



Journées de méthodologie statistique de l'Insee

2022



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.

Journées de méthodologie statistique de l'Insee – 31 mars 2022

Plan

Contexte et données

Qui sont les répondants à
chaque vague d'interrogation ?

Comment quantifier l'effet de
l'incitation financière par
rapport à d'autres
caractéristiques de la collecte ?

PANEL D'OBSERVATION

des pratiques et des conditions d'exercice

en médecine générale

DREES
ORS -
UMR SESSTIM
URML



Plan

Contexte et données

Qui sont les répondants à
chaque vague d'interrogation ?

Comment quantifier l'effet de
l'incitation financière par
rapport à d'autres
caractéristiques de la collecte ?

PANEL D'OBSERVATION

des pratiques et des conditions d'exercice

en médecine générale

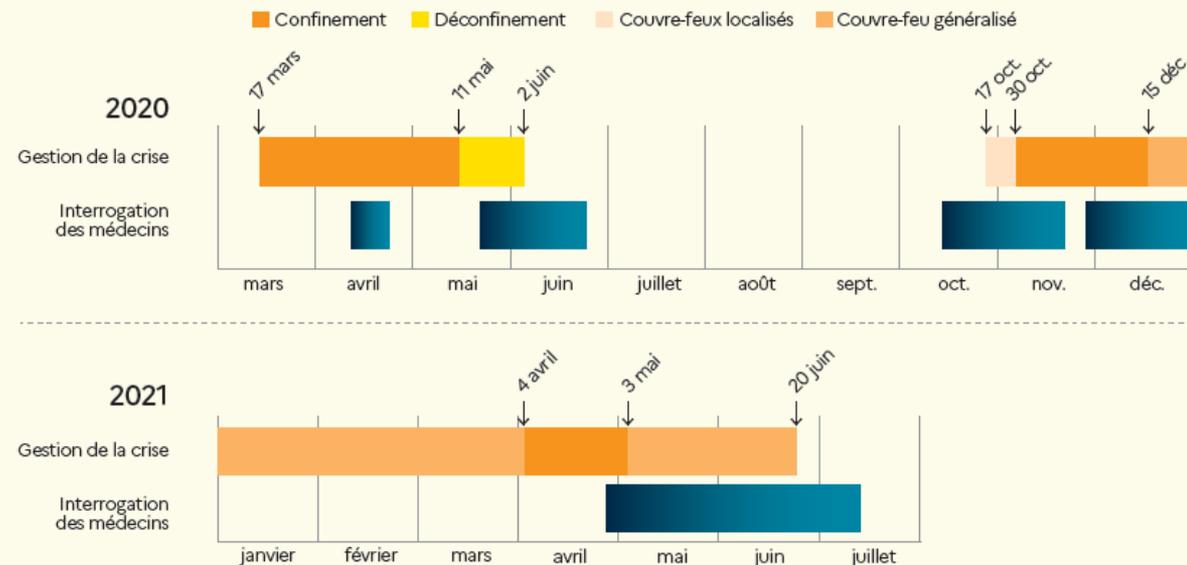
DREES
ORS -
UMR SESSTIM
URML



Le « panel médecins généralistes »

Une enquête par questionnaire auprès de médecins généralistes libéraux

Calendrier des vagues d'enquêtes sur la crise sanitaire menées auprès des médecins généralistes en 2020 et 2021



Enquête par *vagues*

Modes de collecte : *Web*
puis téléphone OU
téléphone seulement

2686 médecins sollicités
hors échantillon full CATI

Thématiques : crise
sanitaire (activité, opinions
gestion de la crise...),
prévention, vaccination,
antibiorésistance

Les répondants sont incités financièrement

Sauf quand il ne reste plus de sous



*4 vagues
d'interrogation
supplémentaires
par rapport aux
prévisions initiales,
sans indemnisation
des médecins*

 *Expérience
non randomisée,
tous les répondants
d'une vague sont
indemnisés*

Objectifs et littérature

- Grâce à cette **expérience naturelle**, estimer l'effet de l'indemnisation financière sur la propension à répondre
 - La prochaine fois, est-ce qu'on continue avec cette indemnisation pour les répondants, rare dans la statistique publique française ?
 - Plus généralement, quels sont les déterminants de l'attrition pour cette enquête ?
- Littérature sur le sujet abondante (en population générale **et pour les professionnels de santé**)
 - 👉 Incitations (😊 Taux de réponse 😊 Structure de l'échantillon)
 - 👉 Incitations prépayées plutôt que conditionnelles à la réponse
 - 👉 Incitations monétaires plutôt que non-monétaires
- Difficultés : expérience non randomisée / on ne regarde que l'attrition



Plan

Contexte et données

Qui sont les répondants à
chaque vague d'interrogation ?

