



CORRECTION DE LA NON-RÉPONSE ET ESTIMATION DES PRÉVALENCES :

RÉSULTATS ISSUS DE TROIS COHORTES ÉPIDÉMIOLOGIQUES CIBLANT LES RISQUES PROFESSIONNELS

N. Soullier, B. Geoffroy-Perez, A. Gueguen, L. Benezet, J. Chatelot, M. Zins, G. Santin

Colloque Sondages, Lyon, 24 octobre 2018



CONTEXTE

INTRODUCTION (1/4)



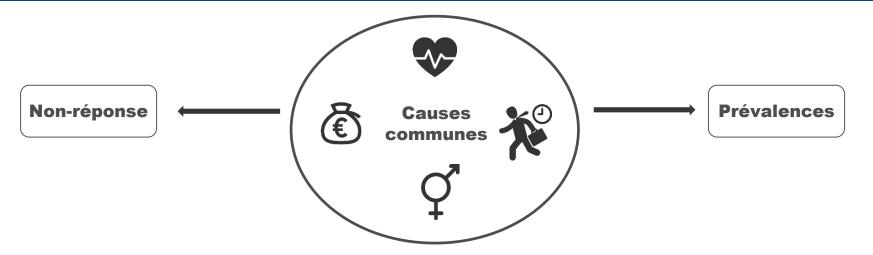
BAISSE DU TAUX DE RÉPONSE AUX ENQUÊTES

- ✓ Les enquêtes = un outil utile de collecte de données
- ✓ Baisse du taux de réponse aux enquêtes → aussi les enquêtes épidémiologiques
- ✓ Conséquences de la non-réponse totale :
 - Perte de précision
 - Présence potentielle de biais de non-réponse



INTRODUCTION (2/4)





ESTIMATION DE PRÉVALENCE

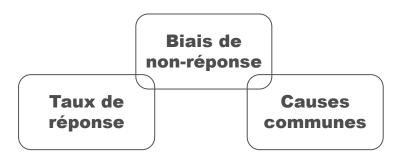
- ✓ <u>Cadre : enquêtes de **surveillance épidémiologique** avec pour objectif de produire des statistiques extrapolables à une population d'intérêt (p. ex. **prévalence**)</u>
- ✓ Biais de non-réponse = fonction du taux de réponse et de la covariance entre la variable d'intérêt et la probabilité de réponse
 - Il y a un biais potentiel s'il y a des **causes communes** associées à la non-réponse et aux variables d'intérêt.
 - De précédents travaux ont montré un lien entre la non-réponse totale aux enquêtes sur la santé et les caractéristiques démographiques, socio-économiques, l'état de santé et le statut professionnel.

INTRODUCTION (3/4)



LIMITER LES BIAIS DE NON-RÉPONSE

- ✓ Méthode a priori : définir des protocoles qui permettent d'augmenter le taux de réponse (relances, multimode...)
- ✓ Méthode a posteriori : utiliser des informations liées à la fois à la probabilité de réponse et aux variables d'intérêt pour réduire les biais de non-réponse
 - Nécessite des données collectées sur l'ensemble de l'échantillon
 - Approche à prévoir a priori (autorisations, consentement...)



INTRODUCTION (4/4)



OBJECTIF

- ✓ Développer une méthodologie pour corriger la non-réponse totale dans les estimations de prévalence
- ✓ Trois cohortes épidémiologiques ciblant les risques professionnels
- ✓ Echantillons appariés avec des données issues de systèmes d'information (SI) existants
 - SI santé : Sniiram
 - SI profession : assurances vieillesse



