès de la téléconsultation en France. Toutefois, ce type d'événements n'advient que très rarement, ainsi il serait intéressant de poursuivre l'étude de l'évolution de la pratique de téléconsultation, une fois les effets de la Covid-19 atténués. Cela serait alors possible en intégrant dans l'étude les données de consommation de consultation et de téléconsultation datant de 2021, désormais disponibles sur le site de l'assurance maladie.

De plus, plusieurs approximations ont été réalisées, notamment lors de la production des scénarios d'évolution de la sinistralité prenant compte de la corrélation linéaire entre les fréquences de consultation et de téléconsultation pour en déduire la prime pure de consultation à l'horizon 2070 ;

Enfin, il aurait été intéressant de poursuivre l'étude de l'influence de la région ou du département sur la sinistralité de téléconsultation dans le temps.

Bibliographie

- ADOU, A.-E. (2020). Modélisation du risque hospitalisation sur les données open source de la Sécurité sociale. Mémoire d'actuariat. EURIA, Univ. de Bretagne Occidentale.
- BANQUE DE FRANCE (2022). Projections Macroéconomiques France. URL : https://www.banque-france.fr/sites/default/files/medias/documents/projections-macroeconomiques_septembre-2022.pdf.
- BENAGLIA, T., CHAUVEAU, D., HUNTER, D. R. et YOUNG, D. S. (2010). mixtools: an R package for analyzing mixture models. *Journal of statistical software* 32, p. 1-29.
- BENTATA, P. (2022). Le recours à la téléconsultation et à la téléexpertise : quel impact économique attendre en France ? URL : <https://www.institutmolinari.org/wp-content/uploads/2022/01/etude-telemedecine-janvier2022.fr.pdf>.
- BREUIL-GENIER, P. et GOFFETTE, C. (2006). La durée des séances des médecins généralistes. URL : <https://drees.solidarites-sante.gouv.fr/sites/default/files/er481.pdf>.
- De FOUCAUD, I. (2020). Le boom de la téléconsultation lié au Covid va-t-il durer ? URL : https://www.challenges.fr/entreprise/sante-et-pharmacie/le-boom-de-la-teleconsultation-lie-au-covid-va-t-il-durer_727124.
- DONNET, S. (2019). Cours de Modèles Linéaires. Université Paris Dauphine.
- DREES (2021). Panorama statistique : Cohésion sociale - Travail - Emploi. URL : <https://drees.solidarites-sante.gouv.fr/sites/default/files/panorama/historique.htm>.
- DREES (2022). Démographie des professionnels de santé. URL : <https://drees.shinyapps.io/demographie-ps/>.
- DRESS (2020). En 2018, les territoires sous-dotés en médecins généralistes concernent près de 6 % de la population. URL : <https://drees.solidarites-sante.gouv.fr/sites/default/files/er1144.pdf>.
- FERRARA, L. (2011). Séries Temporelles Avancées Polycopié de Cours. URL : http://lo.ferrara.free.fr/polycop_M2_TSA_1112.pdf.
- FONDS DE LA COMPLÉMENTAIRE SANTÉ SOLIDAIRE (2019). Rapport d'Activité. URL : https://www.complementaire-sante-solidaire.gouv.fr/fichier-utilisateur/fichiers/RA_2019_VF.pdf.
- GARNIER, A. (2020). Apport des *Open Data* pour évaluer les impacts de la réforme 100% Santé. Mémoire d'actuariat. ISFA, Univ. Claude Bernard Lyon 1.
- GRUN, B. et LEISCH, F. (2007). FlexMix: An R package for finite mixture modelling.
- GUIDE DGOS (2019). Facturation des téléconsultations et télé expertises en établissement de santé. URL : https://solidarites-sante.gouv.fr/IMG/pdf/guide_facturation_tlm_en_etablissement_de_sante.pdf.
- HAUTE AUTORITÉ DE SANTÉ (2019). Téléconsultation et téléexpertise : guide de bonnes pratiques. URL : https://www.has-sante.fr/jcms/c_2971632/fr/teleconsultation-et-teleexpertise-guide-de-bonnes-pratiques.
- HOUDAYER, C. (2021). Tarification en santé sous contrainte budgétaire d'un euro par jour à l'aide de fonctions d'utilité. Mémoire d'actuariat. Univ. Paris Dauphine, Univ. PSL.
- INSEE (2019). Projection de population 2021-2070 pour la France. URL : <https://www.insee.fr/fr/statistiques/5894886?sommaire=5760764>.

