

ESTIMATION DE PRÉVALENCES DANS CONSTANCES : PREMIÈRES EXPLORATIONS

// ESTIMATION OF PREVALENCES IN CONSTANCES, FIRST EXPLORATIONS

Gaëlle Santin (gaelle.santin@inserm.fr), Eléonore Herquelot, Alice Guéguen, Matthieu Carton, Diane Cyr, Marie Genreau, Stephen Goldberg, Julie Gourmelen, Iris Hourani, Anna Ozguler, Pauline Pascal, Céline Ribet, Elie Salem, Marcel Goldberg, Marie Zins

Unité Cohortes épidémiologiques en population, UMS 11 Inserm-Université Versailles-Saint Quentin, Villejuif, France

Soumis le 29.03.2016 // Date of submission: 03.29.2016

Résumé // Abstract

Introduction – Constances est une cohorte épidémiologique généraliste avec des objectifs analytiques et descriptifs.

Objectif – Estimer des prévalences extrapolables aux personnes âgées entre 30 et 69 ans en 2013 entrant dans le champ de Constances, en prenant en compte la non-participation et en combinant deux années d'invitation.

Méthodes – À partir des invités tirés au sort en 2012 et 2013 âgés de 30 à 69 ans (n=471 152), la participation est étudiée en fonction des données sociodémographiques, des données du Sniiram (Système national d'information inter-régimes de l'Assurance maladie) et de celles de la Cnav (Caisse nationale d'assurance vieillesse). Plusieurs prévalences sont estimées selon le sexe par repondération : sans prise en compte de la non-participation, puis en prenant en compte la non-participation *via* les données sociodémographiques, du Sniiram et de la Cnav.

Résultats – Le taux de participation est de 7,3%. De nombreuses variables du Sniiram et de la Cnav sont associées à la participation dans Constances. La comparaison des prévalences montre que la prise en compte de la non-participation a un impact faible à important, dont le sens peut différer, que ce soit pour des variables mesurées au Centre d'examen de santé ou par auto-questionnaire.

Conclusion – Ces résultats préliminaires montrent que l'apport des données du Sniiram et de la Cnav est important pour étudier les effets de sélection dans Constances et qu'il est en général nécessaire de les prendre en compte pour estimer des prévalences. La combinaison des années d'invitation est à réaliser avec précaution.

Introduction – CONSTANCES is a general purpose epidemiological cohort with analytical as well as descriptive objectives.

Aim – To estimate prevalences which can be extrapolated to people aged 30 to 69 in 2013, within the scope of CONSTANCES by taking into account non-participation and combining two years of invitation.

Methods – From a sample of invited subjects selected randomly in 2012 and 2013 and aged 30 to 69 years (n=471,152), participation was studied as a function of sociodemographic data, as well as data from the SNIIRAM (national inter-scheme health insurance information system) and the CNAV (National Old-Age Insurance Fund). Different sex-specific prevalences were estimated by reweighting: firstly, without accounting for non-participation, and secondly, accounting for non-participation using sociodemographic, SNIIRAM and CNAV data.

Results – The participation rate is 7.3%. Many variables from SNIIRAM and CNAV are associated with participation in the CONSTANCES cohort. A comparison of prevalences shows that taking non-participation into account has a low to important impact, whose direction can differ, whether the variables are measured at the health examination center or collected by self-administered questionnaire.

Conclusion – These preliminary results show that the contribution of SNIIRAM and CNAV data is important in order to study the selection effects in the cohort, and that it is generally required to take them into account to estimate prevalences. The combination of years of invitation should be conducted with caution.

Mots-clés : Biais de sélection, Non-réponse, Bases médico-administratives, Pondération, Prévalence

// **Keywords**: Selection bias, Non-response, Medico-administrative databases, Reweighting, Prevalence

Introduction

Constances est une cohorte épidémiologique généraliste¹. Elle a été construite pour constituer une infrastructure de recherche afin de faciliter des travaux d'épidémiologie analytique et permettre des études de santé publique et de surveillance épidémiologique. C'est

dans ce cadre que de nombreux programmes de santé publique vont s'appuyer sur les données de Constances².