MATÉRIEL ET MÉTHODES

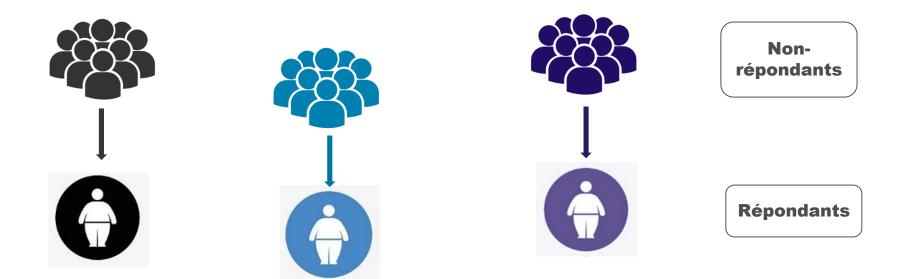
PRÉSENTATION DES COHORTES



	Constances	Coset-MSA	Coset-RSI	
Population couverte	Affiliés au RGSS, à la LMDE, CAMIEG, MGEN et MFPs	Travailleurs du monde agricole (salariés et non-salariés)	Commerçants, artisans, professions libérales (non-salariés)	
Mise en œuvre	Inserm-UVSQ-CNAM	Santé publique France-DST en partenariat avec le Régime concerné		
Modalités d'inclusion	Tirage au sort			
Mode de recueil de données	Actif : Questionnaire (auto et administré), bilan de santé dans un centre de santé (CES)	Actif: auto-questionnaire		
	Passif : Donnée	es issues des systèmes d'information existants		
Phase pilote (année)	2009	2010	2012	
Extension nationale (année)	2012-2019	2017	2017	
Effectif initial tiré au sort	Plus de 2 000 000	180 000	300 000	
Effectif attendu de répondants au questionnaire	200 000	30 000	30 000	
Effectif suivi passivement	200 000 + 400 000	30 000 + 60 000	30 000 + 60 000	
Données analysées (effectif invité)	Invités en 2012-2013 âgés de 30 à 69 ans (n=471 152)	Pilote (n=10 000)	Pilote (n=20 000)	
Répondants	31 642 (7%)	2 363 (24%)	2 661 (13%)	

CORRECTION DE LA NON-RÉPONSE TOTALE (1/2)





CORRECTION PAR REPONDÉRATION

Principe : augmenter le poids des répondants afin qu'ils représentent également les nonrépondants

- Etape 1 : modélisation de la non-réponse
- <u>Etape 2</u>: constitution de groupes homogènes de réponse à partir des probabilités prédites à l'étape 1

CORRECTION DE LA NON-RÉPONSE TOTALE (2/2)



SOUS L'HYPOTHÈSE MAR

« Missing At Random » (manquant au hasard) : il y a indépendance entre la probabilité de réponse et la variable d'intérêt conditionnellement aux causes communes.

- Après prise en compte de ces causes communes dans la correction de la non-réponse,
 la prévalence estimée sur l'échantillon des répondants est sans biais.
- Identifier et recueillir les causes communes (variables auxiliaires) à la fois pour les répondants et les non-répondants.
- Information des personnes, possibilité de s'opposer au recueil de ces données

DONNÉES AUXILIAIRES COMMUNES



<u>SniirAM</u>	<u>Données socio-</u> <u>démographiques</u>
Remboursements de soins :	
Recours à un professionnel de santé, par type ;	Sexe
Remboursements de médicaments, par type (classification « Anatomique, Thérapeutique et Chimique » (ATC));	• Age
Nombres d'actes de biologie, pharmacie, CCAM et LPP, par type ;	
Montant des prestations.	
Hospitalisations:	
Nombre d'hospitalisations, type d'établissement (public ou privé), motif (chirurgical ou non chirurgical), diagnostics (CIM10);	
Durée d'hospitalisation (Constances).	
Affections de longue durée (ALD) (sauf Coset-MSA)	
Absentéisme pour raison de santé - Indemnités journalières (sauf Constances) :	
Montant total des versements, durée totale ;	
Type d'arrêts (pour maladie, accident du travail ou maternité).	