- INSTITUT DES ACTUAIRES (2021). Innovation *Open Data* : Comment les nouvelles bases de données santé en open data peuvent-elles être source d'inspiration pour l'assurance santé de demain. URL : https://www.institutdesactuaires.com/global/gene/link.php?doc_id=11745&fg=1.
- JAKOUBOVITCH, S., BOURNOT, M.-C., CERCIER, E. et TUFFREAU, F. (2012). Les emplois du temps des médecins généralistes. URL : <https://drees.solidarites-sante.gouv.fr/sites/default/files/er797-2.pdf>.
- L'ASSURANCE MALADIE (2022). La téléconsultation. URL : <https://www.ameli.fr/seine-saint-denis/exercice-coordonne/exercice-professionnel/telemedecine/teleconsultation>.
- LEPAGE, E. (2018). Téléconsultation remboursée : "Je n'avais pas vu de médecin depuis cinq ans". URL : <https://www.nouvelobs.com/sante/20180914.OBS2396/teleconsultation-remboursee-je-n-avais-pas-vu-de-medecin-depuis-cinq-ans.html>.
- MAITRA, S. (2019). Time-series Analysis with VAR VECM: Statistical approach. URL : <https://towardsdatascience.com/vector-autoregressions-vector-error-correction-multivariate-model-a69daf6ab618>.
- MALACHANE, L. (2019). L'histoire de la télémedecine. URL : <https://www.leah.care/blog/histoire-de-la-telemedecine/#:~:text=L%20histoire%20de%20la%20t%C3%A9l%C3%A9medecine&text=La%20t%C3%A9l%C3%A9assistance%20m%C3%A9dicale%20par%20t%C3%A9l%C3%A9phone,du%20radiologue%20Jacob%20Gershon%2DCohen>.
- MCADAM, A. (2020). General Linear Models in R. URL : <https://static1.squarespace.com/static/5eb33c095018927ea433a883/t/5fa0ef7532a6f7749776f2ac/1604382594791/glm-in-R.pdf>.
- MEDAVIZ (2020). La télémedecine, une solution contre le renoncement aux soins. URL : <https://www.medaviz.com/la-telemedecine-une-solution-contre-le-renoncement-aux-soins/#:~:text=En%20int%C3%A9grant%20ou%20en%20r%C3%A9int%C3%A9grant,que%20certaines%20%E2%80%9Ccontraintes%20individuelles%E2%80%9D>.
- MINISTÈRE DE LA SANTÉ ET DE LA PRÉVENTION (2022). Présentation de la Sécurité sociale. URL : <https://solidarites-sante.gouv.fr/affaires-sociales/securite-sociale/article/presentation-de-la-securite-sociale>.
- OPUSDATA (2021). Les chiffres *Open Data*. URL : <https://opusline.fr/wp-content/uploads/2021/02/OPUSDATA-3.pdf>.
- ORDRE NATIONAL DES MÉDECINS (2021). Atlas de la démographie médicale en France. URL : https://www.conseil-national.medecin.fr/sites/default/files/external-package/analyse_etude/1grhel2/cnom_atlas_demographie_medicale_2020_tome1.pdf.
- R CORE TEAM (2022). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. URL : <https://www.R-project.org/>.
- ROUVIÈRE, L. (2015). Sélection-"validation" de modèles. URL : https://perso.univ-rennes2.fr/system/files/users/rouviere.l/chapitre3_glm.pdf.
- Vector Autoregressive Models for Multivariate Time Series (2006). *Modeling Financial Time Series with S-PLUS®*. New York, NY : Springer New York, p. 385-429. URL : https://doi.org/10.1007/978-0-387-32348-0_11.

Annexes

A.1 Annexe 1 : Référentiel d'équivalence entre les actes de consultation et de téléconsultation