Comment quantifier l'effet de
l'incitation financière par
rapport à d'autres
caractéristiques de la collecte ?

PANEL D'OBSERVATION

des pratiques et des conditions d'exercice

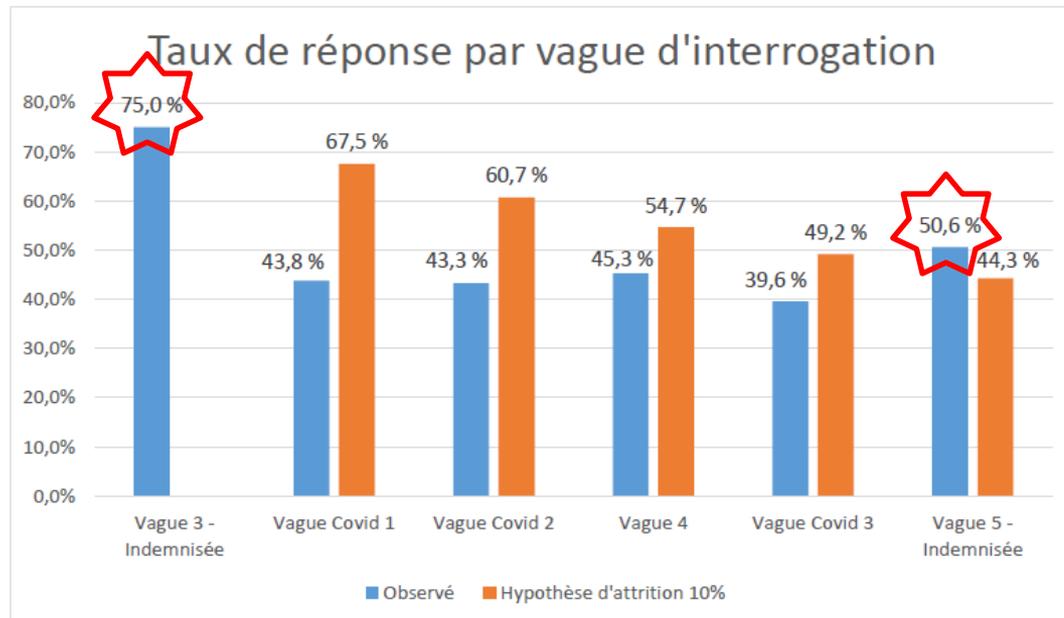
en médecine générale

DREES
ORS -
UMR SESSTIM
URML



Taux de réponse par vague d'interrogation

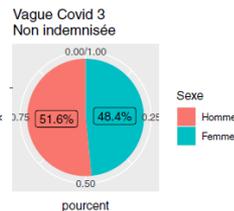
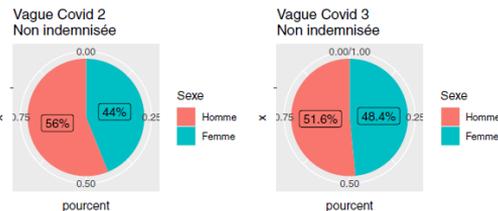
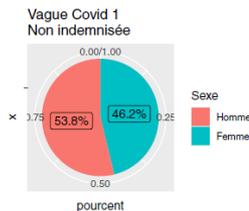
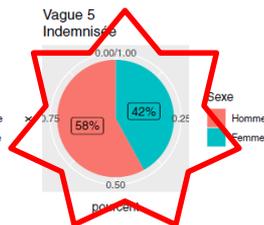
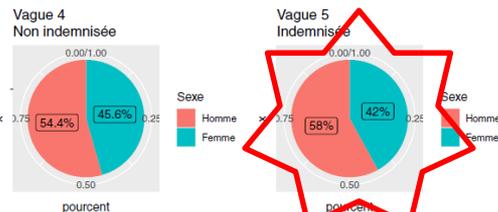
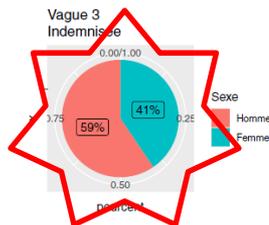
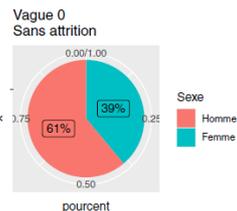
Plus élevé pour les vagues indemnisées