DONNÉES AUXILIAIRES SPÉCIFIQUES



<u>Constances</u>	<u>Coset-MSA</u>	<u>Coset-RSI</u>	
Cnav	Bases contrats / cotisations et AT-MP* de la MSA	Base carrière du RSI	
Variables sociodémographiques	Variables sociodémographiques	Variables sociodémographiques	
Département (résidence)	Département (lieu de travail)	Caisse d'affiliation	
Régime d'affiliation	Statut (salarié / non salarié)	Groupe professionnel (artisan / commerçant / profession	
<u>Données socioprofessionnelles</u>	Données professionnelles	libérale)	
 Catégorie socioprofessionnelle Actif / Inactif 	 Caractéristiques de la dernière activité professionnelle en date en tant qu'affilié à 	<u>Données professionnelles</u>	
 Revenus d'emploi-retraite Revenus d'allocations 	la MSA (secteur d'activité, durée d'emploi) ; Nombre d'emplois salariés dans l'année.	Ancienneté au RSIPériode d'emploi terminée au	
		RSI au moment du tirage au sort	
	AT-MP*		
	 Nombre d'AT-MP* reconnus, par gravité, par année. 		

^{*}Accidents du Travail – Maladies Professionnelles

MODÉLISATION DE LA NON-RÉPONSE



RÉGRESSION LOGISTIQUE

Processus de sélection :

- 1. Exclusion des prévalences rares : minimum requis de répondants et de non-répondants (10 ou 30)
- 2. Modélisation de la non-réponse par type de variables : régressions logistiques multivariées avec les variables socio-démo / professionnelles / santé → sélection des variables significatives à 0,20
- 3. Construction du modèle final : régression logistique multivariée avec sélection pasà-pas parmi les variables sélectionnées à l'étape 2 (p < 0,05)

GROUPES HOMOGÈNES DE RÉPONSE (1/3)



PRINCIPE

- √ Regrouper les individus qui ont des profils de réponse proche
 - Dans chaque groupe, les réponses des répondants représenteront aussi les nonrépondants

GROUPES HOMOGÈNES DE RÉPONSE (1/3)



PRINCIPE

- √ Regrouper les individus qui ont des profils de réponse proche
 - Dans chaque groupe, les réponses des répondants représenteront aussi les nonrépondants

MÉTHODE DES SCORES

- √ GHR = quantiles de la distribution de la probabilité prédite
 - Limite le poids du modèle
 - Pouvoir trouver des répondants pour représenter les non-répondants pour les probabilités de réponse extrêmes

GROUPES HOMOGÈNES DE RÉPONSE (1/3)



PRINCIPE

- √ Regrouper les individus qui ont des profils de réponse proche
 - Dans chaque groupe, les réponses des répondants représenteront aussi les nonrépondants

MÉTHODE DES SCORES

- √ GHR = quantiles de la distribution de la probabilité prédite
 - Limite le poids du modèle
 - Pouvoir trouver des répondants pour représenter les non-répondants pour les probabilités de réponse extrêmes

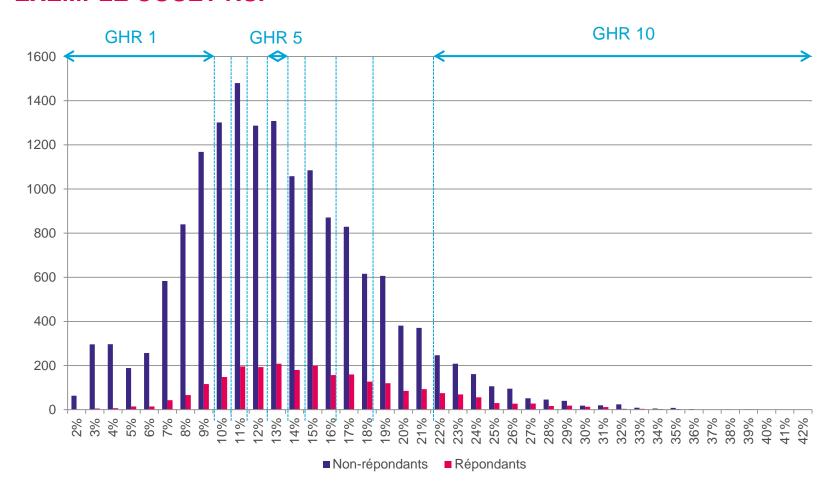
DONNÉES MANQUANTES

- ✓ Données auxiliaires non recueillies :
 - Pour les personnes ayant refusé l'accès à leurs données (1% à 3%)
 - Pour les personnes non informées / non contactées (5% à 10%)
- ✓ Pas de probabilité prédite pour ces personnes → GHR spécifique supplémentaire

