
Quel est l'effet de l'incitation financière pour répondre auprès d'une population de médecins généralistes ?

Maxime BERGEAT (*)

(*) Drees, Bureau des Professions de Santé

maxime.bergeat@sante.gouv.fr

Mots-clés. Analyse de données, apprentissage supervisé, données d'enquêtes, incitation financière, médecins généralistes.

Domaine. Collecte de données d'enquêtes.

Résumé

La Drees, en collaboration avec les Observatoires régionaux de santé (ORS) et les Unions régionales de professionnels de santé, médecins libéraux (URPS-ML) des régions Provence-Alpes-Côte d'Azur (Paca) et Pays de la Loire, a mis en place en 2018 un panel national de médecins généralistes libéraux, interrogés sur leur activité, leurs pratiques et leurs conditions de travail tous les six à neuf mois. Lors de chaque vague d'interrogation, les médecins répondants sont indemnisés pour le temps passé à répondre. Pendant la crise sanitaire de la Covid-19, le panel d'observation des pratiques et des conditions d'exercice en médecine générale a été mobilisé à quatre reprises supplémentaires par rapport aux prévisions initiales, de manière notamment à suivre l'activité des médecins généralistes pendant l'épidémie. Pour ces interrogations supplémentaires, les médecins répondants n'ont pas été indemnisés pour leur réponse.

Dans cette communication, on étudie l'effet de l'incitation financière sur la propension à répondre des médecins interrogés. Dans la littérature, il est mis en évidence que l'utilisation d'une indemnisation financière conduit en général à des taux de réponse supérieurs, et souvent également à des profils de répondants plus diversifiés. Dans le cadre du panel d'observation des pratiques et des conditions d'exercice en médecine générale, après avoir constaté que les vagues d'enquête indemnisées donnent lieu à des taux de réponse plus importants, on dresse une typologie des profils-type de l'attrition grâce à des techniques d'analyse factorielle, mettant notamment en évidence une catégorie de médecins « difficiles à convaincre », pour qui le recours à une incitation financière pourrait susciter de meilleurs comportements de réponse.

Dans un second temps, on cherche à étudier l'effet conjoint des caractéristiques individuelles des médecins (sexe, âge, volume usuel d'activité, lieu d'exercice) et des variables décrivant la collecte (nombre de relances, modes de collecte utilisés, recours ou non à une indemnisation financière) afin de quantifier l'effet du recours à une indemnisation financière par rapport aux autres caractéristiques de la collecte des personnes interrogées. Pour cela, on utilise des modèles prédictifs de type *machine learning*, permettant de ne pas être contraint par des hypothèses restrictives sur les

variables incluses dans les modèles, et on utilise des méthodes de *feature importance* permettant de quantifier l'importance de chaque covariable pour prédire le comportement de réponse des médecins.

L'incitation financière semble jouer un rôle important pour prédire la réponse d'un médecin lors d'une vague d'interrogation, bien que moindre que d'autres caractéristiques de la collecte comme sa durée. Toutefois, le fait qu'on ait pas eu recours à une expérience randomisée, et que l'analyse soit effectuée durant un contexte sanitaire exceptionnel ayant fortement impacté l'activité médicale, ne permet pas de donner des conclusions très affirmées sur le rôle de l'incitation financière pour expliquer les comportements de réponse. La communication est conclue par une discussion autour des avantages et inconvénients du recours à des approches prédictives pour classer l'importance des covariables utilisées dans un modèle où on cherche à expliquer une variable d'intérêt : même si l'intérêt semble relativement limité dans le cas d'usage présenté dans cette communication, le recours à ces approches, qui sont simples à mettre en œuvre et ont le mérite de ne pas reposer sur des hypothèses fortes, peut être prometteur pour d'autres problématiques de la statistique publique.

Abstract

This paper analyses attrition in a survey of French general practitioners. Non-response is studied for six survey waves taking place in 2020 and 2021. First three attrition profiles are defined based on factor analysis. Then machine learning models are estimated in order to predict non-response with covariates based on both individual and survey characteristics. We use feature importance methods to rank covariates used in the models based on their predictive power. The goal of the analysis is to assess whether using financial incentives for survey respondents is useful in order to increase response rates.

Introduction

La Drees, en collaboration avec les Observatoires régionaux de santé (ORS) et les Unions régionales de professionnels de santé, médecins libéraux (URPS-ML) des régions Provence-Alpes-Côte d'Azur (Paca) et Pays de la Loire, a mis en place en 2018 un panel national de médecins généralistes libéraux, interrogés sur leur activité, leurs pratiques et leurs conditions de travail tous les six à neuf mois. Environ 3000 médecins sont inclus dans ce panel.

Lors de chaque vague d'interrogation, les médecins répondants sont indemnisés pour le temps passé à répondre (à hauteur de 25 euros, soit la base de remboursement d'une consultation médicale, pour une durée de questionnaire de 15 minutes). Pendant la crise sanitaire de la Covid-19, le panel d'observation des pratiques et des conditions d'exercice en médecine générale a été mobilisé à quatre reprises supplémentaires par rapport aux prévisions initiales, de manière notamment à suivre l'activité des médecins généralistes pendant l'épidémie. Pour ces interrogations supplémentaires, les médecins répondants n'ont pas été indemnisés pour leur réponse.

Dans cette communication, on cherche à utiliser cette expérience naturelle pour mesurer l'effet de l'indemnisation sur la propension à répondre à une vague d'interrogation du panel d'observation des pratiques et des conditions d'exercice en médecine générale. Plus généralement, on veut identifier l'impact de l'indemnisation comparativement à d'autres variables qui caractérisent à la fois l'effort porté dans la collecte (durée, nombre de relances, modes de collecte) et les personnes interrogées (caractéristiques sociodémographiques et de leurs lieux d'exercice).

Les enquêtes de la statistique publique française recourent rarement à l'indemnisation financière pour inciter les personnes interrogées à répondre. La pratique est toutefois plus répandue dans d'autres pays. Dans le cadre de l'enquête européenne sur le budget des ménages (enquête Budget

de famille à l’Insee), plusieurs pays indemnisent les ménages répondants (EUROSTAT, 2020). Par exemple, pour l’Enquête permanente sur le budget des ménages au Luxembourg, les ménages répondants sont indemnisés entre 65 € et 105 € selon la taille du ménage.

Il existe une littérature assez abondante à propos de l’effet de l’indemnisation financière sur la propension à répondre des personnes interrogées. La synthèse de différentes expérimentations, principalement américaines, effectuée dans (SINGER et YE, 2013), montre que l’utilisation d’indemnisation financière va généralement de pair avec des taux de réponse supérieurs. Plusieurs expériences randomisées comparent également l’efficacité d’une incitation prépayée, c’est-à-dire envoyée avant d’avoir reçu la réponse, plutôt que promise sous réserve d’avoir répondu. Dans les expériences discutées dans (SINGER et YE, 2013), les taux de réponse observés sont généralement plus élevés lorsque l’incitation est prépayée, par rapport à une incitation conditionnelle à la réponse. Par exemple, aux Pays-Bas et au Royaume-Uni, des expérimentations ont été conduites pour les personnes interrogées dans le cadre de l’Enquête emploi (*Labour Force Surveys*) : le fait d’envoyer des timbres gratuits aux personnes interrogées permet d’améliorer les taux de réponse observés (BARNES et collab., 2008). En France, une expérimentation conduite dans le cadre de l’enquête internationale PIAAC sur les compétences des adultes (BALLINI et collab., 2022) met en évidence un effet positif du recours à une indemnisation sur la propension à répondre des personnes interrogées. Par ailleurs, les auteurs semblent constater un effet de l’indemnisation sur les variables d’intérêt de l’enquête, avec des scores de compétence en moyenne plus élevés pour les individus ayant obtenu une indemnisation financière par rapport aux autres.

On peut également s’interroger sur l’effet de l’indemnisation financière sur la distribution des réponses obtenues. En effet, recourir à une indemnisation peut modifier la structure des réponses obtenues, soit directement (les personnes incitées sont de meilleure humeur, pouvant affecter leurs réponses), soit indirectement *via* la composition de l’échantillon. Pour ce qui est des effets de composition de l’échantillon, dans les différentes expérimentations présentées dans (SINGER et YE, 2013), il n’est pas mis en évidence que l’incitation financière détériore la composition de l’échantillon, en faisant que les individus les plus disposés à répondre le sont d’autant plus lorsqu’ils bénéficient d’une incitation financière. L’effet sur la composition de l’échantillon de répondants selon qu’il y ait ou non une indemnisation financière est discuté dans cette communication.