Néanmoins, Constances ne s'inscrit pas dans les standards des enquêtes descriptives à visée représentative pour plusieurs raisons. Le plan de sondage de Constances diffère selon les années et les Centres

d'examen de santé (CES) de la Sécurité sociale en raison de contraintes logistiques diverses. Ces différences conduisent à des poids de sondage dispersés, qui peuvent eux-mêmes engendrer des estimations de variance élevées³. Par ailleurs, le recueil de données de santé mesurées implique que les personnes se rendent dans un CES, ce qui conduit à un taux de participation très faible (environ 7%) généralement observé pour ce type d'enquête⁴. En effet, les personnes participantes doivent accepter de se déplacer dans un CES qui peut être éloigné de leur domicile et difficile d'accès. Cette non-participation peut entraîner des biais importants qu'il faut minimiser autant que possible. En général, les enquêtes épidémiologiques sont sujettes à des biais de sélection, les jeunes, les personnes en moins bonne santé ou socialement défavorisées y participant moins^{5,6}.

Une étude a montré précédemment que le recours à des bases de données administratives telles que le Sniiram (Système national d'information inter-régimes de l'Assurance maladie) était très utile pour corriger la non-participation dans une enquête de surveillance épidémiologique lorsque ces informations étaient disponibles pour les participants et les non-participants⁷.

Enfin, plusieurs projets déposés dans le cadre de l'appel à projets « Constances » avec des objectifs de santé publique requièrent l'estimation de prévalences ou de pathologies dont certaines sont rares ; il sera dans ce cas préférable de combiner plusieurs années d'invitation pour avoir des effectifs suffisamment importants⁸. Ceci nécessite de satisfaire des conditions de stabilité de la population et de la mesure de la variable d'intérêt.

L'objectif de ce travail est de présenter les problèmes rencontrés pour estimer des prévalences dans Constances et les premiers traitements effectués pour corriger la non-participation et combiner deux années d'invitation.

La population sur laquelle seront inférées les prévalences est celle des personnes âgées de 30 à 69 ans en 2013, vivantes au 31 janvier 2014, résidant dans un département couvert par la cohorte Constances et affiliées au régime général de l'Assurance maladie, à la Camieg (Caisse d'assurance maladie des industries électriques et gazières), à la MFP Services (union de mutuelles issues des fonctions publiques d'État, Territoriale et Hospitalière) ou la MGEN (Mutuelle générale de l'Éducation nationale). Les résultats ne sont pas présentés pour les moins de 30 ans. En effet, leur taux de participation était particulièrement faible et il était délicat de faire l'hypothèse que les jeunes participants puissent représenter tous les jeunes de la population d'intérêt.

Matériel et méthodes

La population cible de Constances correspond aux personnes résidant dans un département couvert par la cohorte Constances, affiliées au régime général (au sens large) et âgées de 18 à 69 ans au moment de leur invitation.

Plan de sondage

La base de sondage correspond aux personnes du Répertoire national inter-régimes des bénéficiaires de l'assurance maladie (RNIAM) géré par la Cnav (Caisse nationale d'assurance vieillesse), âgées de 18 à 69 ans, appartenant au régime général (au sens strict) ou à l'une des sections locales mutualistes ayant signé une convention avec Constances (Camieg, LMDE – La mutuelle des étudiants – MFP Services et MGEN), affiliées à l'une des Caisses primaires d'assurance maladie correspondant à un département couvert par Constances et résidant dans l'un d'entre eux¹.

L'inclusion s'étalant sur six ans (2012 à 2017), les clés NIR ont été partitionnées en six groupes distincts et, chaque année, une base de sondage est constituée en sélectionnant les personnes appartenant au groupe de l'année en cours. Elle comprend l'âge et le sexe des personnes et est appariée avec des données du Système national de gestion des carrières géré par la Cnav : dernière catégorie socioprofessionnelle (PCS) connue et activité/inactivité de la personne (ces deux dernières informations combinées seront appelées par la suite typologie d'activité professionnelle).

Tirage au sort des invités

Chaque année, un tirage au sort stratifié à probabilités inégales est réalisé selon les strates suivantes : département de résidence, affiliation, âge, sexe, typologie d'activité professionnelle.