GROUPES HOMOGÈNES DE RÉPONSE (2/3)



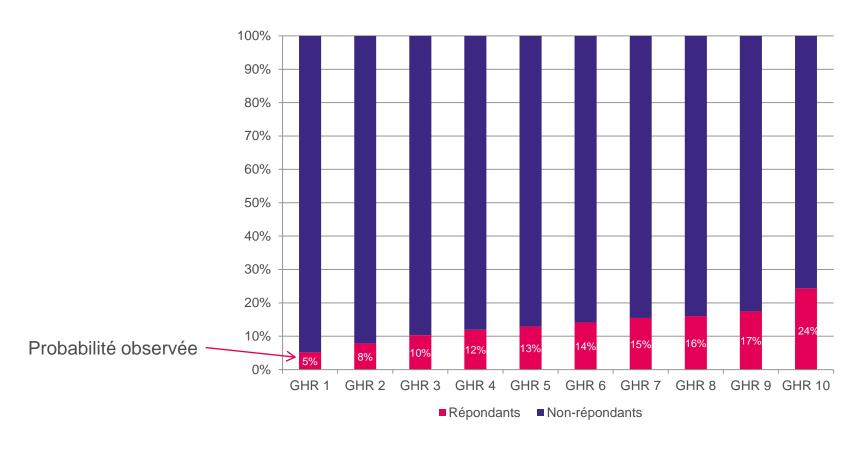
EXEMPLE COSET-RSI



GROUPES HOMOGÈNES DE RÉPONSE (2/3)



EXEMPLE COSET-RSI



REPONDÉRATION



POIDS CORRIGÉS POUR LA NON-RÉPONSE

- √ Facteur correctif = 1 / taux de réponse observé dans le GHR
- ✓ Poids corrigé pour la non-réponse = Poids sondage * facteur correctif

EXEMPLE DE COSET-RSI

	Minimum	Maximum
Taux de réponse observé	5%	24%
Facteur correctif	4,1	19,4
Poids de sondage	14,0	17,2
Poids corrigés	57,4	334,0

✓ Rapport max/min des poids corrigés : 5,8

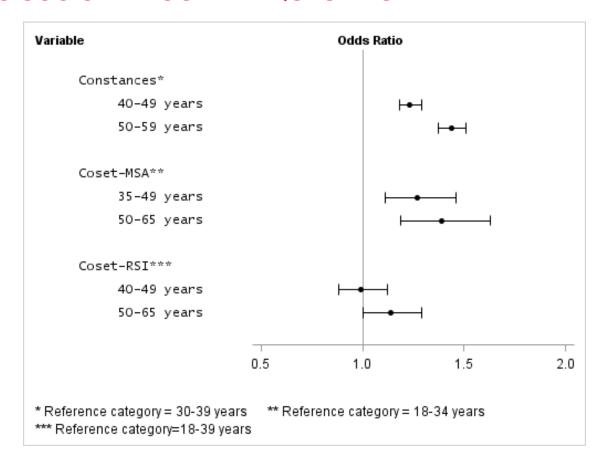


RÉSULTATS

FACTEURS ASSOCIÉS À LA NON-RÉPONSE (1/6)



VARIABLES SOCIO-DÉMOGRAPHIQUES : ÂGE



FACTEURS ASSOCIÉS À LA NON-RÉPONSE (2/6)



VARIABLES SOCIO-DÉMOGRAPHIQUES : SEXE

En univarié : les femmes répondent plus

Cohorto	Taux de réponse		Odds ratio [IC 95%]	
Cohorte	Hommes	Femmes	(ref=hommes)	
Constances	7,0%	7,4%	1,06 [1,03 ; 1,08]	
Coset-MSA	21,6%	28,0%	1,41 [1,28 ; 1,56]	
Coset-RSI	12,2%	15,4%	1,31 [1,20 ; 1,42]	

En multivarié : pas d'association ou association inverse (Constances)

 Association captée par d'autres variables, en particulier les variables de consommation de soins

FACTEURS ASSOCIÉS À LA NON-RÉPONSE (3/6)