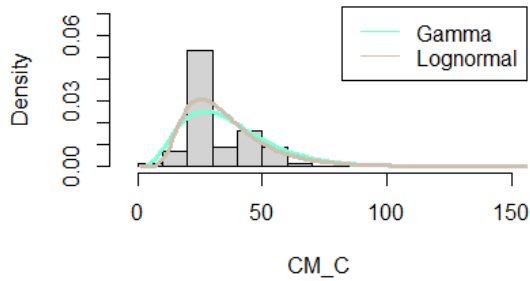
Prestation de Consultation	Catégories	Correspondance	Degré de confiance	Commentaires
1098 = CONSULTATION CCMU 3	Consultation dont l'équivalence est inconnue	3103 = TÉLÉCONSULTATION LORS D'UN SOIN	8/10	"La CCMU (Classification Clinique des Malades des Urgences) modifiée classe selon 7 degrés de gravité les patients de l'urgence pré-hospitalière (SMUR) et de l'accueil hospitalier. C'est le médecin SMUR ou de l'accueil qui détermine à la fin de l'examen clinique initial ce degré." (Source : http://medicalcul.free.fr/ccmu.html)
1099 = CONSULTATION CCMU 4 ET 5	Consultation sans équivalence	-	8/10	Les consultations correspondant au niveau CCMU 4 et 5, effectuée par un médecin urgentiste dans un SU privé font partie des consultations complexes (source : https://www.csmf.org/niveau-3-consultations-complexes/). Néanmoins, selon ameli.fr : "Ces consultations se font uniquement en présence du patient (pas de téléconsultation possible pour ces prises en charge spécifiques)." (Source : https://www.ameli.fr/assure/remboursements/rembourse/consultations-telemedecine/metropole#text_35809)
1104 = CONSULTATION OBLIGATOIRE ENFANT	Consultation sans équivalence	-	7/10	"Les visites médicales de l'enfant sont des examens obligatoires et spécifiques qui portent sur : courbe de croissance ; psychomotricité ; vaccinations ; évaluation de la capacité à effectuer certains gestes (marcher, attraper, serrer ...) ; vie affective de l'enfant ; dépistage précoce des anomalies ou déficiences ; comportements et environnements favorables à la santé, en particulier l'activité physique et sportive." (Source : https://www.service-public.fr/particuliers/vosdroits/F35490/27dFicheParent-F967#2). Il est considéré que ces examens ne peuvent pas être réalisés à distance.
1105 = CONSULTATION COMPLEXE	Consultation sans équivalence	-	10/10	"Ces consultations se font uniquement en présence du patient (il n'existe pas de téléconsultations possibles pour ces prises en charge spécifiques)." (Source : https://www.ameli.fr/assure/remboursements/rembourse/consultations-telemedecine/metropole#text_35809)
1106 = MAJORATION CONSULTATION COMPLEXE				
1107 = CONSULTATION TRES COMPLEXE ENFANT				
1108 = MAJORATION CONSULTATION TRES COMPLEXE				
1109 = CONSULTATION SPECIALISTE MEDICINE GENERALE	Consultation dont l'équivalence est connue	1192 = TÉLÉCONSULTATION GÉNÉRALISTE	9,5/10	La facturation des téléconsultations effectuées par un médecin généraliste ou spécialiste en médecine générale est bien incluse dans le guide de tarification de la téléconsultation. De plus l'acte C correspond à une consultation au cabinet effectuée par un médecin généraliste. (Sources : https://www.ameli.fr/medecin/exercice-liberal/remuneration/consultations-actes/tarifs/tarifs-generalistes/tarifs-metropole ; https://solidarites-sante.gouv.fr/IMG/pdf/guide_facturation_tim_en_etablissement_de_sante.pdf)
1110 = CONSULTATION MEDICINE GENERALE				
1111 = CONSULTATION COTEE C				
1112 = CONSULTATION COTEE CS				
1113 = CONSULTATION COTEE CNP		1191 = TÉLÉCONSULTATION TOUTES SPÉCIALITÉS		"L'acte CS correspond à une consultation au cabinet effectuée par le médecin spécialiste qualifié et le médecin spécialiste qualifié en MG. L'acte CNP correspond à une consultation effectuée par un psychiatre/neurologue/neuropsychiatre. Ainsi ces actes ont une correspondance dans le guide de facturation." (Sources : https://www.ameli.fr/medecin/exercice-liberal/remuneration/consultations-actes/tarifs/tarifs-specialistes/metropole ; http://infoslettre.info/24/IL108/docs/3-PLAQUETTE_PSY_NEURO_NEUROPSY.pdf ; https://solidarites-sante.gouv.fr/IMG/pdf/guide_facturation_tim_en_etablissement_de_sante.pdf)
1114 = CONSULTATION SPECIFIQUE CARDIOLOGIE	Consultation sans équivalence	-	7/10	Cette consultation peut être considérée comme complexe (ou très complexe), ce qui n'admet pas d'équivalent parmi les actes de téléconsultation. (Source : https://www.ameli.fr/assure/remboursements/rembourse/consultations-telemedecine/metropole#text_35809)
1115 = CONSULTATION BILAN	Consultation sans équivalence	-	9/10	"Il est possible de prétendre au bilan santé de la CPAM tous les 5 ans. L'examen est personnalisé en fonction de l'âge, des antécédents familiaux, des habitudes de vie et des facteurs de risque. En moyenne, la consultation dure 2h30. L'examen est totalement gratuit, vous n'avez aucune avance de frais à faire. [...] C'est une succession d'examens biologiques, dentaire, médical complet, paramédical, entretien sur le mode vie, etc. (Source : https://www.aide-sociale.fr/bilan-sante-cpam-gratuit/). Il est considéré que ces examens ne peuvent pas être réalisés à distance.
1117 = CONSULTATION DES SPECIALISTES COTEE C2	Consultation dont l'équivalence est connue	1191 = TÉLÉCONSULTATION TOUTES SPÉCIALITÉS	9,9/10	*APY (ex C2, 5) en Neuro et psy : Avis ponctuel de consultant pour les Psys, neuro-Psys, Neuros : consultation pour avis (délai de 4 mois sans voir le patient sauf si nécessité d'1 ou 2 CNPSY dans les semaines suivantes) *APC (ex C2) : Avis ponctuel de consultant Spé qualifié (y compris gen). (Sources : https://www.syndicatavenispe.fr/exercice-pro/durant-la-vie-professionnelle/cotation-des-consultations/ ; https://www.fmfpro.org/tarifs-medecins-specialistes-metropole.html)
1118 = CONSULTATION DES PSYCHIATRES COTEE C2,5				
1131 = MAJORATION CONSULTATION APPAREILLAGE	Consultation sans équivalence	-	9/10	"Cette consultation s'effectue conjointement avec le médecin de MPR et l'appareilleur et se déroule en plusieurs temps : Indication de l'appareillage ; Après examen clinique et paraclinique, éventuellement complétés d'examen spécifique à l'appareillage (prise d'empreinte...). L'indication d'appareillage est posée ; Prescription de l'appareillage ; Réalisation de l'appareillage ; Suivi du patient appareillé : Les patients appareillés nécessitent une surveillance clinique voir paraclinique régulière afin d'adapter l'appareillage aux changements de morphologie ou à l'évolution de la pathologie." (Source : http://www.nantes-mpr.com/autres_rubriques/appareil.html). Il est considéré que ces examens ne peuvent pas être réalisés à distance.