Plusieurs analyses sont consacrées aux effets de l’incitation sur les propension à répondre des professionnels de santé et des médecins (VANGHEEST et collab., 2007; CHO et collab., 2013). L’utilisation d’incitations monétaires semble avoir sauf exceptions un impact significatif sur l’augmentation des taux de réponse, ce qui n’est pas le cas des incitations non-monétaires. Par exemple, la méta-analyse présentée dans (CHO et collab., 2013) à partir de l’étude de 154 enquêtes auprès de professionnels de santé conclut à l’existence d’un impact significativement positif « toutes choses égales par ailleurs » pour le taux de réponse de l’enquête du fait d’avoir mis en place une incitation financière¹.

Dans cette communication, on cherche à mesurer l’effet du recours à une incitation financière sur la propension à répondre à une vague d’interrogation du Panel d’observation des pratiques et des conditions d’exercice en médecine générale. On s’intéresse ici à un effet de non-attribution plutôt qu’au mécanisme complet de non-réponse, dans le sens où chaque questionnaire n’est proposé qu’aux médecins ayant accepté en 2018 de faire partie du dispositif d’interrogation. Par ailleurs, le cadre théorique n’est pas celui d’une expérience randomisée, car certaines vagues d’interrogation font l’objet d’une indemnisation tandis que d’autres ne le sont pas : ainsi, l’ensemble des médecins interrogés se voit ou non proposer une indemnisation lors de chaque vague d’enquête. L’identification de l’effet propre de l’indemnisation financière est ainsi plus complexe à identifier. Quels sont les mécanismes de l’attrition dans une enquête auprès de médecins généralistes fran-

¹ Dans cette étude, l’unité d’analyse est l’enquête, et on contrôle notamment de la date et du lieu de réalisation de celle-ci, du ou des modes de collecte, des métiers des professionnels de santé enquêtés, et du nombre de relances effectuées.

gais? Quel est l'impact de l'incitation financière sur la réponse de ces médecins? Pour étudier ces questions, on propose une démarche en deux temps.

- D'abord, on dresse une typologie des profils d'attrition pour les six vagues d'enquête étudiées dans cette communication.
- On discute ensuite une méthode pour expliquer les comportements d'attrition, utilisant des approches prédictives de type *machine learning* pour classer les variables permettant d'expliquer la propension à répondre selon leur pouvoir prédictif.

1 Qui sont les répondants à chaque vague d'interrogation?

Dans un premier temps, on cherche à caractériser, grâce à une approche descriptive, l'attrition dans le panel d'observation des pratiques et des conditions d'exercice en médecine générale. Mis en place en 2018, il était initialement prévu d'interroger les médecins de ce panel une à deux fois par an sur leurs pratiques et leurs conditions d'exercice, sur des thèmes variés comme l'organisation du cabinet, l'utilisation des outils de e-médecine, l'impact de la baisse de la démographie médicale sur les pratiques, la prise en charge des soins non programmés, les pratiques de prévention ou l'antibiorésistance. Dans le cadre de la crise sanitaire liée à la Covid-19, des enquêtes supplémentaires ont été réalisées au cours de l'année 2020 pour étudier l'activité des médecins généralistes, ainsi que leurs opinions sur la gestion de la crise (MONZIOLS et collab., 2020; BERGEAT et collab., 2021).

Le champ de l'enquête comprend l'ensemble des médecins généralistes tirant tout ou partie de leurs revenus d'une activité libérale, installés au 1^{er} janvier 2018, n'exerçant pas un mode d'exercice particulier (MEP) exclusif et ayant au moins 200 patients « médecin traitant » au 1^{er} janvier 2018. Les médecins ont été tirés au sort dans le Répertoire Partagé des Professions de Santé (RPPS) apparié à des données de l'Assurance Maladie. Le tirage a été stratifié selon le sexe, l'âge en tranches (moins de 50 ans, entre 50 et 59 ans et 60 ans ou plus), le volume d'activité compris comme le nombre de consultations et visites réalisées en 2017 en tranches (inférieur au premier quartile, entre le premier et le troisième quartile et au-dessus du troisième quartile), la région et la densité médicale de la zone d'exercice. Les médecins exerçant en régions Provence-Alpes-Côte d'Azur et Pays de la Loire sont sur-représentés tout comme les médecins exerçant en zones sous-denses. Un médecin est considéré comme exerçant en zone sous-dense lorsque l'indicateur d'Accessibilité Potentielle Localisée (APL) du territoire de vie-santé dans lequel il exerce est inférieur à 2,8 consultations/an/habitant. Au total, 3 304 médecins ont accepté de participer à ce quatrième panel, pour 12 022 médecins sollicités.

Pour chaque vague d'interrogation, les médecins ayant accepté de participer au panel sont interrogés de façon privilégiée grâce à une collecte par Internet, suivie de relances téléphoniques après quelques semaines de terrain pour ceux n'ayant pas encore répondu. Une partie de l'échantillon (représentant un sixième des médecins échantillonnés) est interrogée uniquement par téléphone, de manière à pouvoir analyser l'effet du mode de collecte sur le profil des répondants et sur les réponses obtenues. Dans cette communication, les médecins de cet échantillon dit *full CATI* ne sont pas inclus dans l'analyse, afin de ne pas attribuer des effets liés au mode de collecte à l'indemnisation financière. Par ailleurs, les médecins qui n'ont pas été contactés lors de la vague d'interrogation du printemps 2021 (car sortis du champ depuis leur inclusion dans le Panel, suite par exemple à un départ en retraite, ou ayant choisi de se désinscrire, l'enquête ne revêtant pas de caractère obligatoire) ne sont pas inclus dans cette analyse, qui porte sur 2686 médecins.

On considère six vagues d'interrogation réalisées entre décembre 2019 et juillet 2021. Parmi ces enquêtes, deux ont fait l'objet d'une indemnisation et quatre n'ont pas été indemnisées, comme indiqué dans la Figure 1.



FIGURE 1 : Calendrier des vagues d'enquête du Panel d'observation des pratiques et des conditions d'exercice en médecine générale

1.1 Taux de réponse selon la vague d'interrogation

Dans un premier temps, on étudie simplement les taux de réponse observés pour chaque vague d'interrogation afin de mesurer l'attrition. On compare ces taux de réponse observés avec les taux de réponse qu'on pourrait attendre selon une hypothèse d'attrition de 10 % par vague d'interrogation, correspondant aux hypothèses faites lors de la préparation de l'enquête et à ce qui a été observé lors de l'édition précédente du panel d'observation des pratiques et des conditions d'exercice en médecine générale².

Les taux de réponse sont présentés dans la Figure 2. On constate effectivement que, pour les vagues d'interrogation ne donnant pas lieu à une indemnisation, les taux de réponse observés sont inférieurs aux taux de réponse qu'on aurait pu attendre avec une hypothèse d'attrition de 10 % entre la vague d'interrogation de l'hiver 2019-2020 (vague 3) et les vagues suivantes. Pour l'interrogation menée au printemps 2021 et qui a donné lieu à une indemnisation, on constate que la situation est inverse, avec +6 points d'écart pour le taux de réponse observé par rapport à ce qui est attendu avec l'hypothèse sur l'attrition³. Ainsi, même si on est pas ici dans le cadre d'une expérience randomisée, comme une indemnisation a été proposée à l'ensemble des personnes interrogées lors de chaque vague d'enquête indemnisée, et que les conditions de collecte des données sont différentes entre chaque interrogation, on peut penser en première approche que le recours à une indemnisation financière permet de stimuler les taux de réponse observés.

1.2 Caractérisation des répondants à chaque vague d'interrogation

Pour préciser ce constat, étudions désormais la structure des personnes répondantes pour chaque vague d'interrogation. Les résultats présentés ici sont pondérés en utilisant le poids corrigé de la non-réponse et calé lors de la première vague d'interrogation du panel. Ainsi, la déformation de la structure des répondants entre chaque vague d'interrogation correspond seulement à la

² Toutefois, on peut noter que ces hypothèses ont été effectuées dans un contexte avec des vagues d'interrogation moins nombreuses et moins fréquentes, en effet les vagues dites « Flash Covid » représentent des interrogations supplémentaires par rapport à ce qui avait été prévu initialement.

³ On rappelle qu'on s'intéresse ici à un taux de réponse calculé sur la population des personnes contactées pour l'enquête et qui ont accepté de participer au panel. Ainsi, seul le phénomène d'attrition est observé.