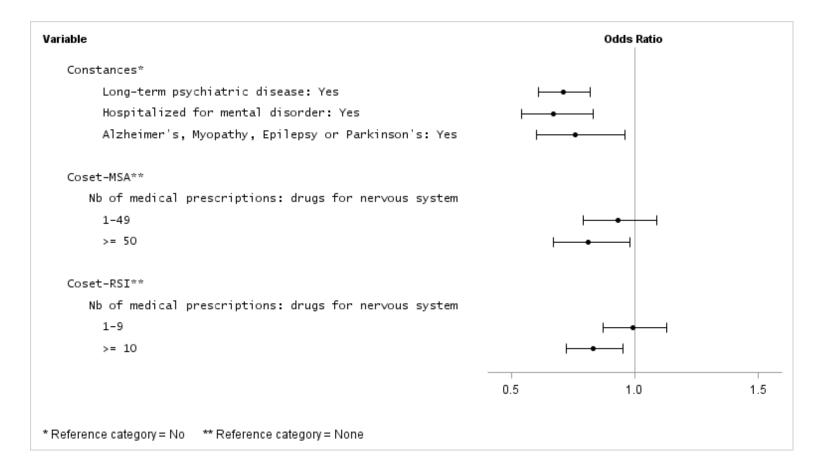
VARIABLES SOCIO-PROFESSIONNELLES

- Les catégories socio-professionnelles les plus aisées ont une propension à répondre plus élevée.
 - Coset-MSA : les salariés du tertiaire et les exploitants polyculture
 - Coset-RSI : les professions libérales
 - Constances : les cadres
- La stabilité dans l'emploi influe positivement sur la réponse à l'enquête.
 - Coset-MSA : durée du dernier emploi ≥ 10 ans
 - Coset-RSI : durée cumulée des périodes d'emploi ≥ 10 ans
 - Constances : travailleurs du secteur public (CAMIEG, MGEN, MFPs)

FACTEURS ASSOCIÉS À LA NON-RÉPONSE (4/6)



VARIABLES SANTÉ: SANTÉ MENTALE



FACTEURS ASSOCIÉS À LA NON-RÉPONSE (5/6)



VARIABLES SANTÉ GÉNÉRALE

• Avoir eu recours à un praticien de ville (généraliste, dentiste, spécialiste) est associé positivement à la probabilité de réponse.

Nombre de consultations chez un dentiste		OR	95% IC
	0	1	-
Constances	1	1,72	[1,65; 1,80]
Constances	2-3	1,77	[1,69; 1,85]
	> 3	1,74	[1,65; 1,83]
	0	1	-
Coset-MSA	1-9	1,32	[1,17; 1,48]
	≥ 10	1,18	[1,01;1,39]
Coset-RSI	0	1	-
Coset-RSI	≥1	1,31	[1,17; 1,46]

FACTEURS ASSOCIÉS À LA NON-RÉPONSE (6/6)



VARIABLES SANTÉ: HOSPITALISATION

 Avoir été hospitalisé ou avoir une maladie grave est associé négativement à la probabilité de réponse.

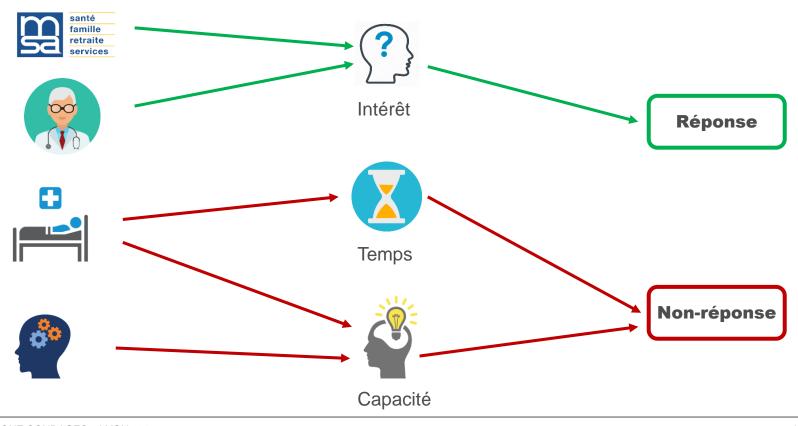
Nombre d'hospitalisations		OR	95% IC
	Aucune	1	-
	Ambulatoire seulement	1,09	[1,03;1,14]
	Moins de 6 nuits	0,95	[0,91;1,00]
Constances	6 nuits ou plus	0,85	[0,78;0,92]
	Séances seulement	0,95	[0,74; 1,22]
	Moins de 6 nuits + séances	0,91	[0,73;1,13]
	6 nuits ou plus + séances	0,61	[0,48 ; 0,78]
	Aucune	1	-
Coset-MSA	1	0,83	[0,72;0,95]
	≥ 2	0,95	[0,79;1,15]
Coost DSI	Aucune dans le public	1	-
Coset-RSI	≥ 1 hospitalisation dans le public	0,82	[0,71;0,94]