FIGURE 19 : Référentiel d'équivalence entre les prestations de consultation et de téléconsultation (1/2)

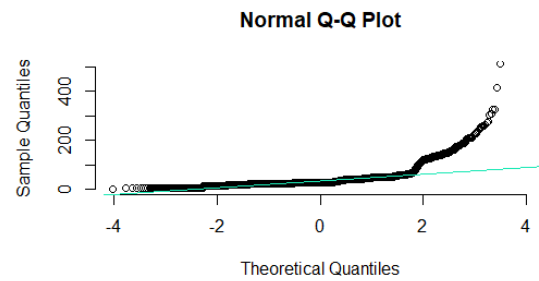
Prestation de Consultation	Catégories	Correspondance	Degré de confiance	Commentaires
1114 = CONSULTATION SPECIFIQUE CARDIOLOGIE	Consultation sans équivalence	-	7/10	Cette consultation peut être considérée comme complexe (ou très complexe), ce qui n'admet pas d'équivalent parmi les actes de téléconsultation. (Source : https://www.ameli.fr/assure/remboursements/rembourse/consultations-telemedecine/metropole#text_35809)
1115 = CONSULTATION BILAN	Consultation sans équivalence	-	9/10	"Il est possible de prétendre au bilan santé de la CPAM tous les 5 ans. L'examen est personnalisé en fonction de l'âge, des antécédents familiaux, des habitudes de vie et des facteurs de risque. En moyenne, la consultation dure 2H30. L'examen est totalement gratuit, vous n'avez aucune avance de frais à faire. [...] C'est une succession d'examens biologiques, dentaire, médical complet, paramédical, entretien sur le mode vie, etc. (Source : https://www.aide-sociale.fr/bilan-sante-cpam-gratuit/). Il est considéré que ces examens ne peuvent pas être réalisés à distance.
1117 = CONSULTATION DES SPECIALISTES COTEE C2	Consultation dont l'équivalence est connue	1191 = TÉLÉCONSULTATION TOUTES SPÉCIALITÉS	9,9/10	*APY (ex C2, 5) en Neuro et psy : Avis ponctuel de consultant pour des Psys, neuro-Psys, Neuros : consultation pour avis (délai de 4 mois sans revoir le patient sauf si nécessité d'1 ou 2 CNPSY dans les semaines suivantes) *APC (ex C2) : Avis ponctuel de consultant Spe qualifié (y compris gen). (Sources : https://www.sindicatavenirspe.fr/exercice-pro/durant-la-vie-professionnelle/cotation-des-consultations/ ; https://www.fmfpro.org/tarifs-medecins-specialistes-metropole.html)
1118 = CONSULTATION DES PSYCHIATRES COTEE C2,5				
1131 = MAJORATION CONSULTATION APPAREILLAGE	Consultation sans équivalence	-	9/10	"Cette consultation s'effectue conjointement avec le médecin de MPR et l'appareilleur et se déroule en plusieurs temps : Indication de l'appareillage ; Après examen clinique et paraclinique, éventuellement complétés d'examen spécifique à l'appareillage (prise d'empreinte...), l'indication d'appareillage est posée ; Prescription de l'appareillage ; Réalisation de l'appareillage ; Suivi du patient appareillé : Les patients appareillés nécessitent une surveillance clinique voir paraclinique régulière afin d'adapter l'appareillage aux changements de morphologie ou à l'évolution de la pathologie. (Source : http://www.nantes-mpc.com/autres_rubriques/appareil.htm). Il est considéré que ces examens ne peuvent pas être réalisés à distance.
1132 = MAJORATION CONSULTATION ENDOCRINO	Consultation dont l'équivalence est connue	1191 = TÉLÉCONSULTATION TOUTES SPÉCIALITÉS	9,5/10	Equivalence existante avec la facturation de l'acte Téléconsultation (TC) : "Téléconsultation effectuée par un médecin spécialiste (hors médecin généraliste, psychiatre, neurologue, neuropsychiatre)." (Source : https://solidarites-sante.gouv.fr/IMG/pdf/guide_facturation_tlm_en_etablissement_de_sante.pdf)
1134 = MAJORATION PREMIERE CONSULTATION FAMILLE	Consultation sans équivalence	-	5/10	La seule correspondance trouvée dans la littérature est la suivante : il s'agit d'une "consultation annuelle de synthèse familiale pour le psychiatre et pédopsychiatre pour un enfant présentant une pathologie psychiatrique grave relevant d'une ALD" faisant partie des consultations complexes de la NGAP. Ainsi, selon le site officiel de Ameli, il n'existe pas d'équivalent parmi les actes de téléconsultation. (Source : https://www.csmf.org/niveau-3-consultations-complexes/)
1135 = MAJORATION CONSULTATION ANNUELLE FAMILLE				
1140 = CONSULTATION SPECIFIQUE DE DEPISTAGE	Consultation sans équivalence	-	5/10	Source : schéma à la page 2 : https://www.mgfrance.org/images/utilitaires-medicaux/ppp-presentation.pdf
1148 = REMUNERATION FORFAITAIRE PAR CONSULTATION POUR LE SUIVI DES PERSONNES AGEES	Consultation dont l'équivalence est connue	1191 = TÉLÉCONSULTATION TOUTES SPÉCIALITÉS	8/10	"Cette rémunération s'adresse désormais aux médecins pour le suivi des patients de plus de 80 ans dont ils ne sont pas le médecin traitant. La nouvelle « MPA » (majoration personnes âgées) concerne l'ensemble des médecins libéraux conventionnés pour les patients dont ils ne sont pas le médecin traitant, quelle que soit leur spécialité." (Source : https://www.ameli.fr/medecin/exercice-liberal/remuneration/consultations-actes/tarifs/tarifs-generalistes/majoration-personnes-agees#:~:text=La%20r%C3%A9mun%C3%A9ration%20MPA%20est%20calcul%C3%A9e,de%20patients%20par%205%20%E2%82%AC.)