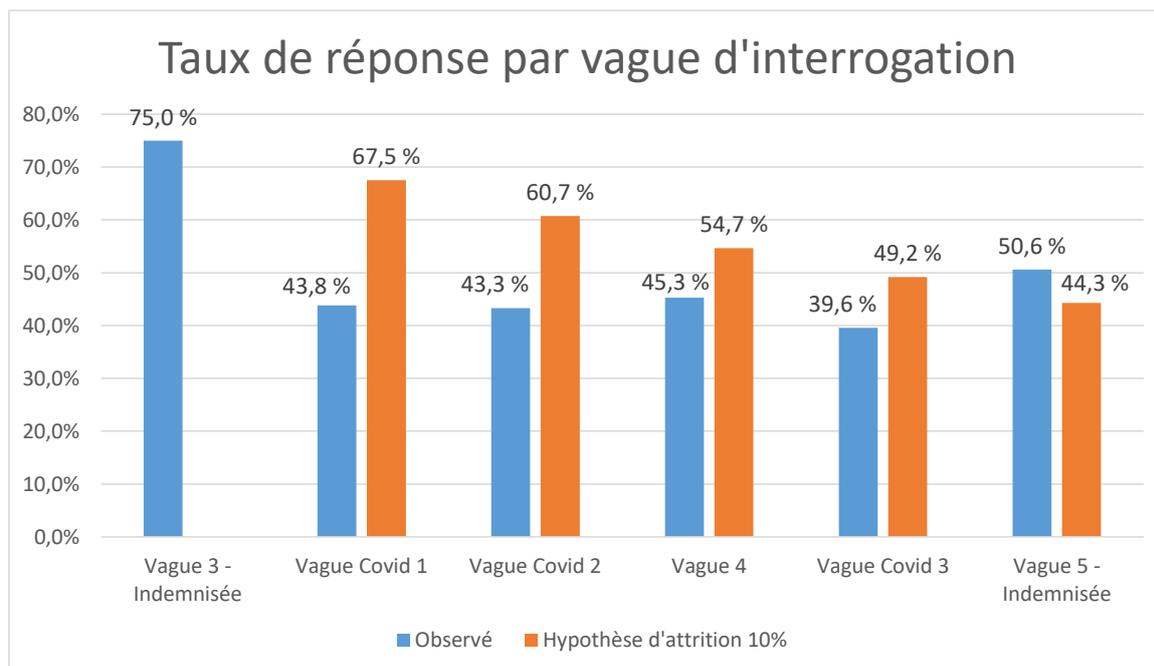


FIGURE 2 : Taux de réponse observés et prévisionnels pour les vagues d'interrogation du panel en 2020 et 2021

déformation **due à l'attrition**, et non à l'inclusion dans le panel⁴. Les variables utilisées pour le traitement de la non-réponse et le calage sur marges sont le sexe, l'âge du praticien, son volume d'activité, sa région d'exercice (Provence-Alpes-Côte d'Azur, Pays de la Loire ou autre région, ces deux premières régions étant sur-représentées dans l'échantillon afin de permettre des exploitations au niveau régional) et l'exercice ou non dans une zone à faible densité médicale.

On constate ici (voir Figures 3 et 4) que les échantillons des vagues indemnisées semblent moins déformés par rapport aux échantillons des vagues d'interrogation où aucune indemnisation n'est proposée aux médecins répondants. En particulier, les répondants des vagues où il n'y a pas eu d'indemnisation sont plus souvent des femmes et sont en moyenne plus jeunes, par rapport aux autres vagues d'interrogation ayant donné lieu à une indemnisation des répondants. Par exemple, la base de sondage est constituée à 39 % de femmes, et les deux vagues indemnisées correspondent à un échantillon féminisé à respectivement 41 % et 42 %, lorsqu'on pondère les données sans prise en compte de l'attrition. Ainsi, il semblerait à première vue que l'indemnisation permet d'obtenir une structure de l'échantillon de répondants moins déformée que si on n'en utilise pas, ce qui est un résultat assez souvent constaté dans la littérature d'après la méta-analyse de SINGER et YE (2013). Ainsi, les répondants supplémentaires obtenus grâce au recours à une indemnisation semblent à première vue plutôt être des répondants différents des autres.

Pour préciser l'étude des comportements de réponse sur la période décembre 2019-juillet 2021, on réalise une **Analyse en Composantes Principales (ACP)** afin de synthétiser l'information portée par les variables dichotomiques indiquant que la personne interrogée a ou non répondu lors de chaque vague d'interrogation. Les variables supplémentaires introduites ici sont le sexe et la tranche d'âge des médecins, leur volume d'activité usuel, ainsi que leur région d'exercice et le fait qu'il exercent ou non dans une zone sous-dense. Après avoir donné des éléments sur

⁴ L'analyse présentée dans cette section a été répétée sans pondérer les données, et les messages obtenus sont similaires à ceux présentés ici.

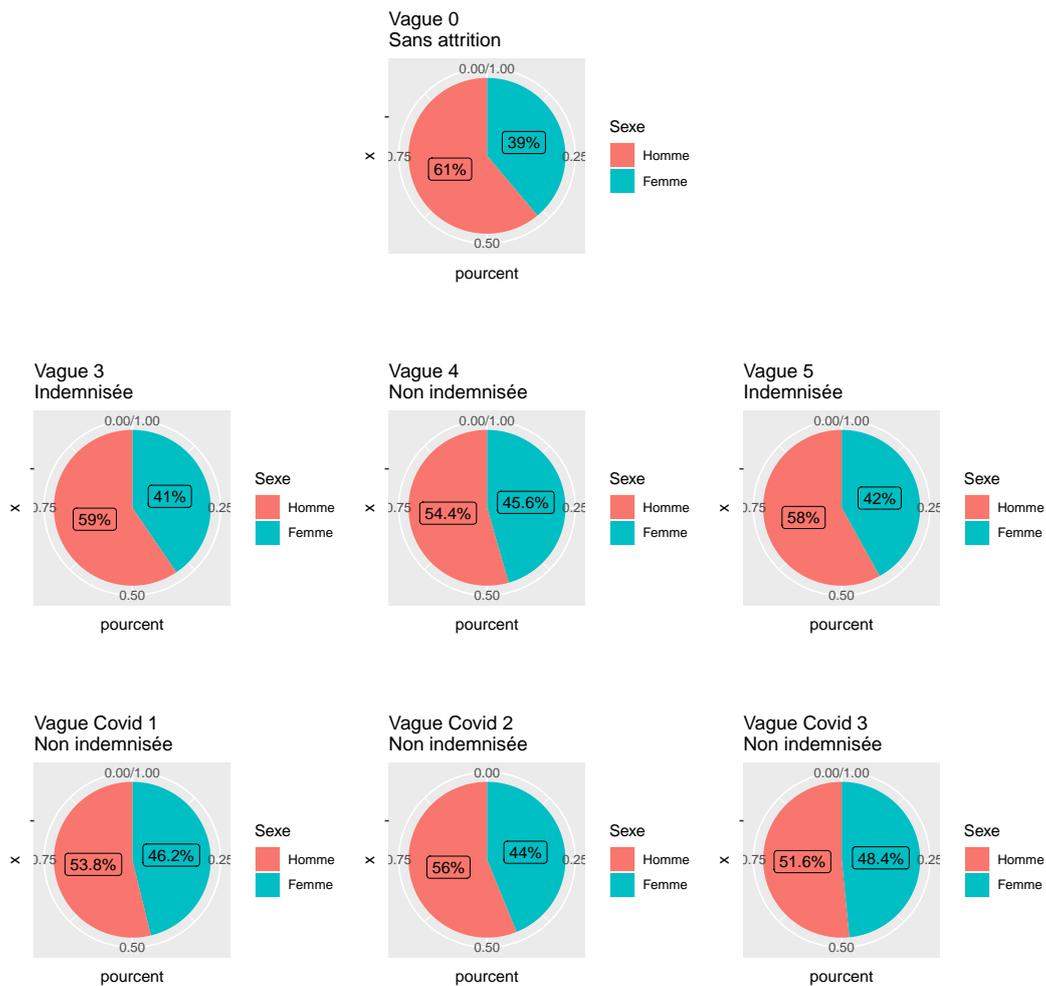


FIGURE 3 : Distribution des répondants à chaque vague d'interrogation selon le sexe