FACTEURS ASSOCIÉS À LA NON-RÉPONSE : CONCLUSION



En résumé: Les modèles incluent des variables différentes selon la cohorte, mais on peut souligner des effets similaires.

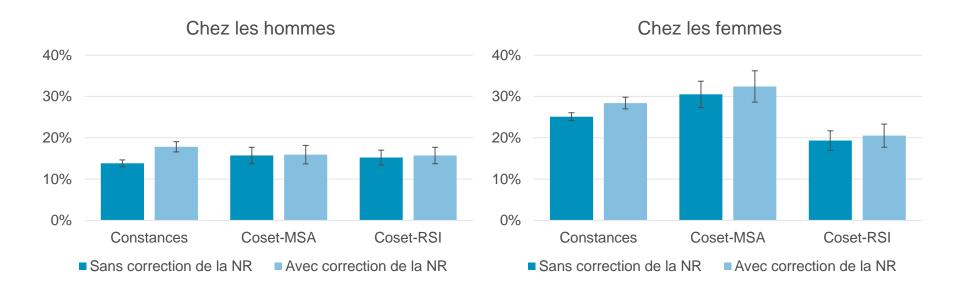
INTERPRÉTATION



ESTIMATION DES PRÉVALENCES (1/3)



SYMPTOMATOLOGIE DÉPRESSIVE (CES-D)

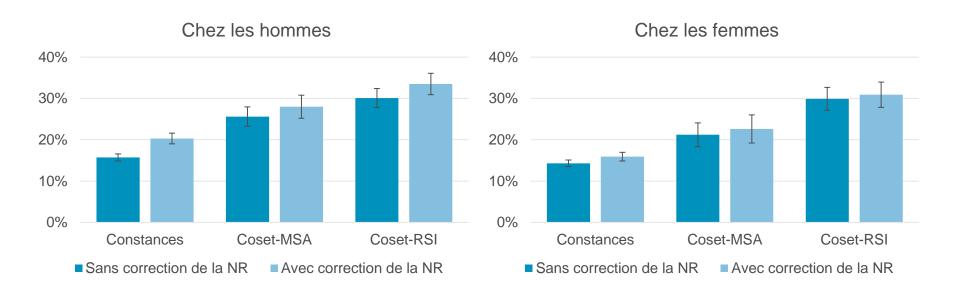


- ✓ Niveaux différents selon les cohortes → populations d'étude différentes
- ✓ La correction de la non-réponse augmente la prévalence dans les 3 cohortes → cohérent avec le modèle de non-réponse

ESTIMATION DES PRÉVALENCES (2/3)