- Contextes de collecte différents (relances téléphoniques, collecte plus longue)
- Contexte sanitaire évolutif sur la période (→ Impacte la disponibilité des médecins)

Qui sont les répondants à chaque vague d'interrogation ?

Lors des vagues indemnisées, structure moins déformée par rapport à la population cible

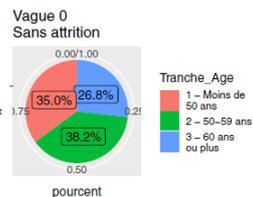


Poids utilisé : poids calé
« avant attrition » de la première
vague d'interrogation

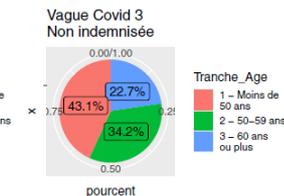
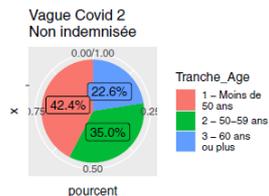
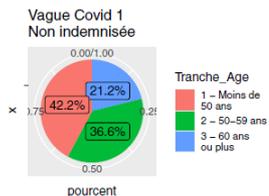
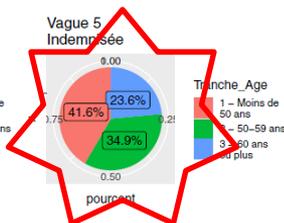
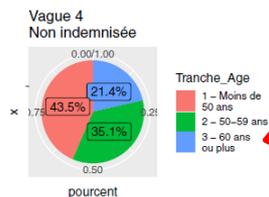
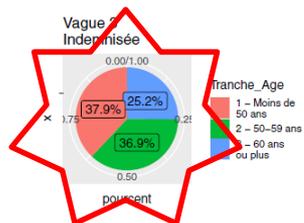


Qui sont les répondants à chaque vague d'interrogation ?

Lors des vagues indemnisées, structure moins déformée par rapport à la population cible

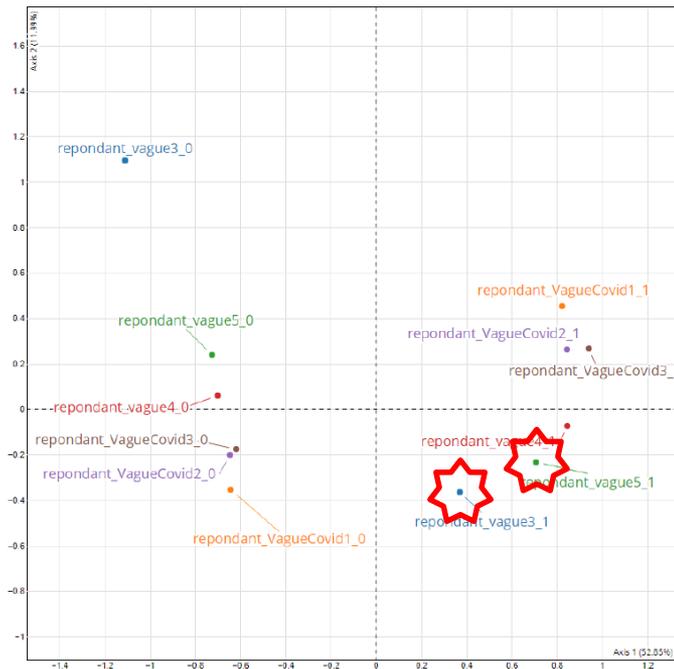


Poids utilisé : poids calé
« avant attrition » de la première
vague d'interrogation



Les profils de l'attrition

Y a-t-il des médecins qui ne répondent que si on leur donne de l'argent ?