1153 = MAJORATION CONSULTATION POUR INSUFFISANT CARDIAQUE APRES HOSPITALISATION	Consultation dont l'équivalence est inconnue	3105 = TÉLÉCONSULTATION DOMICILE	7/10	"MSH : Majoration pour consultation de suivi de sortie d'hospitalisation. Pendant la période de l'épidémie du Covid-19, cet acte est réalisable en téléconsultation et remboursé à 100%. Dans le logiciel de facturation, sélectionner l'exonération de type soins particuliers exonérés". (Source : https://omniprat.org/fiches-pratiques/majoration-consultation-de-suivi-de-sortie-d-hospitalisation-msh/#:~:text=Il%20%C3%A9value%20l%20niveau%20d,reconnaissance%20des%20signes%20d'alarme)
1154 = MAJORATION CONSULTATION SUIVI APRES HOSPITALISATION PATIENTS A FORTE COMORBIDITE				
1168 = CONSULTATION DE CONTRACEPTION ET PREVENTION	Consultation sans équivalence	-	9/10	"Première consultation de contraception et de prévention des maladies sexuellement transmissibles pour les jeunes filles de 15 à 18 ans dénommée par le médecin généraliste, le gynécologue ou le pédiatre sous réserve de respecter le TO" et elle fait partie des consultations complexes (Source : https://www.csmf.org/niveau-3-consultations-complexes/)
1434 = PREVENTION BUCCO-DENTAIRE: CONSULTATION - MATER	Consultation sans équivalence	-	8,5/10	"La femme enceinte peut bénéficier de cet examen de prévention bucco-dentaire à compter du 4e mois de grossesse jusqu'au 12e jour après l'accouchement. Cet examen comprend : une anamnèse ; un examen bucco-dentaire ; une action de prévention et d'éducation sanitaire pour la femme et le futur enfant. Il peut être complété si nécessaire par : des radiographies intrabucales ; l'établissement d'un programme de soins." (Source : https://www.ameli.fr/chirurgien-dentiste/exercice-liberal/prescription-prise-charge/prise-charge-situation-type-soin/soin-ebd-femmes-enceintes). Il est alors considéré qu'il n'existe pas d'équivalent parmi les actes de téléconsultation.
1438 = BUCCO DENT. CONSULTATION COMPLEXE MATERNITE				
1439 = BUCCO DENT. CONSULTATION ET RADIO PANORAMIQUE MATERNITE				
1922 = CONSULTATION	Consultation dont l'équivalence est connue	1191 = TÉLÉCONSULTATION TOUTES SPÉCIALITÉS	9,5/10	Les majorations de facturation des TC effectuées pendant la nuit ou en des jours fériés existent et sont bien réglementées, ainsi l'hypothèse d'équivalence sera retenue. (Source : https://www.fmfpro.org/la-telemedecine-pour-les-nuls-les-cotations.html)
1933 = MAJORATION CONSULTATION REGULEE DE NUIT				
1934 = MAJORATION CONSULTATION REGULEE MILIEU DE NUIT				
1935 = MAJORATION CONSULTATION REGULEE DIMANCHE, JOURS FERIES ET ASSIMILES				
1941 = MAJORATION CONSULTATION REGULEE SAMEDI APRES MIDI				
1954 = PARTICIPATION ASSURE CONSULTATIONS ET SOINS EXTERNES (C2S + AME)	Consultation dont l'équivalence est connue	3104 = TÉLÉCONSULTATION LIEU DÉDIÉ	9,5/10	"La participation forfaitaire de 24 € s'applique aux actes pratiqués dans un cabinet médical, un centre de santé, un établissement de santé de type hôpital ou clinique ou dans le cadre de consultations externes (ambulatoires). Le forfait de 24 € est à la charge du patient lorsque le tarif des soins qu'il reçoit est : *Supérieur ou égal à 120 € dans la Classification commune des actes médicaux (CCAM) ; *Supérieur ou égal à un coefficient de 60, dans la Nomenclature générale des actes professionnels (NGAP)". (Source : https://www.previsima.fr/question-pratique/quest-ce-que-la-participation-forfaitaire-de-18.html#:~:text=La%20participation%20forfaitaire%20de%2024%20%E2%82%AC%20s'applique%20aux%20actes,de%20consultations%20externes%20(ambulatoires).)
1955 = PARTICIPATION ASSURE CONSULTATIONS ET SOINS EXTERNES (REGIME LOCAL)				
2217 = FORFAIT CONSULTATION URGENGE HOPITAL	Consultation dont l'équivalence est connue	3104 = TÉLÉCONSULTATION LIEU DÉDIÉ	7/10	Les conditions de facturation des téléconsultations et téléexpertises pour patients admis aux urgences sont bien présentes dans le guide de facturation de la téléconsultation. (Source : https://solidarites-sante.gouv.fr/IMG/pdf/guide_facturation_tlm_en_etablissement_de_sante.pdf)
2336 = FORFAIT POUR CONSULTATION EN CENTRE MEDICO-PSYCHO PEDAGOGIQUE	Consultation dont l'équivalence est connue	3104 = TÉLÉCONSULTATION LIEU DÉDIÉ	9/10	C'est une consultation effectuée par un médecin spécialiste dans un lieu dédié.
2414 = CONSULTATION IVG	Consultation dont l'équivalence est connue	1191 = TÉLÉCONSULTATION TOUTES SPÉCIALITÉS	8/10	"L'IVG médicamenteuse en ville peut être réalisée totalement ou partiellement par téléconsultation si la femme le souhaite et le praticien l'estime possible." (Source : http://www.cngof.fr/coronavirus-go-cngof/apercu?path=CONTRACEPTION%2B-%2BIVG%2B-%2BCOVID%252FIVGMEDCOVID.pdf&i=43097)
2426 = CONSULTATION IVG SPECIALISTE				