CONSOMMATION DE TABAC

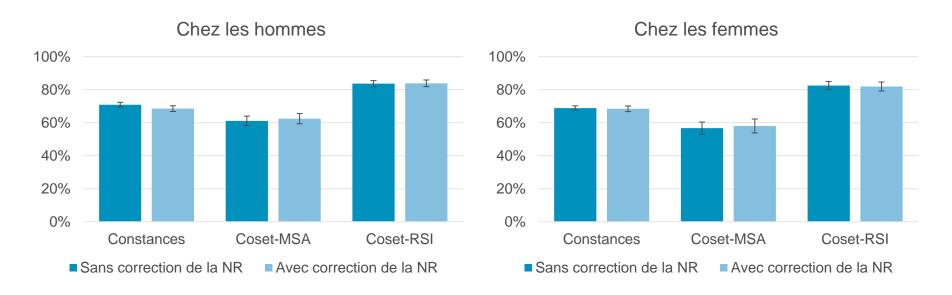


- ✓ Niveaux différents selon les cohortes → populations d'étude différentes
- ✓ La correction de la non-réponse augmente la prévalence dans les 3 cohortes (les fumeurs répondent moins) → associé aux variables de santé incluses dans le modèle de correction de la NR ?

ESTIMATION DES PRÉVALENCES (3/3)



RESPECT MÉRITÉ AU TRAVAIL (ITEM SIEGRIST)



- ✓ La correction de la non-réponse a peu d'effet sur la prévalence (≤ ±2 points).
- ✓ Il n'y a pas suffisamment de variables permettant d'apprécier le ressenti dans le modèle de non-réponse ou alors le respect mérité au travail n'est pas lié à la probabilité de répondre.



- ✓ Présentation des similarités entre les facteurs associés à la non-réponse dans les 3 cohortes
 - Des différences existent, comparaison limitée



- ✓ Présentation des similarités entre les facteurs associés à la non-réponse dans les 3 cohortes
 - Des différences existent, comparaison limitée
- ✓ L'apport des différents types de variables a été étudié dans Constances.
 - Sniiram > Cnav > sociodémographiques



- ✓ Présentation des similarités entre les facteurs associés à la non-réponse dans les 3 cohortes
 - Des différences existent, comparaison limitée
- ✓ L'apport des différents types de variables a été étudié dans Constances.
 - Sniiram > Cnav > sociodémographiques
- ✓ La comparaison des prévalences estimées avec des enquêtes existantes est difficile.
 - Cependant, le sens de l'effet de la correction de la non-réponse est tel qu'attendu selon la littérature



- ✓ Présentation des similarités entre les facteurs associés à la non-réponse dans les 3 cohortes
 - Des différences existent, comparaison limitée
- ✓ L'apport des différents types de variables a été étudié dans Constances.
 - Sniiram > Cnav > sociodémographiques
- ✓ La comparaison des prévalences estimées avec des enquêtes existantes est difficile.
 - Cependant, le sens de l'effet de la correction de la non-réponse est tel qu'attendu selon la littérature
- ✓ Une approche prédictive a également été testée pour la modélisation de la non-réponse dans Constances.
 - Plusieurs méthodes ont été implémentées et donnaient des résultats similaires à l'approche classique.



- ✓ Présentation des similarités entre les facteurs associés à la non-réponse dans les 3 cohortes
 - Des différences existent, comparaison limitée
- ✓ L'apport des différents types de variables a été étudié dans Constances.
 - Sniiram > Cnav > sociodémographiques
- ✓ La comparaison des prévalences estimées avec des enquêtes existantes est difficile.
 - Cependant, le sens de l'effet de la correction de la non-réponse est tel qu'attendu selon la littérature
- ✓ Une approche prédictive a également été testée pour la modélisation de la non-réponse dans Constances.
 - Plusieurs méthodes ont été implémentées et donnaient des résultats similaires à l'approche classique.
- ✓ Un calage a ensuite été effectué pour chaque enquête.
 - Les résultats présentés ici ont utilisé les poids non calés.



- ✓ Le modèle de correction de la non-réponse doit permettre de corriger pour un ensemble large de variables d'intérêt.
 - Défi de trouver les meilleurs indicateurs et de ne pas inclure trop de variables dans le modèle
 - Information auxiliaire riche sur les différents thèmes de l'enquête → plutôt confiant
 - Hypothèse MAR ? Variables explicatives non observables ou non observées