Méthode :

1. Analyse en Composantes Principales
 - a. Unité statistique : médecin interrogé
 - b. Variables actives : a répondu à la vague n

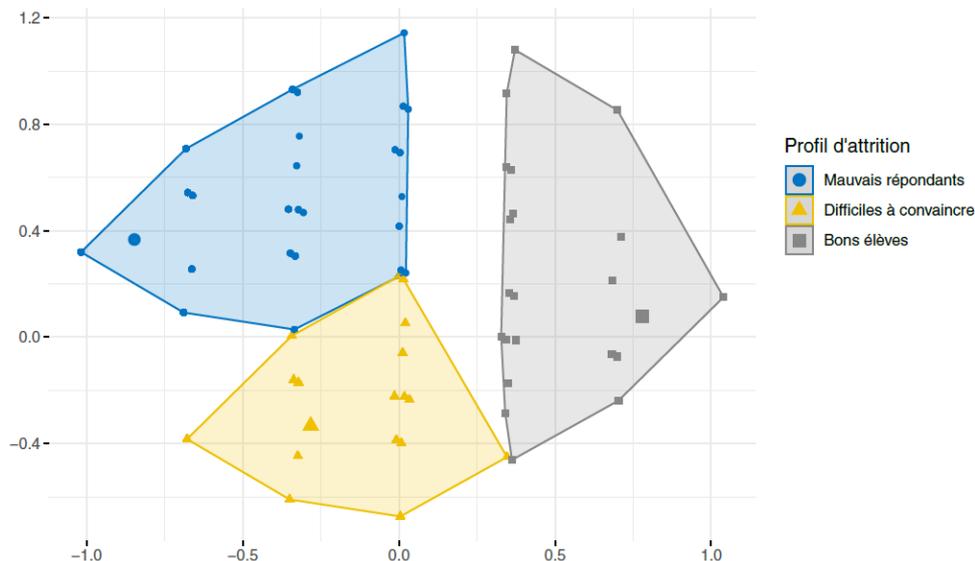
Axe horizontal : réponse vs. non-réponse

*Axe vertical : enquête
« classiques » vs. enquêtes
« Flash »*

Les profils de l'attrition

Y a-t-il des médecins qui ne répondent que si on leur donne de l'argent ?

Classification des médecins interrogés selon leurs comportements vis-à-vis de l'attrition



Méthode :

1. Analyse en Composantes Principales
2. Classification ascendante hiérarchique

Trois profils d'attrition dont des personnes interrogées « difficiles à aller chercher »

Les profils de l'attrition

Y a-t-il des médecins qui ne répondent que si on leur donne de l'argent ?

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

Méthode :

1. Analyse en Composantes Principales
2. Classification ascendante hiérarchique
3. Caractérisation des profils de l'attrition

Les difficiles à convaincre répondent peu par Internet et peu aux vagues non indemnisées (où il y a eu peu de relances téléphoniques...)

Part réponses données par téléphone : calcul seulement sur vagues avec des relances téléphoniques

Plan

Contexte et données

Qui sont les répondants à
chaque vague d'interrogation ?

Comment quantifier l'effet de
l'incitation financière par
rapport à d'autres
caractéristiques de la collecte ?

PANEL D'OBSERVATION

des pratiques et des conditions d'exercice

en médecine générale

DREES
ORS -
UMR SESSTIM
URML



Méthodologie

1. Définition des variables utilisées pour prédire les comportements de réponse
2. Estimation de différents modèles prédictifs (*Machine Learning*) pour prévoir les comportements de réponse
 - a. Choix de quelques modèles
 - b. Optimisation des hyperparamètres
 - c. Comparaison des performances en termes de pouvoir prédictif
3. Calcul de métriques d'importance pour les covariables du modèle
 - a. Voir l'effet de l'indemnisation financière par rapport à d'autres caractéristiques
 - b. Stabilité entre les modèles ?