Un courrier d'invitation est envoyé à chaque personne tirée au sort. Si elle souhaite participer à Constances, elle renvoie un coupon-réponse au CES correspondant à son département de résidence. Les personnes invitées ne reçoivent pas de courrier de relance.

Tirage au sort de non-participants

À l'issue de la collecte des données de questionnaires et des données issues des CES, trois motifs de non-participation sont rencontrés :

- courrier d'invitation non distribuable (10%) ;
- refus exprimé de participer (1%) ;
- non renvoi du coupon-réponse (89%).

Après la collecte, un échantillon de non-participants (cohorte témoin) est constitué en juin de l'année suivant l'année d'invitation. Pour les participants ayant donné leur autorisation et pour la cohorte témoin, un appariement avec des données du Sniiram et de la Cnav est réalisé. La cohorte témoin ne couvre que la troisième catégorie de sujets pour des raisons légales (exclusion des personnes non informées et des refus). En raison du volume de données très important généré par le suivi annuel de la cohorte témoin, cette dernière est constituée après un tirage au sort stratifié à probabilités inégales (mêmes strates que celles définies pour le tirage au sort initial) ; pour une année donnée, la taille de la cohorte témoin est deux fois plus grande que celle des participants.

Estimation des prévalences à l'inclusion pour la population cible de Constances âgée de 30 à 69 ans en 2013

Échantillon et pondérations

L'approche classique pour estimer des prévalences concernant la population d'intérêt consisterait à utiliser les données relatives aux invités de 2013 et à appliquer aux participants une pondération correspondant au produit de l'inverse des probabilités d'inclusion et d'un facteur correctif de la non-participation.

Cependant, pour anticiper la nécessité d'obtenir des estimations précises pour les prévalences rares, nous avons travaillé sur un effectif proche du nombre annuel de personnes incluses dans Constances ; ainsi, nous avons adopté une approche moins classique qui consiste à ajouter, aux personnes invitées en 2013, celles invitées en 2012 qui auraient pu être invitées en 2013. Nous avons donc exclu les personnes décédées et les personnes âgées de moins de 30 ans en 2012. C'est à cette population que l'on se réfèrera par la suite pour les invités de 2012. La combinaison de plusieurs années a été déjà proposée pour estimer des prévalences de pathologies ou d'expositions rares⁸. Cette démarche est acceptable car les deux populations sources d'où sont issues les personnes invitées en 2012 et 2013 sont disjointes et correspondent en grande partie à la même population cible.

Le poids de sondage affecté à une personne correspond à l'inverse de sa probabilité d'inclusion.

Un facteur correctif de la non-participation a été estimé en combinant les invités de 2012 et de 2013. Les deux étapes principales sont :

- 1/ prise en compte des « non-renvois de coupon-réponse » : elle consiste à étudier la participation des sujets pour lesquels Constances a accès aux données Sniiram et Cnav :
 - estimation de la probabilité de participation par régression logistique : la variable à expliquer oppose les participants aux non-participants tirés au sort en 2012 et 2013 (affectés de leur poids issu du tirage au sort de la cohorte témoin). Les variables explicatives (80 initialement) sont les variables de stratification, les variables provenant du Sniiram (consommation de soins, hospitalisation, affection de longue durée (ALD) ; les données concernent les trois années glissantes précédant l'invitation) et de la Cnav (emploi, revenus, couverture médicale universelle (CMU)...). Dans un premier temps, des régressions pas à pas descendantes par groupe de variables ont été réalisées (stratification, Cnav, affiliation, consommation de soins, remboursements, hospitalisation et ALD). Dans un deuxième temps, une régression incluant toutes les variables associées à la participation au seuil de 20% a permis de construire le modèle final ;
 - à partir des probabilités prédites par le modèle final, constitution de 20 groupes

homogènes de réponse (GHR) par la méthode des scores⁹. Un premier facteur correctif de la non-participation est estimé par l'inverse du taux de participation estimé par les GHR ;

- 2/ prise en compte des plis non distribuables (PND) et des refus : compte tenu du faible pourcentage de PND et de refus (10%), un calage direct a été préféré à une modélisation fine de ces deux événements. Ainsi, le facteur correctif de la non-participation précédent est calé sur les variables de stratification. La distribution de référence est estimée grâce aux invités de 2012 et 2013. On obtient de cette manière le facteur correctif final pour la non-participation.