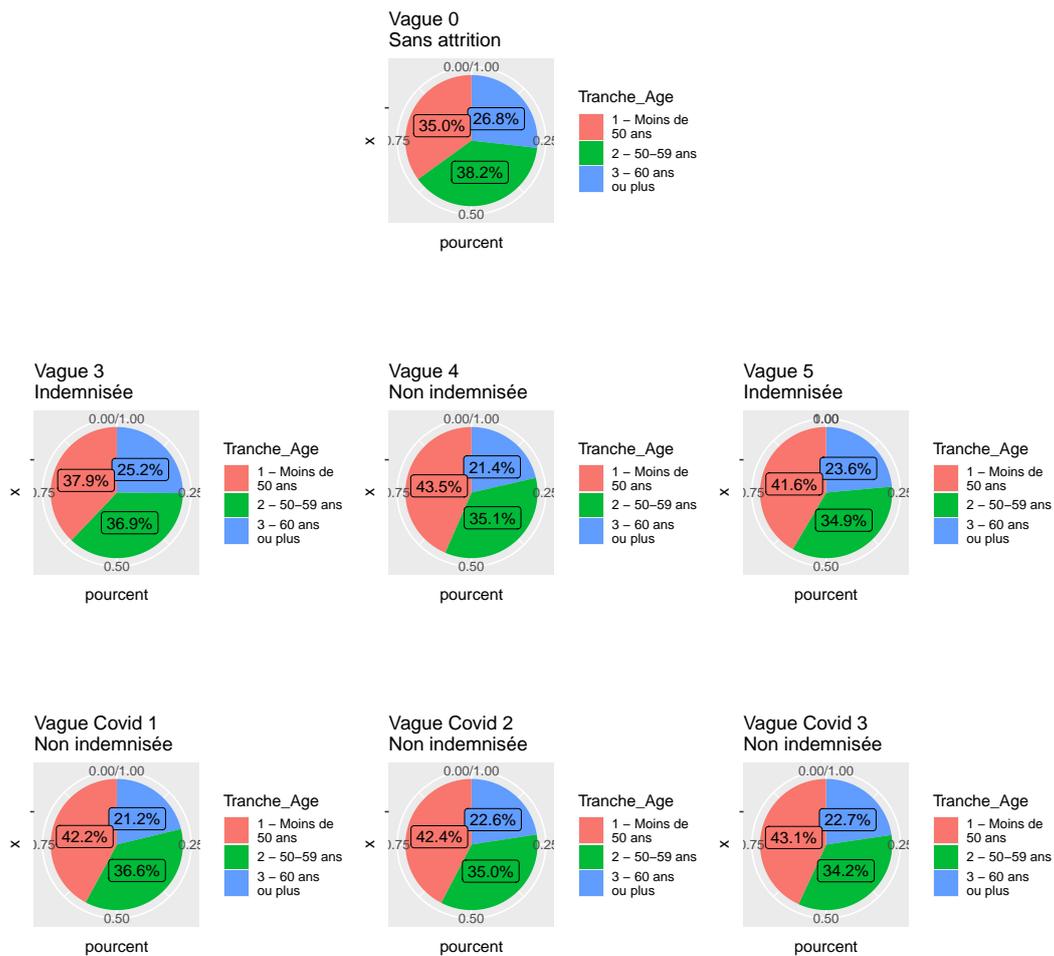


FIGURE 4 : Distribution des répondants à chaque vague d'interrogation selon la tranche d'âge

l'analyse factorielle ici réalisée, on réalise une **Classification Ascendante Hiérarchique** à partir des composantes principales issues de l'ACP. Cela permet de définir une typologie des profils d'attrition des médecins généralistes interrogés, qu'on peut mettre en regard de leurs caractéristiques individuelles et de celles de leur lieu d'exercice.

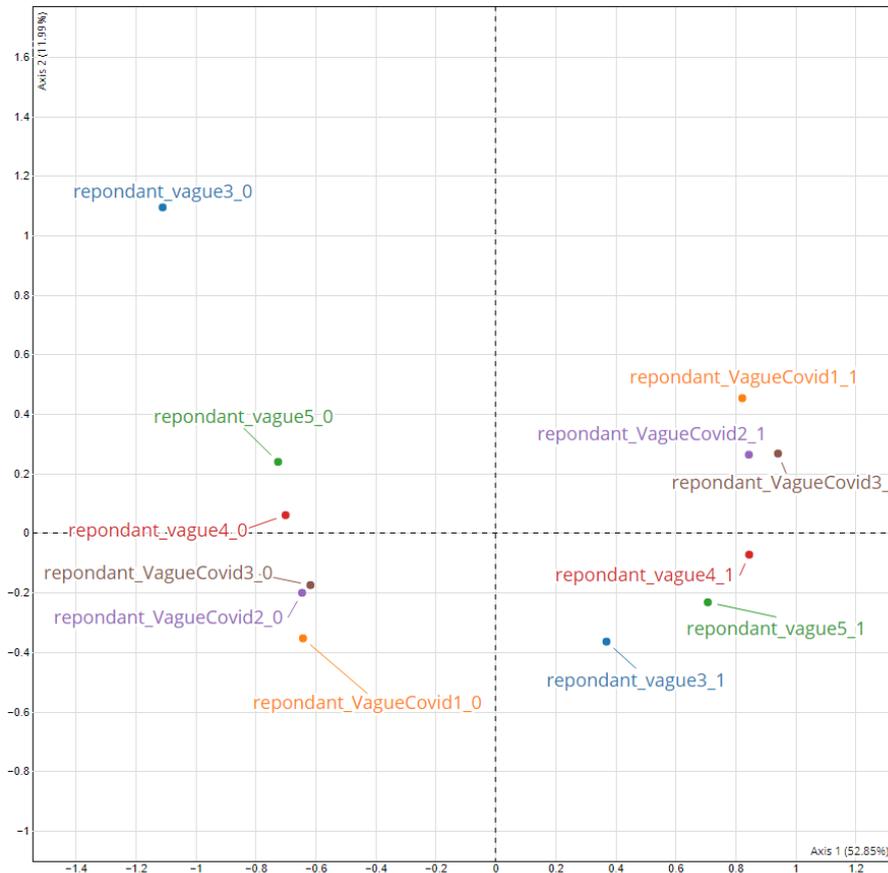


FIGURE 5 : Projections des variables actives dans le premier plan factoriel issu de l'ACP

On représente dans la Figure 5 les projections des variables actives dans le premier plan issu de l'analyse factorielle, qui porte près de deux tiers de l'information contenue dans les six variables de réponses incluses dans l'analyse multivariée (65 % de l'inertie). Dans ce premier plan factoriel, on constate :

- Que l'opposition entre les quadrants de gauche et ceux de droite semble s'opérer selon la réponse aux différentes vagues d'interrogation. Ainsi, les médecins qui se projettent à droite dans le premier plan factoriel ont de meilleurs comportements de réponse que les autres.
- Que les quadrants du haut et du bas à droite semblent opposer les individus qui répondent aux vagues non indemnisées (vagues « Flash Covid » et vague 4) et ceux qui répondent aux vagues traditionnelles indemnisées (vague 3 et vague 5), avec un questionnaire plus long et à propos de thèmes plus structurels sur l'activité et les pratiques des médecins généralistes. Par ailleurs, les vagues 3 et 5 ont fait l'objet d'un effort de collecte plus important (terrain téléphonique plus long, en particulier).

On peut préciser la description des comportements d'attrition en projetant des variables supplémentaires dans le premier plan factoriel issu de l'Analyse en Composantes Principales. Les résultats sont présentés dans la Figure 6.

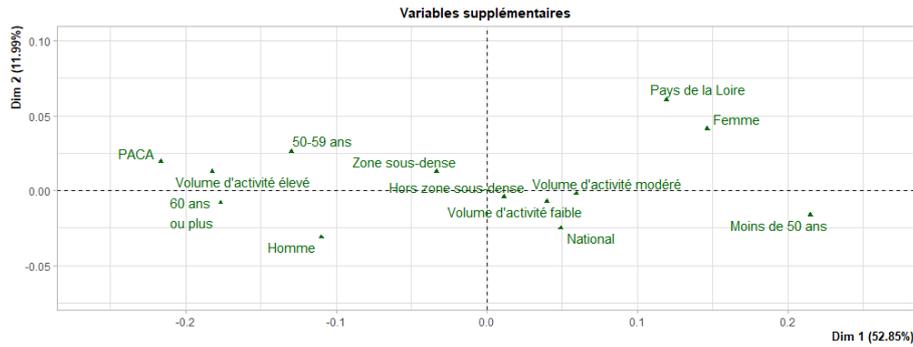


FIGURE 6 : Projections de variables supplémentaires dans le premier plan factoriel issu de l'ACP

On observe des différences qui portent principalement sur le premier axe, décrivant les bons comportements de réponse des médecins interrogés. On retrouve ici des résultats classiques sur les comportements de réponse, indiquant que les femmes et les plus jeunes personnes interrogées répondent plus fréquemment lorsqu'elles sont sollicitées. Il y a également une corrélation entre le fait de répondre peu et d'avoir un volume d'activité élevé en 2017, ce qui peut traduire une disponibilité moindre pour répondre aux enquêtes. Les différences observées sur l'axe vertical sont moins marquées : on peut par exemple noter que les médecins femmes se projettent en moyenne en haut à droite dans le premier plan factoriel, et leurs confrères en bas à gauche : cela pourrait traduire une plus grande difficulté à atteindre les médecins hommes par rapport aux autres, ces derniers pourraient notamment être plus sensibles aux incitations financières.