Variables utilisées dans les modèles prédictifs

Variable	Type de caractéristique	Type de variable	Variable	Type de caractéristique	Type de variable
Sexe	Individuelle	Catégorielle	Numéro Vague	Vague	Continue
Âge (50-59 ans)	Individuelle	Catégorielle	Durée Collecte	Vague	Continue
Âge (60 ans et +)	Individuelle	Catégorielle	Relances tél	Vague	Catégorielle
Activité 2017 (Modérée)	Individuelle	Catégorielle	Durée Collecte tél	Vague	Continue
Activité 2017 (Élevée)	Individuelle	Catégorielle	Nb relances	Vague	Continue
Exercice solo	Individuelle	Catégorielle	Thématique Covid	Vague	Catégorielle
Région (Paca)	Individuelle	Catégorielle	Indemnisation	Vague	Catégorielle
Région (PdL)	Individuelle	Catégorielle	Activité dép Pendant vague	Mixte	Continue
Zone exercice Sous-dense	Individuelle	Catégorielle			



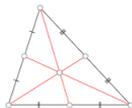
Unité d'observation :
individu x vague d'enquête

SNDS

1. Définition des variables utilisées pour prédire les comportements de réponse
2. Estimation de différents modèles prédictifs (*Machine Learning*) pour prévoir les comportements de réponse
3. Calcul de métriques d'importance pour les covariables du modèle

Performances des modèles utilisés : bof...

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 %



- Trois méthodes de prédiction avec des arbres de décision, une méthode de classification selon le plus proche « centroïde »
- Utilisation de différents échantillons de données pour paramétrer/tester le modèle
- On fait tourner les échantillons de test pour avoir des mesures de performance (un peu) robustes : *validation croisée répétée*

1. Définition des variables utilisées pour prédire les comportements de réponse
2. Estimation de différents modèles prédictifs (*Machine Learning*) pour prévoir les comportements de réponse
3. Calcul de métriques d'importance pour les covariables du modèle

... Mais on peut dire des choses sur l'importance des covariables utilisées pour la prédiction

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

1. Définition des variables utilisées pour prédire les comportements de réponse
2. Estimation de différents modèles prédictifs (*Machine Learning*) pour prévoir les comportements de réponse
3. Calcul de métriques d'importance pour les covariables du modèle

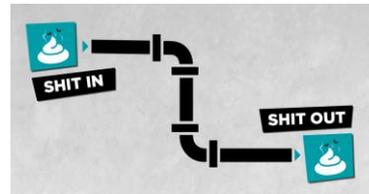
... Mais on peut dire des choses sur l'importance des covariables utilisées pour la prédiction

Variable	Type de caractéristique	Type de variable	Rang d'importance selon la méthode				Rang moyen
			Forêt	GBM	XGB	NSC	
Numéro Vague	Vague	Continue	2	2	10	3	4,25
Durée	Vague	Continue	1	4	1	2	2
Collecte	Vague	Continue	1	4	1	2	2
Relances tél	Vague	Catégorielle	14	17	15	7	13,25
Durée	Vague	Continue	3	1	3	1	2
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

1. Définition des variables utilisées pour prédire les comportements de réponse
2. Estimation de différents modèles prédictifs (*Machine Learning*) pour prévoir les comportements de réponse
3. Calcul de métriques d'importance pour les covariables du modèle

En quelques mots

- 🤖 Une méthode qui utilise une approche **prédictive** pour donner des résultats **descriptifs**
- 🤖 Simple à utiliser, et permet de ne pas avoir (trop) de contraintes dans le choix des covariables (*multicolinéarité*)
- 😐 Dans le cas d'usage présenté ici :
 - 🤖 Des résultats classiques en étudiant de façon descriptive les profils de l'attrition
 - 😐 Dur de conclure sur ce qui est le plus efficace dans l'effort de collecte entre le recours à une indemnisation et d'intensifier la collecte (**expérience non randomisée**)
- 😐 Principales limites
 - 😐 On ne regarde qu'une partie de la non-réponse (**attrition**)
 - 😐 Avec les modèles utilisés, on ne sait pas dans quel sens « jouent » les covariables



- Regarder **rapport coût/bénéfice** de l'indemnisation avec approche prédictive
- Suggestions bienvenues à la pause ou plus tard 😊

Merci de votre attention 😊

Et merci au comité d'organisation des JMS 😊

Le papier :

http://www.jms-insee.fr/2022/S20_2_ACTE_BERGEAT_JMS2022.pdf