FIGURE 20 : Référentiel d'équivalence entre les prestations de consultation et de téléconsultation (2/2)

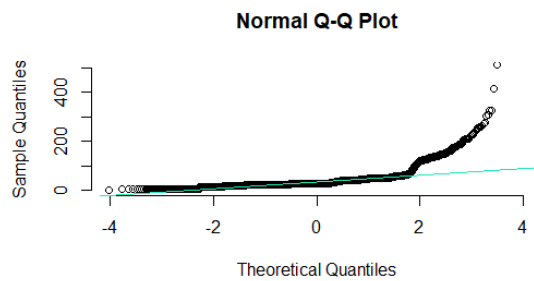
A.2 Annexe 2 : Modélisation GLM de la sinistralité en consultation



(a) Comparaison des deux lois



(b) QQ plot - Loi Lognormale



(c) QQ plot - Loi Gamma

FIGURE 21 : Adéquation graphique des modèles de coût moyen pour les prestations de consultation

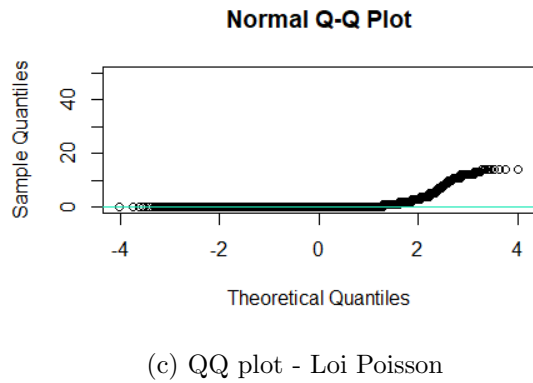
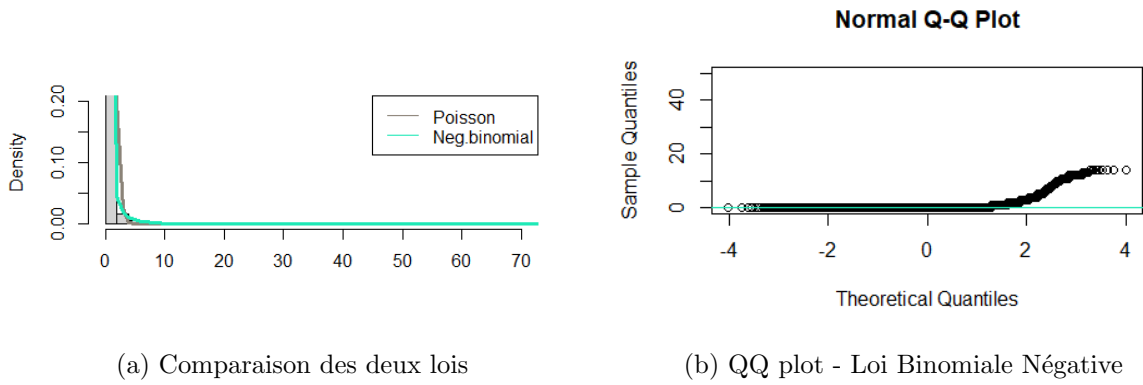


FIGURE 22 : Adéquation graphique des modèles de fréquence pour les prestations de consultation