Pour compléter l'analyse, on procède à une classification ascendante hiérarchique (CAH), afin de regrouper les personnes interrogées avec des comportements proches par rapport à l'attrition étudiée dans cette communication.

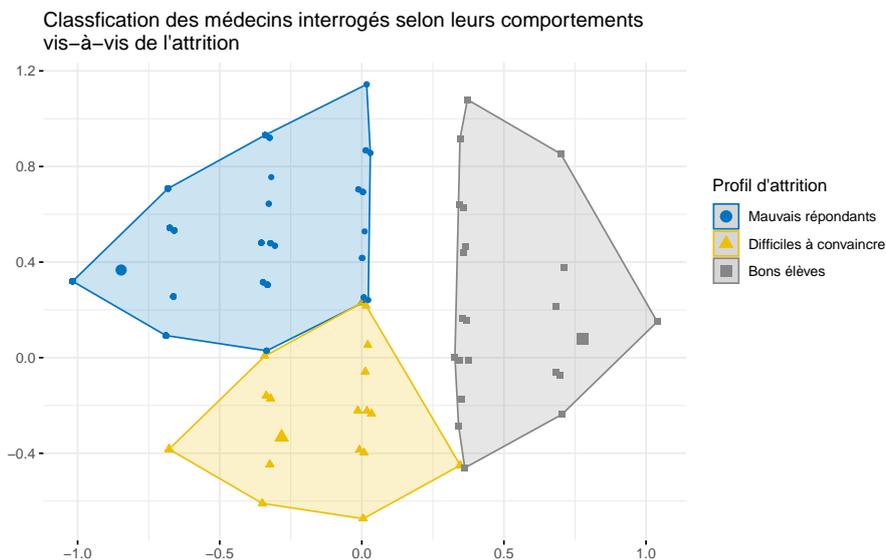


FIGURE 7 : Trois profils de personnes interrogées pour l'attrition

On représente les résultats de la CAH dans la Figure 7, en considérant 3 *clusters* qui correspondent à trois profils d'attrition.

Des éléments décrivant la composition des profils d'attrition sont également donnés dans la Table 1. D'abord, on retrouve les **mauvais répondants**, qui représentent un quart des personnes

Indicateur	Mauvais répondants (24.3 %)	Difficiles à convaincre (36.2 %)	Bons élèves (39.5 %)	Ensemble des médecins interrogés
Part de femmes	36.7 %	39.1 %	50.2 %	42.9 %
Âge moyen au 1 ^{er} janvier 2018	54 ans 4 mois	52 ans 4 mois	49 ans	51 ans 6 mois
Nombre moyen de réponses aux 2 vagues indemnisées	0.14	1.36	1.84	1.26
Nombre moyen de réponses aux 4 vagues non indemnisées	0.36	0.8	3.39	1.72
Proportion de réponses données par téléphone ⁵	61.5 %	51.1 %	14.7 %	28.6 %

TABLE 1 : Éléments descriptifs sur les profils pour l'attrition (données non pondérées)

interrogées dans le cadre de cette enquête sur les médecins généralistes, et se projettent en haut à gauche dans le premier plan factoriel (Figure 5). Ces médecins, ayant accepté d'être intégrés dans le dispositif d'enquêtes en 2018, ont répondu à peu de questionnaires entre l'hiver 2019-2020 et l'été 2021 (0.5 questionnaire en moyenne, pour 6 sollicitations sur cette période). L'étude de la composition par sexe et âge de ce *cluster* permet de retrouver un résultat classique sur les comportements de réponse : il s'agit plus souvent d'hommes et ils sont plus âgés que la moyenne des médecins interrogés.

À l'opposé, on retrouve les **bons élèves**, qui répondent à la plupart des vagues d'enquête pour lesquelles on les sollicite. Ils représentent près de 4 médecins interrogés sur dix, sont plus souvent des femmes et plus jeunes que la moyenne. Ces médecins ont répondu en moyenne à 5.2 vagues d'interrogation parmi les sollicitations qu'ils ont reçues sur la période étudiée.

Enfin, 36 % des médecins appartiennent au profil des **difficiles à convaincre**. Ces derniers répondent deux fois moins souvent que la moyenne pour les vagues non indemnisées, mais un peu souvent que la moyenne pour les deux vagues où ils bénéficient d'une compensation financière suite à leur réponse. Ce profil de médecins semble donc en moyenne plus sensible à l'incitation financière. Pour caractériser les médecins de ce profil, on a estimé un modèle « toutes choses égales par ailleurs » parmi les médecins des profils difficiles à convaincre ou bons élèves (estimation d'un modèle logit prenant en compte comme covariables des informations disponibles dans la base de sondage ou recueillies lors de la première vague d'interrogation à laquelle tous les médecins étudiés dans cette communication ont répondu), afin de chercher à caractériser les difficiles à convaincre. Il semble alors que les hommes, les médecins plus âgés, exerçant seul ou en Provence-Alpes-Côte d'Azur font en moyenne plus souvent partie du *cluster* des difficiles à convaincre que de celui des bons élèves.

Le groupe des difficiles à convaincre semble également avoir des comportements de réponse différents par rapport aux autres profils de répondants. « Plus difficiles à aller chercher », ils sont en particulier plus nombreux à répondre par téléphone. Ainsi, lorsqu'ils sont répondants à une enquête (pour les vagues d'interrogation où il y a eu des relances téléphoniques et où l'information sur le mode de réponse est disponible), ils répondent par téléphone dans plus de la moitié des cas, alors que moins de 30 % des réponses sont obtenues par téléphone pour ces vagues d'enquête. Ainsi, il semblerait que le groupe des **difficiles à convaincre** est à la fois plus sensible à l'incitation financière, mais aussi aux relances téléphoniques.

Parmi les personnes difficiles à convaincre, est-il plus efficace de leur proposer une incitation financière ou des relances par téléphone pour améliorer leur propension à répondre ? La première analyse descriptive menée dans cette partie montre qu'il y a un profil de personnes interrogées, en moyenne plus âgées, pour lesquelles il est plus difficile de recueillir des réponses. Il n'est toutefois pas aisé en première intention de distinguer parmi les potentielles méthodes utilisées (ajouts de relances, proposition de différents modes de collecte pour permettre notamment de répondre

par téléphone à ceux qui seraient moins prompts à répondre par Internet, incitation financière) celles qui sont les plus efficaces pour obtenir plus de réponses. Dans la deuxième partie, on discute une approche visant à quantifier l'importance des différentes variables caractérisant les répondants et la collecte pour prédire les taux de réponse, dans l'optique notamment d'étudier l'effet de l'indemnisation financière.

2 Comment quantifier l'effet de l'incitation financière par rapport à d'autres caractéristiques de la collecte et des médecins interrogés ?

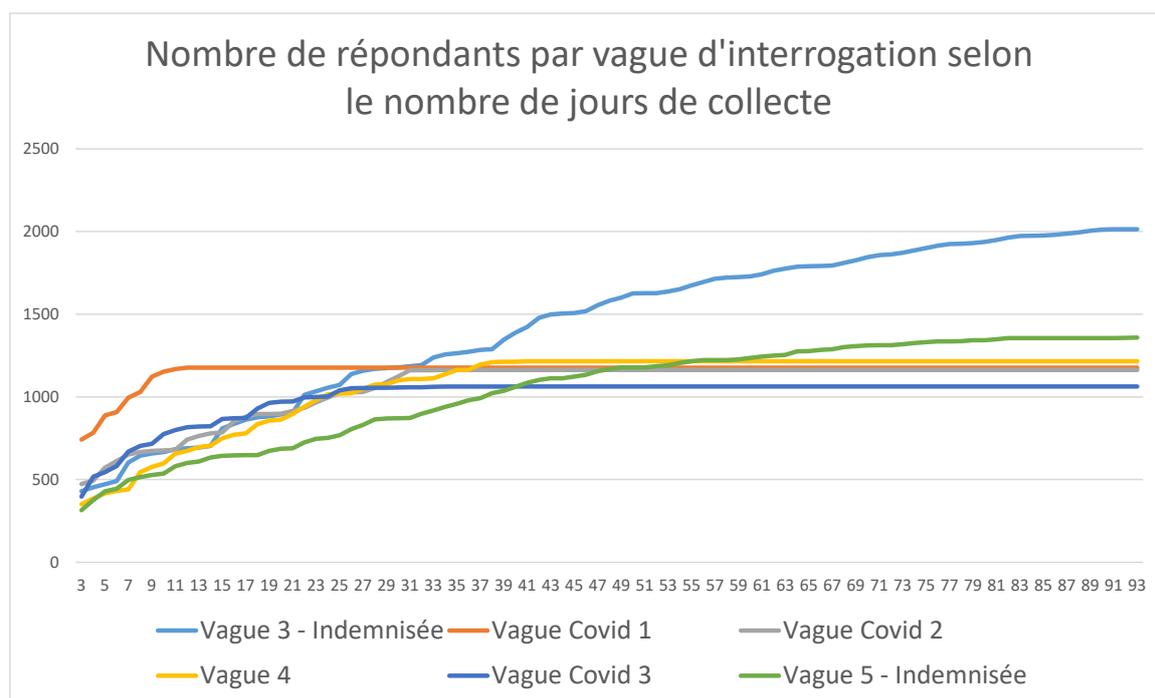


FIGURE 8 : Nombre de répondants selon le nombre de jour de collecte et selon la vague d'interrogation