Le poids final est égal au produit du poids de sondage et du facteur correctif final pour la non-participation.

Variables d'intérêt

Elles ont été sélectionnées en fonction de leur source et de leur nature diverse (santé, comportement à risque, composante subjective, en lien avec des inégalités sociales de santé etc.). Elles sont issues :

- des données mesurées au CES : LDL-cholestérol, glycémie, statut pondéral ;
- des données de questionnaire : symptomatologie dépressive mesurée par l'échelle CES-D (*Center for Epidemiologic Studies – Depression Scale*)¹⁰, statut tabagique et, pour les personnes salariées en activité professionnelle, le respect mérité au travail (issu du questionnaire Siegrist¹¹).

Prévalences estimées

Pour chaque variable d'intérêt, les prévalences pondérées ont été estimées pour les hommes et pour les femmes sans ($\bar{y}_{\text{poids_sond}}$) et avec correction de la non-participation ($\bar{y}_{\text{poids_cor}}$).

En raison de la faible proportion de données manquantes partielles, ces dernières ont été traitées sous l'hypothèse qu'elles étaient manquantes complètement au hasard.

L'estimation de la variance prend en compte le plan de sondage et la non-participation.

Les changements relatifs liés à la correction de la non-participation ont été estimés par :

$$CR = \frac{\bar{y}_{\text{poids_cor}} - \bar{y}_{\text{poids_sond}}}{\bar{y}_{\text{poids_sond}}} * 100$$

Résultats

Taux de participation

La proportion de PND est de 7,5% parmi les 471 152 invités en 2012 ou 2013 et âgés d'au moins 30 ans au 1^{er} janvier 2013. Le taux de refus est de 1,5% parmi les invités ayant reçu le courrier d'invitation. Parmi les invités n'ayant pas refusé explicitement de participer à Constances, le taux de participation est de 7,4% (n=31 642).

Facteurs associés à la participation

Les données Sniiram et Cnav sont disponibles pour 91,3% des participants, soit 28 914 personnes, en grande partie à cause de procédures d'échange de fichiers non consolidées. Les cohortes témoins de 2012 et 2013 comprennent 60 782 personnes.

La probabilité de participer à Constances est associée, dans le modèle multivarié, à des variables socio-démographiques, socioprofessionnelles, relatives au recours aux soins, aux hospitalisations et aux ALD, soit 36 variables conduisant à 80 degrés de liberté. Les principales sont présentées figures 1, 2 et 3 ; les variables relatives aux causes d'hospitalisation ainsi que les ALD les moins fréquentes ne sont pas intégrées dans les figures.

Concernant les variables sociodémographiques, dans le modèle final, la probabilité de participer est associée aux CES ; elle est plus élevée chez les hommes et augmente avec l'âge.

La probabilité de participation est par ailleurs associée à la catégorie sociale (OR=0,5 ; IC95%: [0,4-0,6] chez les ouvriers *versus* les cadres). Elle est plus élevée chez les personnes ayant un revenu important lié à l'emploi ou à la retraite (OR=1,4 [1,2-1,6]) pour les revenus supérieurs à 2 400 euros *versus* aucun ; elle est en revanche moins élevée chez les personnes ayant eu au moins un remboursement au titre de la CMU (OR=0,8 [0,7-0,9]).