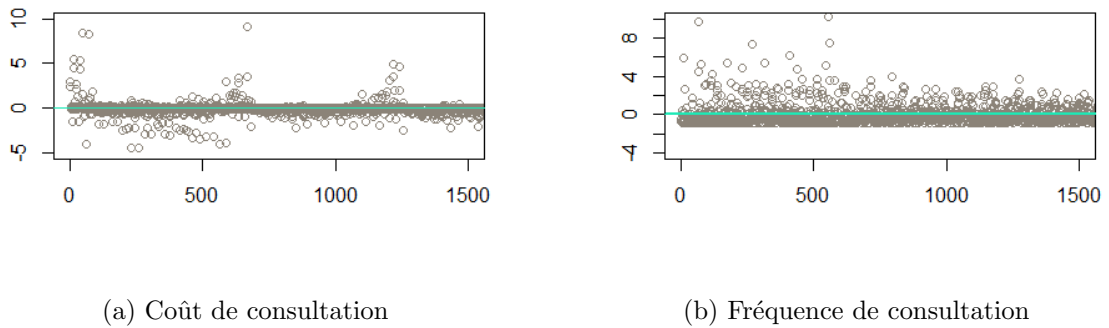


FIGURE 23 : Résidus de Pearson

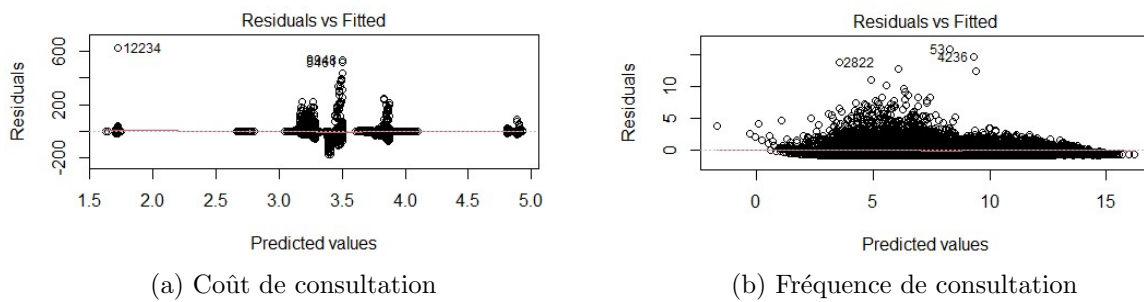


FIGURE 24 : *Residuals vs Fitted*

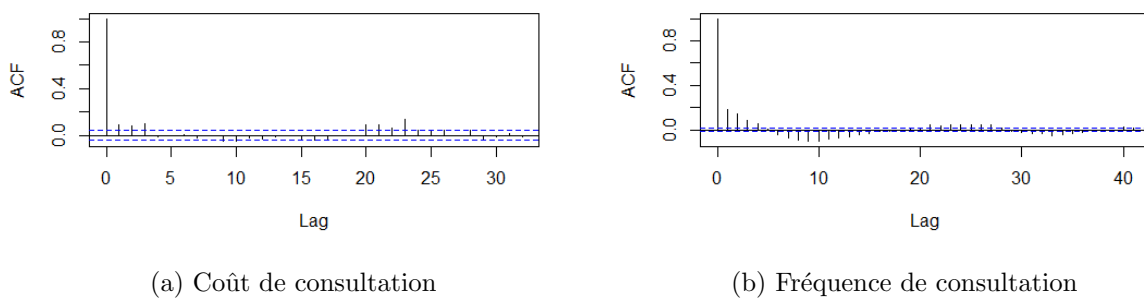


FIGURE 25 : Autocorrélogramme des résidus

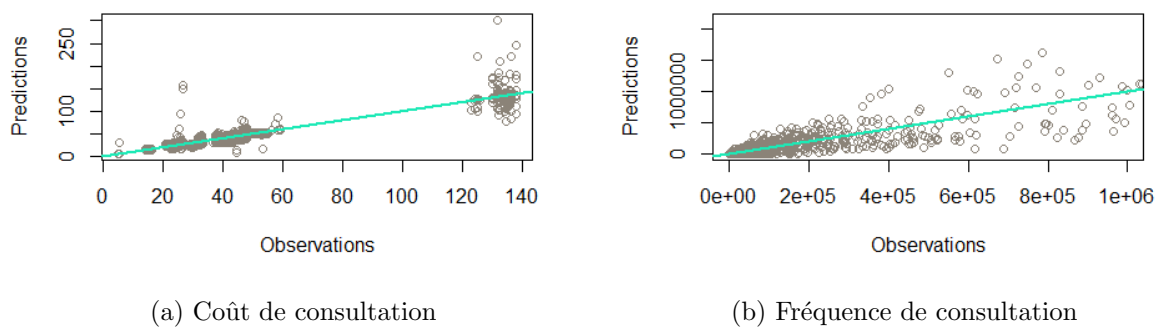


FIGURE 26 : Predictions vs Observations

A.3 Annexe 3 : Evolution des sinistralités de consultation et de téléconsultation entre 2018 et 2070, selon les scénarios élaborés