On peut observer dans la Figure 8 le nombre de répondants pour chaque vague d'interrogation⁶, selon la durée de la collecte. On constate ici effectivement que, même si ce sont les vagues indemnisées qui conduisent au plus grand nombre de répondants et aux plus hauts taux de réponse à la fin de la collecte, il s'agit également de vagues d'interrogation où la durée de la collecte a été plus longue et où il y a eu par ailleurs des relances téléphoniques pour les personnes n'ayant pas répondu par Internet pendant les premières semaines de la collecte. Pour la vague non indemnisée Covid 1, dont la collecte a eu lieu en avril 2020 durant le premier confinement de la population en France et où l'activité des médecins généralistes était inférieure à leur activité habituelle (MONZIOLS et collab., 2020), le rythme des remontées de questionnaires lors de la fin

⁶ Comme précisé dans la première partie, on ne considère ici que les médecins contactés pour la dernière interrogation étudiée, hormis ceux contactés uniquement par téléphone - échantillon *full CATI*. Cela permet de raisonner à champ constant et de ne pas prendre en compte les éventuels biais liés au mode de collecte de l'enquête.

de la période de collecte est d'ailleurs encore très important, et supérieur à ce qu'on observe pour les vagues indemnisées.

Dans cette partie, on propose une méthode pour classer les variables qui caractérisent les individus interrogés ainsi que l'effort de collecte (durée de la collecte, présence ou absence de relances téléphoniques, recours à une indemnisation financière pour les répondants...) selon leur pouvoir prédictif du comportement de réponse individuel. On recourt pour cela à des techniques d'apprentissage supervisé (*machine learning*), permettant de classer les covariables des modèles selon leur pouvoir prédictif (*feature importance*).

2.1 Définition de la méthode

Pour classer les variables selon l'importance de leur pouvoir prédictif, on s'intéresse aux comportements de réponse des médecins interrogés entre fin 2019 et le printemps 2021, en réunissant l'ensemble des personnes interrogées durant six vagues d'interrogation. La variable à expliquer correspond au fait d'avoir ou non répondu à l'enquête pour une vague d'interrogation. Les covariables utilisées pour la prédiction de la réponse à une vague d'enquête sont les suivantes :

- Des variables qui caractérisent les médecins interrogés et qui sont fixes pour chaque individu :
 - Sexe
 - Âge au 1^{er} janvier 2018 (Moins de 50 ans, 50-59 ans, 60 ans ou plus)
 - Volume d'activité du médecin défini comme le nombre de consultations et visites réalisées en 2017 en tranches (inférieur au premier quartile, entre le premier et le troisième quartile et au-dessus du troisième quartile)
 - Le fait que le médecin exerce seul ou en groupe (variable collectée lors de la première vague d'interrogation auprès de l'ensemble des médecins interrogés)
- Des informations sur le lieu d'exercice des médecins interrogés :
 - Région d'exercice (Provence-Alpes-Côte d'Azur, Pays de la Loire ou autre région)
 - Densité de la zone d'exercice (Un médecin est considéré comme exerçant en zone sous-dense lorsque l'indicateur d'Accessibilité Potentielle Localisée (APL) du territoire de vie-santé dans lequel il exerce est inférieur à 2,8 consultations/an/habitant)
- Des informations qui caractérisent la vague de collecte et qui sont identiques pour l'ensemble des individus interrogés lors d'une vague :
 - Une variable continue qui est le numéro de la vague de collecte. Cette information est notamment utile pour caractériser le comportement « normal » d'attrition où on s'attend à une diminution du nombre de réponses au fil du temps.
 - La durée de la collecte en nombre de jours
 - La présence ou non de relances téléphoniques pour la collecte
 - Le cas échéant, la durée du terrain de relances téléphoniques
 - Le nombre de relances effectués par mail ou SMS lors de la vague d'interrogation
 - Le fait que le questionnaire soit principalement consacré aux thématiques liées à la Covid-19⁷
 - Le recours ou non à une indemnisation financière

⁷ La thématique était annoncée dans les mails-avis, et par ailleurs ces questionnaires sont plus courts.

- Une information qui dépend à la fois du lieu d'exercice des médecins interrogés et de la période de collecte, pour mesurer **l'activité des médecins pendant la période de collecte**. Il s'agit du nombre moyen par jour et par médecin généraliste de consultations, visites et téléconsultations réalisées dans le département d'exercice du médecin. Cette information est utilisée pour avoir une idée de la disponibilité des médecins pendant la période de collecte (il y a eu en effet d'importantes évolutions de l'activité des médecins selon les évolutions locales de la crise sanitaire et les mesures mises en place pour l'endiguer, comme les confinements de la population). Les données administratives de remboursement du Système National des Données de Santé permettent de calculer cette variable.

Parmi les variables incluses dans l'analyse, certaines sont mesurées au niveau du médecin (et fixes sur la période analysée pour les six vagues d'enquête), et d'autres mesurées au niveau de la vague de collecte (et fixes pour l'ensemble des médecins interrogés). Par ailleurs, certaines variables sont très corrélées entre elles, comme par exemple celles sur l'existence de relances téléphoniques durant la collecte et le recours à une indemnisation financière (par exemple, il n'y a pas de vague d'interrogation où les médecins sont indemnisés et sans relances téléphoniques). Il n'est ainsi pas aisé de recourir à des méthodes économétriques pour étudier le lien entre la variable d'intérêt (la réponse à l'enquête) et les différentes covariables, d'où le choix ici proposer d'utiliser une approche prédictive.

Quatre méthodes d'apprentissage supervisé sont testées dans cette communication. Le choix des méthodes utilisées repose notamment sur le fait qu'il existe une mesure permettant de quantifier le pouvoir prédictif des covariables incluses dans le modèle (**explicabilité** des modèles). Intuitivement, la mesure de l'importance des covariables incluses dans les modèles est effectuée en comparant les pouvoirs prédictifs des modèles lorsque la covariable est ou non incluse dans la modélisation. Les modèles suivants sont estimés⁸ :

- Trois méthodes utilisant des arbres de décision pour prédire la variable d'intérêt :
 - La méthode des forêts aléatoires au sens de BREIMAN (2001). Pour prédire une variable d'intérêt, les forêts aléatoires consistent en l'agrégation d'un ensemble d'arbres de décision pour la classification, avec une agrégation *bootstrap* (*bagging* ou *bootstrap aggregating*) permettant d'utiliser différentes observations et différentes covariables pour les arbres de décision qui composent la forêt.
 - Deux approches qui utilisent le *boosting* plutôt que le *bagging* pour combiner les prédictions des arbres de décision. Dans le *boosting*, on prend en compte les prédictions réalisées lors des étapes précédentes afin de pondérer les probabilités d'appartenance au sous-échantillon d'observations utilisé pour construire l'arbre de décision suivant. On considère deux algorithmes de *boosting* dans cette communication : la machine à *boosting* de gradient ou *Gradient Boosting Machine* et l'implémentation plus récente XGBoost ou *eXtreme Gradient Boosting* (CHEN et GUESTRIN, 2016).
- Une technique fondée sur la classification « selon le plus proche centroïde » ou *Nearest Shrunken Centroids* (TIBSHIRANI et collab., 2002). Pour prédire la variable d'intérêt, on effectue ici un partitionnement des données à partir des covariables, et on s'intéresse aux centroïdes (barycentres) selon la valeur de la variable à prédire, en l'occurrence le comportement de réponse. Pour une nouvelle observation dont on veut prédire le comportement de réponse selon les covariables, la valeur prédite correspond à la valeur du centroïde le plus « proche » de l'observation.