- ✓ Le modèle de correction de la non-réponse doit permettre de corriger pour un ensemble large de variables d'intérêt.
 - Défi de trouver les meilleurs indicateurs et de ne pas inclure trop de variables dans le modèle
 - Information auxiliaire riche sur les différents thèmes de l'enquête → plutôt confiant
 - Hypothèse MAR ? Variables explicatives non observables ou non observées
- ✓ Les données du Sniiram sont des données de remboursement.
 - Utilisation de données administratives/comptables à des fins épidémiologiques
 - Très nombreuses variables → création d'indicateurs généralistes



- ✓ Le modèle de correction de la non-réponse doit permettre de corriger pour un ensemble large de variables d'intérêt.
 - Défi de trouver les meilleurs indicateurs et de ne pas inclure trop de variables dans le modèle
 - Information auxiliaire riche sur les différents thèmes de l'enquête → plutôt confiant
 - Hypothèse MAR ? Variables explicatives non observables ou non observées
- ✓ Les données du Sniiram sont des données de remboursement.
 - Utilisation de données administratives/comptables à des fins épidémiologiques
 - Très nombreuses variables → création d'indicateurs généralistes
- ✓ Disposer du NIR dans la base de sondage a permis un appariement quasi-complet avec le Sniiram et les données professionnelles.



- ✓ Le modèle de correction de la non-réponse doit permettre de corriger pour un ensemble large de variables d'intérêt.
 - Défi de trouver les meilleurs indicateurs et de ne pas inclure trop de variables dans le modèle
 - Information auxiliaire riche sur les différents thèmes de l'enquête → plutôt confiant
 - Hypothèse MAR ? Variables explicatives non observables ou non observées
- ✓ Les données du Sniiram sont des données de remboursement.
 - Utilisation de données administratives/comptables à des fins épidémiologiques
 - Très nombreuses variables → création d'indicateurs généralistes
- ✓ Disposer du NIR dans la base de sondage a permis un appariement quasi-complet avec le Sniiram et les données professionnelles.
- ✓ La méthodologie de correction de la non-réponse a été mise en place sur 3 grandes cohortes, ce qui renforce les résultats et montre l'intérêt de la méthode.

PERSPECTIVES



AMÉLIORATION DES INDICATEURS ET DU MODÈLE DE NON-RÉPONSE

- ✓ Prise en compte de la temporalité des indicateurs de santé notamment.
- √ Modèles stratifiés par sexe

MODÉLISATION LONGITUDINALE DE LA NON-RÉPONSE

✓ Prise en compte de l'attrition

CONCLUSION



RÉSULTATS OBTENUS SUR 3 GRANDES COHORTES

- √ Méthodologie commune
- √ Mise en exergue de facteurs communs associés à la non-réponse

CONCLUSION



RÉSULTATS OBTENUS SUR 3 GRANDES COHORTES

- ✓ Méthodologie commune
- ✓ Mise en exergue de facteurs communs associés à la non-réponse

IMPORTANCE DE PRENDRE EN COMPTE LA NON-RÉPONSE

- ✓ Dans l'estimation de prévalences dans les enquêtes sur la santé
- ✓ La participation aux enquêtes diminue.
- ✓ La non-réponse dépend de l'état de santé.

CONCLUSION



RÉSULTATS OBTENUS SUR 3 GRANDES COHORTES

- ✓ Méthodologie commune
- ✓ Mise en exergue de facteurs communs associés à la non-réponse.

IMPORTANCE DE PRENDRE EN COMPTE LA NON-RÉPONSE

- ✓ Dans l'estimation de prévalences dans les enquêtes sur la santé
- ✓ La participation aux enquêtes diminue.
- ✓ La non-réponse dépend de l'état de santé.

ATOUT RÉEL DES BASES DE DONNÉES MÉDICO-ADMINISTRATIVES

- ✓ Information liée à la santé des répondants et des non-répondants
- ✓ A prévoir en amont, notamment pour récupérer le NIR.

REMERCIEMENTS



PARTENAIRES DES 3 COHORTES :

Coset-MSA: MSA

Coset-RSI: RSI

 Constances : CnamTS, CNAV, ANR, Investissements d'avenir, Université de Versailles-Saint Quentin, Université Paris Descartes

EQUIPE PROJET PREVALEST

IRESP/INCA

REMERCIEMENTS



MERCI À TOUS POUR VOTRE ATTENTION