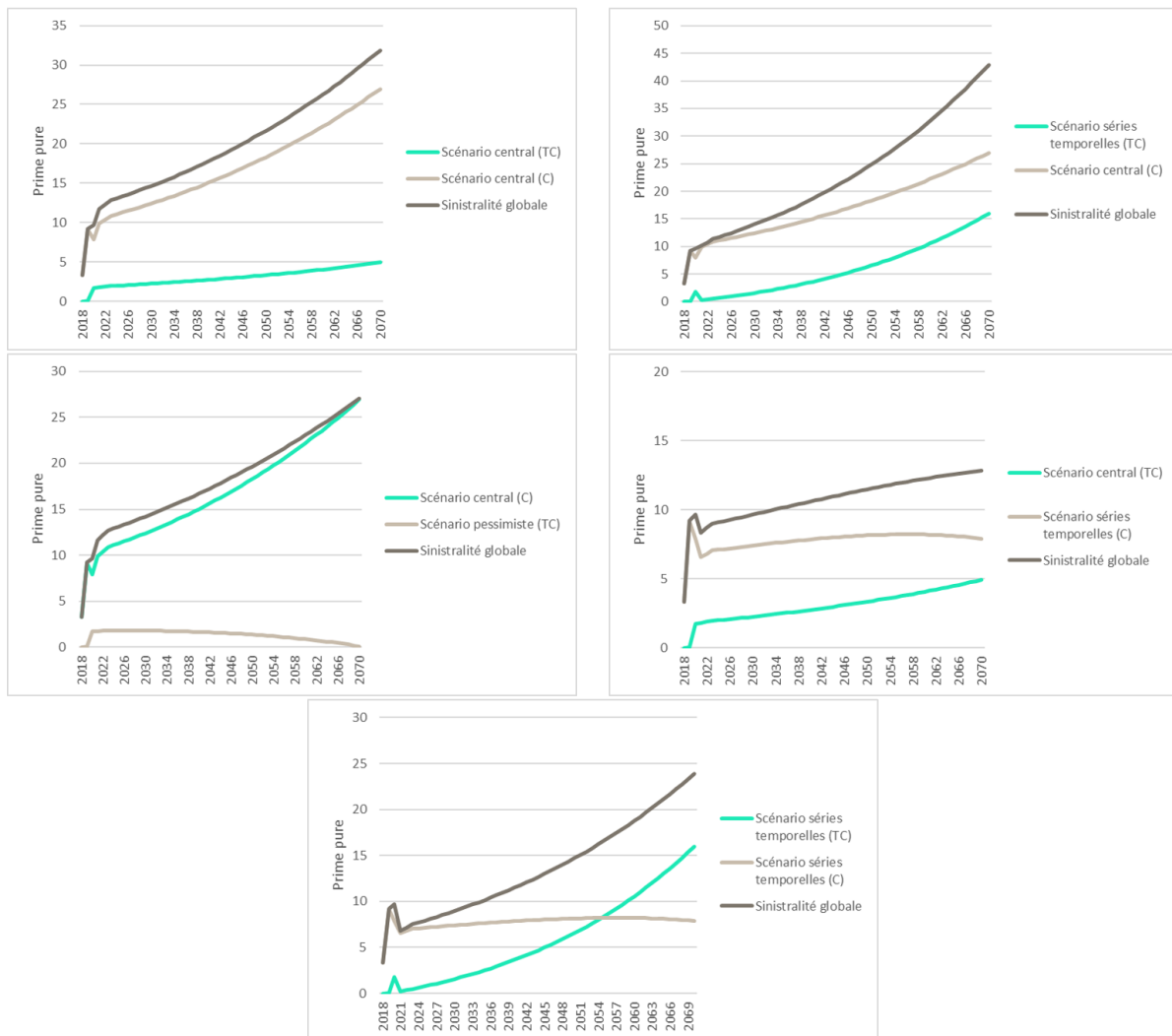


FIGURE 27 : Evolution des sinistralités globale et individuelles de consultation et de téléconsultation entre 2018 et 2070, selon les différents scénarios élaborés (1/2)

A.3. ANNEXE 3 : INFLUENCE DES SCÉNARIOS DE PROJECTION SUR LES SINISTRALITÉS FUTURES13

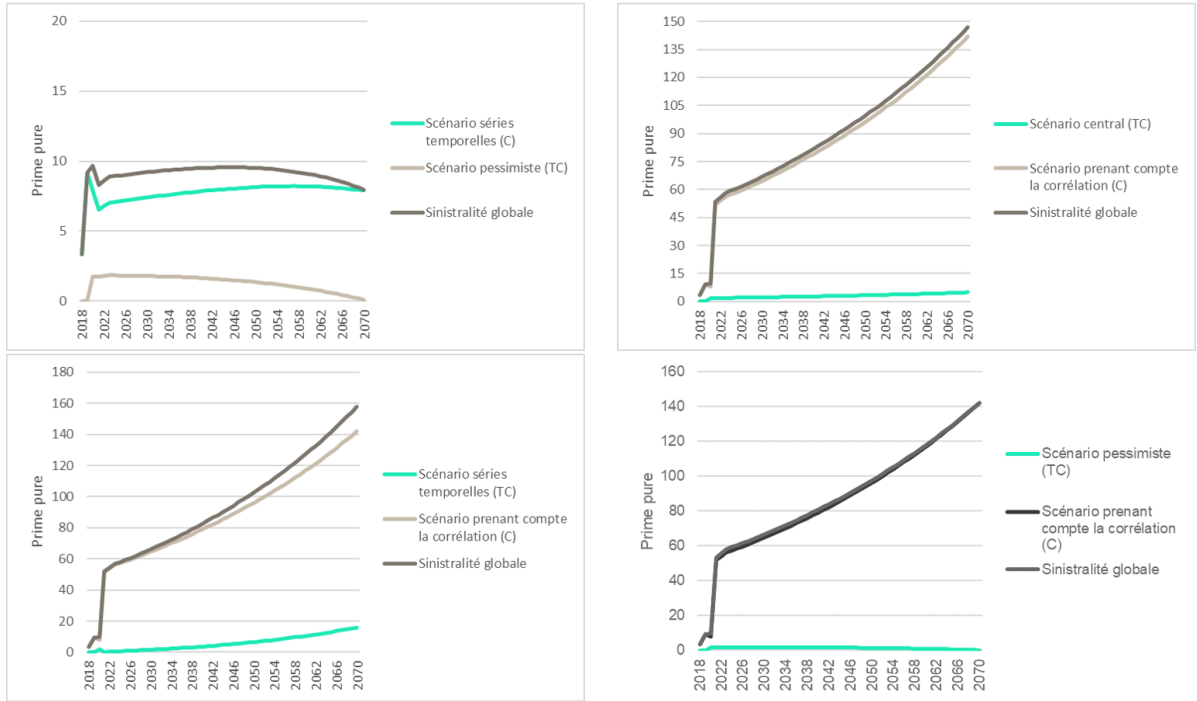


FIGURE 28 : Evolution des sinistralités globale et individuelles de consultation et de téléconsultation entre 2018 et 2070, selon les différents scénarios élaborés (2/2)