Dans la suite de la communication, les 4 méthodes de prédiction ici utilisées sont respectivement appelées Forêt, GBM, XGB et NSC. Pour améliorer le pouvoir prédictif des 4 méthodes d'apprentissage supervisé présentées ici, on procède à un affinage des hyperparamètres des modèles (par

⁸ Les estimations sont réalisées avec le *package* R *caret*.

exemple, paramètres sur les arbres de décision constitutifs des forêts aléatoires ou des méthodes de *boosting*). Pour choisir les hyperparamètres (*hyperparameter tuning*), on teste des grilles de paramètres de manière à augmenter la performance des modèles, mesurée grâce à l'exactitude des prévisions réalisées (proportion de prévisions justes). Pour mesurer empiriquement l'exactitude de la prévision, on procède par validation croisée avec 5 blocs : l'échantillon des données est séparé en 5 parts égales, puis 4/5 de l'échantillon est utilisé pour l'apprentissage du modèle, et les autres observations sont utilisées pour tester les performances de ce dernier. On répète cette opération 5 fois de manière à faire tourner l'échantillon de test et à avoir une estimation plus robuste de la performance du modèle.

Pour les trois méthodes ensemblistes reposant sur des arbres de décision (Forêt, GBM et XGB), une mesure d'influence des covariables est estimée pour chaque arbre de décision construit dans le cadre de la prédiction, et on moyenne ensuite ces mesures pour avoir une mesure relative d'influence globale par covariable. Pour chaque arbre de décision, l'importance relative d'une covariable x_j utilisée pour prédire la variable d'intérêt est mesurée ainsi (FRIEDMAN, 2001) :

$$\text{Importance}_{x_j}^{\text{Arbre}} = \sum_{\text{Splits selon } x_j} \text{Exactitude}_{\text{Avec split}} - \text{Exactitude}_{\text{Sans split}}$$

Ainsi, on mesure pour chaque arbre de décision le gain de performance obtenu en termes d'exactitude lorsque la covariable x_j est utilisée dans l'arbre pour séparer deux nœuds-fils.

Pour la méthode NSC, l'importance des covariables utilisées dans le modèle est mesurée par la distance entre le barycentre global des individus et les barycentres selon la modalité de la covariable à laquelle on s'intéresse.

Enfin, pour les quatre méthodes de prédiction testées dans cette communication, les covariables utilisées dans les modèles sont classées selon leur niveau d'importance mesuré avec les méthodes décrites ci-dessus. On calcule enfin un « rang moyen » qu'on assimile à l'importance du pouvoir prédictif des covariables. Par ailleurs, l'étude des rangs d'importance de chaque covariable selon la méthode d'apprentissage supervisé utilisée permet de discuter la robustesse de la méthode proposée ici pour classer les covariables permettant d'expliquer la propension à répondre des médecins généralistes.

2.2 Quelques résultats

Les méthodes utilisées comprennent une part d'aléatoire, par exemple pour constituer les échantillons d'apprentissage et de test lors de la phase de choix des hyperparamètres. L'utilisation d'une graine permet d'avoir des résultats reproductibles. En première analyse, il semble que les résultats esquissés dans cette partie, présentés pour une itération de la méthode, restent relativement stables en itérant à nouveau les estimations avec un aléa différent⁹.

Méthode	Exactitude		
	Minimale	Moyenne	Maximale
Forêt	60.7 %	61.9 %	62.4 %
GBM	61.7 %	62.7 %	63.8 %
XGB	61.6 %	62.7 %	63.6 %
NSC	57.8 %	58.9 %	60.5 %

TABLE 2 : Exactitude des modèles de prédiction selon le jeu de données utilisé pour le test du modèle

⁹ Ces contrôles de robustesse nécessitent toutefois d'être précisés et effectués plus systématiquement pour évaluer la robustesse de la méthode proposée dans cette communication.

Après avoir optimisé les hyperparamètres des modèles, les échantillons de test des données initiales (5 blocs de données) sont utilisés à nouveau pour mesurer la précision du modèle sur ces cinq échantillons, en utilisant le critère d'exactitude, c'est-à-dire le nombre de prédictions qui s'avère juste. Les résultats sur l'exactitude sont donnés dans la Table 2.

On note que l'exactitude des modèles utilisés pour prédire les comportements de réponse pour chaque vague d'enquête est de l'ordre de 60 % pour les quatre approches testées dans cette communication. Le taux de réponse observé ici parmi les médecins interrogés est de 49.6 %, il est donc rassurant de constater que l'exactitude observée ici est supérieure à ce qu'on aurait observé avec un modèle extrêmement naïf qui renverrait toujours la catégorie la plus probable (ici, de ne pas répondre), qui aurait une exactitude de 50.4 %.

Il semble que les modèles utilisant des arbres de décision ont des performances légèrement meilleures que le modèle NSC fondé sur le critère du plus proche centroïde (exactitude moyenne de 59 % pour les cinq échantillons de test, contre 62-63 % pour les trois autres modèles). Les modèles ensemblistes GBM et XGB qui reposent sur le *boosting* semblent conduire à des niveaux d'exactitude très légèrement meilleurs que le modèle de forêts aléatoires testé ici. Toutefois, ces résultats doivent être interprétés avec précaution, notamment car il y a eu peu de tests de robustesse pour la validation des modèles. Par ailleurs, les performances des modèles restent globalement assez mauvaises, car les prédictions réalisées sont erronées généralement dans 4 cas sur 10¹⁰.

Les résultats sur le rang d'importance des variables pour les 4 modèles prédictifs de la propension à répondre à une vague d'interrogation du panel d'observation des pratiques et des conditions d'exercice en médecine générale sont présentés dans la Table 3. À la lecture de cette table, on peut tirer quelques enseignements :

- Tout d'abord, on peut noter une relative stabilité dans les classements obtenus pour les rangs d'importance de chaque covariable, et ce pour les 4 modèles d'apprentissage supervisé testés. Même si théoriquement, il n'y a pas de propriété théorique pour la convergence des estimateurs d'importance des covariables du modèle, il est plutôt rassurant de constater que les résultats semblent aller dans le même sens pour ces résultats préliminaires.
- Sur les quatre variables pour lesquelles on observe le plus fort pouvoir prédictif (rang moyen le plus faible), trois caractérisent la vague d'interrogation. La durée de collecte (totale et durée du terrain téléphonique) semblent avoir un rôle important pour prédire les comportements de réponse. La variable du « numéro de vague », qui est une variable utilisée comme continue variant entre 1 et 6 et utilisée pour modéliser l'attrition « classique » du dispositif d'enquêtes, semble également centrale pour prédire les comportements de réponse des médecins généralistes. Le mode d'exercice, seul ou en cabinet de groupe¹¹, représente également une information importante pour prédire les comportements de réponse des médecins.
- En revanche, les autres variables qui caractérisent les médecins interrogés semblent moins importantes pour prédire les comportements de réponse, par rapport aux variables caractérisant l'effort de collecte et l'activité effectivement observée dans le département d'exercice des médecins durant une vague d'enquête. Par exemple, le volume d'activité « usuel » mesuré en 2017 est une caractéristique qui apparaît peu prédictive des comportements pour l'attrition dans le cas d'usage présenté ici. Cela pourrait être lié au fait qu'on introduit

¹⁰ La performance des modèles n'est toutefois pas une fin en soi pour cette analyse, dont l'objectif est d'étudier le pouvoir explicatif des variables utilisées pour la prédiction des comportements de réponse.

¹¹ Cette variable n'est pas disponible dans la base de sondage, mais elle a été collectée lors de la vague d'inclusion de l'enquête, et l'information est donc disponible pour l'ensemble des médecins interrogés en 2020 et 2021.

Variable	Type de caractéristique	Type de variable	Rang d'importance selon la méthode				Rang moyen
			Forêt	GBM	XGB	NSC	
Sexe	Individuelle	Catégorielle	7	7	5	8	6,75
Âge (50-59 ans)	Individuelle	Catégorielle	10	6	6	12	8,5
Âge (60 ans et +)	Individuelle	Catégorielle	11	9	7	10	9,25
Activité 2017 (Modérée)	Individuelle	Catégorielle	15	13	13	14	13,75
Activité 2017 (Élevée)	Individuelle	Catégorielle	12	10	9	11	10,5
Exercice solo	Individuelle	Catégorielle	5	3	2	6	4
Région (Paca)	Individuelle	Catégorielle	8	5	4	9	6,5
Région (PdL)	Individuelle	Catégorielle	17	16	16,5	16	16,375
Zone exercice Sous-dense	Individuelle	Catégorielle	16	15	14	17	15,5
Numéro Vague	Vague	Continue	2	2	10	3	4,25
Durée Collecte	Vague	Continue	1	4	1	2	2
Relances tél	Vague	Catégorielle	14	17	15	7	13,25
Durée Collecte tél	Vague	Continue	3	1	3	1	2
Nb relances	Vague	Continue	13	12	12	15	13
Thématique Covid	Vague	Catégorielle	9	14	16,5	5	11,125
Indemnisation	Vague	Catégorielle	6	8	8	4	6,5
Activité dép Pendant vague	Mixte	Continue	4	11	11	13	9,75

TABLE 3 : Importance des covariables utilisées dans les modèles prédictifs pour 4 modèles d'apprentissage supervisé

dans les modèles une variable plus objective sur l'activité des médecins dans leur département d'exercice pendant la période de collecte, qui pourrait mieux refléter la disponibilité des médecins qui a pu fortement varier selon leur lieu d'exercice et les évolutions de la crise de la Covid-19.

- Pour ce qui est de la variable caractérisant le recours à une indemnisation financière, il est difficile de donner une conclusion tranchée. En effet, le rang d'importance moyen qui mesure le pouvoir prédictif de cette variable par rapport aux autres introduites dans le modèle, est moyen (proche du rang d'importance de la variable qui distingue les médecins exerçant en Provence-Alpes-Côte d'Azur des autres, et de la variable distinguant les médecins hommes de leurs consœurs).

3 Discussion

Cette communication vise à étudier les mécanismes de l'attrition dans une enquête auprès de médecins généralistes en France, notamment avec l'objectif de mesurer l'effet de recourir à une incitation financière pour susciter la réponse des personnes interrogées. Dans un premier temps, où l'unité d'analyse est une personne interrogée, on dresse une typologie des profils-type de l'attrition. En particulier, la classification réalisée met en évidence un groupe de « difficiles à convaincre », qui seraient potentiellement plus prompts à réagir à une incitation financière. Dans la seconde partie du travail, on recourt à une analyse où l'unité d'observation est la personne interrogée **durant une des six vagues d'enquête analysées**, permettant ainsi d'étudier conjointement le lien entre la réponse pour une vague donnée avec les caractéristiques individuelles mais aussi de la vague d'enquête. On recourt pour cela à des approches prédictives, et on utilise l'explicabilité des modèles d'apprentissage supervisé qu'on considère, afin de classer les covariables des modèles selon leur degré d'importance. Il ressort comme attendu qu'au-delà des caractéristiques individuelles, l'effort mis dans la collecte, et notamment la durée de la collecte par Internet et par téléphone, détermine fortement les taux de réponse ensuite obtenus. Pour ce qui est de l'indemnisation financière, y recourir semble également être une variable centrale pour expliquer les comportements de réponse (on observe effectivement des taux de réponse plus importants pour les vagues d'enquête indemnisées), dont le pouvoir prédictif serait plus important que les variables individuelles caractérisant les médecins dans l'exemple discuté dans cette communication (mis à part le type d'exercice, seul ou en cabinet de groupe).

La méthode proposée ici pour classer les variables utilisées dans un modèle prédictif selon leur importance a le double avantage d'être simple à utiliser, et de ne pas reposer sur des hypothèses fortes (notamment sur les corrélations entre covariables introduites dans les modèles) pour pouvoir réaliser l'estimation. On ne dispose pas en revanche de propriétés théoriques sur la convergence des estimateurs d'importance qu'on calcule, et par ailleurs dans l'exemple présenté ici, la performance des modèles prédictifs des comportements de réponse est assez mauvaise. En revanche, il est possible de faire varier les méthodes de prédiction utilisées, les hyperparamètres des modèles et les façons de calculer l'importance des covariables, afin d'évaluer empiriquement la pertinence de recourir à des méthodes prédictives de type *machine learning* pour classer les variables expliquant une variable d'intérêt selon leur degré d'importance. Ainsi, le recours à cette méthode utilisant des modèles de *machine learning* pour classer les paramètres les plus explicatifs d'une variable d'intérêt pourrait s'avérer utile pour d'autres cas d'usage de la statistique publique.

Cette méthode possède de nombreuses limites, par exemple dues au fait que la mesure d'importance qu'on calcule n'indique pas dans quel sens jouent les covariables utilisées pour prédire les comportements de réponse, contrairement aux coefficients qui peuvent être estimés à partir d'un modèle économétrique. Par ailleurs, il est difficile de conclure sur l'intérêt de l'incitation financière pour limiter l'attrition dans le cadre de l'enquête auprès des médecins généralistes utilisée dans cette communication. Le fait de ne pas avoir effectué d'expérience randomisée, et que les interrogations soient menées pendant une crise sanitaire exceptionnelle avec de fortes variations épidémiques sur la période, rendent difficile de tirer des conclusions affirmées sur l'intérêt de recourir ou non à des incitations financières dans les dispositifs d'enquête ultérieurs. Sur la base des modèles prédictifs ici estimés, des travaux complémentaires seront réalisés visant à quantifier le gain en termes de taux de réponse prédit lorsqu'on recourt à une indemnisation financière, ce qui pourra être mis en regard du coût de cette dernière.

Enfin, cette communication se limite à l'étude de l'attrition, et non au mécanisme d'inclusion dans le dispositif d'enquêtes (le taux d'inclusion des médecins est de 27 %, c'est-à-dire que près de trois médecins sollicités sur quatre n'ont pas souhaité participer au dispositif ou sont hors du champ de l'enquête), pour lequel le recours à une indemnisation pourrait être plus efficace que pour la limitation de l'attrition pour des médecins déjà engagés dans le processus de collecte des

données. Il est également important de garder en tête les effets collatéraux du recours ou non à une incitation financière (SINGER et YE, 2013) : s’il devient usuel pour la statistique publique en France de recourir à des incitations financières pour les répondants à une enquête, cela a aussi un effet pour les autres enquêtes, que ce soit pour la statistique publique, car les personnes interrogées risquent alors d’être habituées à être indemnisées lorsqu’on les sollicite pour répondre, et de refuser plus souvent de répondre en absence d’indemnisation.

Références

- BALLINI, C., , A. LOUVET et F. MURAT. 2022, « Motiver les enquêtés à l’aide d’incitations financières : une expérimentation conduite en France dans le cadre de l’enquête internationale PIAAC », *Journées de Méthodologie Statistique de l’Insee*.
- BARNES, W., G. BRIGHT et C. HEWAT. 2008, « Making sense of Labour Force Survey response rates », *Economic and Labour Market Review*, vol. 2, p. 32–42.
- BERGEAT, M., H. CHAPUT, P. VERGER, D. SCRONIAS et collab.. 2021, « Comment les médecins généralistes prennent-ils en charge les patients atteints de la Covid-19? », *Drees, Études et Résultats*, vol. 1177.
- BREIMAN, L. 2001, « Random forests », *Machine learning*, vol. 45, n° 1, p. 5–32.
- CHEN, T. et C. GUESTRIN. 2016, « XGBoost: A scalable tree boosting system », dans *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, p. 785–794.
- CHO, Y. I., T. P. JOHNSON et J. B. VANGHEEST. 2013, « Enhancing surveys of health care professionals: a meta-analysis of techniques to improve response », *Evaluation and the health professions*, vol. 36, n° 3, p. 382–407.
- EUROSTAT. 2020, « Household Budget Survey, 2015 Wave, EU Quality Report », *European Commission*.
- FRIEDMAN, J. H. 2001, « Greedy function approximation: a gradient boosting machine », *Annals of Statistics*, p. 1189–1232.
- MONZIOLS, M., H. CHAPUT, P. VERGER, D. SCRONIAS, B. VENTELOU et collab.. 2020, « Comment les médecins généralistes ont-ils exercé leur activité pendant le confinement lié au Covid-19? », *Drees, Études et Résultats*, vol. 1150.
- SINGER, E. et C. YE. 2013, « The use and effects of incentives in surveys », *The Annals of the American Academy of Political and Social Science*, vol. 645, n° 1, p. 112–141.
- TIBSHIRANI, R., T. HASTIE, B. NARASIMHAN et G. CHU. 2002, « Diagnosis of multiple cancer types by shrunken centroids of gene expression », *Proceedings of the National Academy of Sciences*, vol. 99, n° 10, p. 6567–6572.
- VANGHEEST, J. B., T. P. JOHNSON et V. L. WELCH. 2007, « Methodologies for improving response rates in surveys of physicians: a systematic review », *Evaluation and the health professions*, vol. 30, n° 4, p. 303–321.