Pour les variables relatives au recours aux soins, la probabilité de réponse est plus élevée chez les personnes ayant recours à un médecin généraliste (OR=1,3 [1,2;1,4]

au moins 6 fois *versus* aucune) ou spécialiste (OR=1,9 [1,8;2,0]) au moins 4 fois *versus* aucune) ou à un dentiste (OR=1,7 [1,6;1,8] au moins 4 fois *versus* aucune) ; elle est en revanche moins élevée chez les personnes ayant eu des visites à domicile de médecin généraliste (OR=0,3 [0,2;0,4]) au moins 2 fois *versus* aucune). Pour ce qui concerne les hospitalisations, la probabilité de participer est moins élevée chez les personnes ayant été hospitalisées au moins 7 jours (OR=0,9 [0,8;0,9]). Plusieurs ALD sont associées à une probabilité de participer moins élevée : diabète (OR=0,6 [0,5;0,7]), maladie cardiovasculaire (OR=0,7 [0,6;0,8]), infection par le VIH (OR=0,6 [0,4;0,8]), maladie d'Alzheimer, myopathie, épilepsie ou maladie de Parkinson (OR=0,8 [0,6;1,0]), affection psychiatrique de longue durée (OR=0,7 [0,6;0,8]).

Estimation des prévalences sans et avec correction de la non-participation

Le tableau présente les estimations des prévalences des variables d'intérêt selon le sexe, sans et avec correction de la non-participation, ainsi que les changements relatifs associés à chaque modalité de variable.

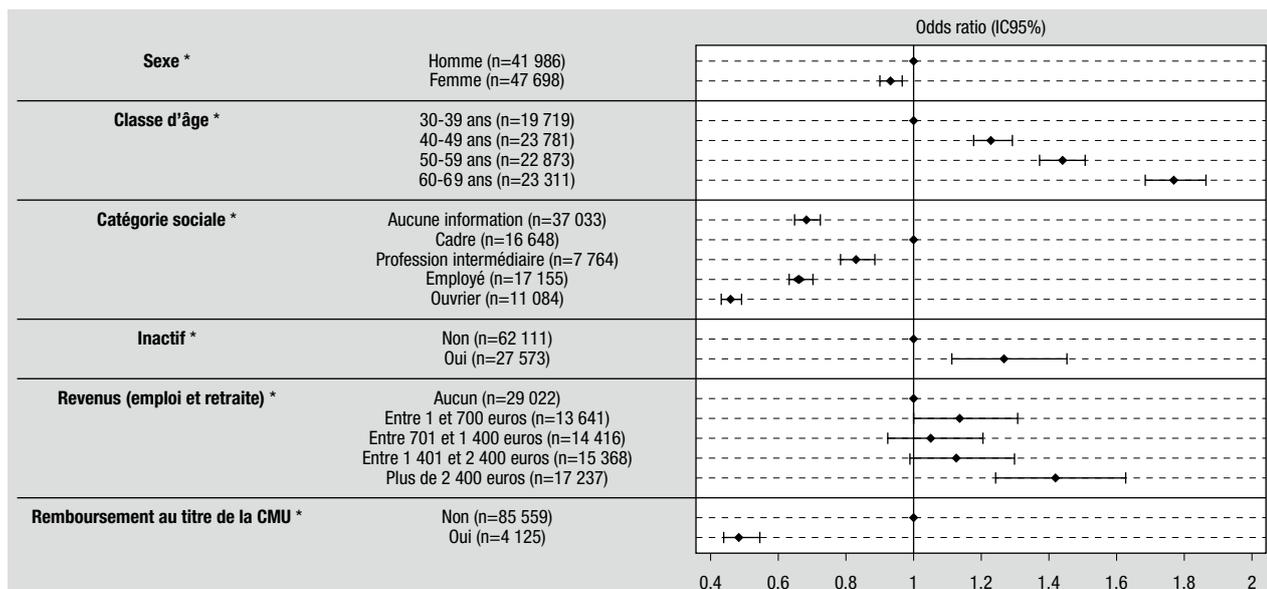
Les changements relatifs sont faibles à importants et peuvent différer selon le sexe. Le sens des changements est négatif ou positif selon les variables.

Variables mesurées au CES

Les changements relatifs sont importants (supérieurs à 15%) pour l'hyperglycémie et l'obésité, les prévalences sans prise en compte de la non-participation étant sous-estimées. Pour le LDL-cholestérol, les changements relatifs sont faibles

Figure 1

Association de la participation dans Constances avec des variables socioéconomiques et le recours aux soins (soins de ville, hospitalisation, ALD). Partie 1 : variables socioéconomiques



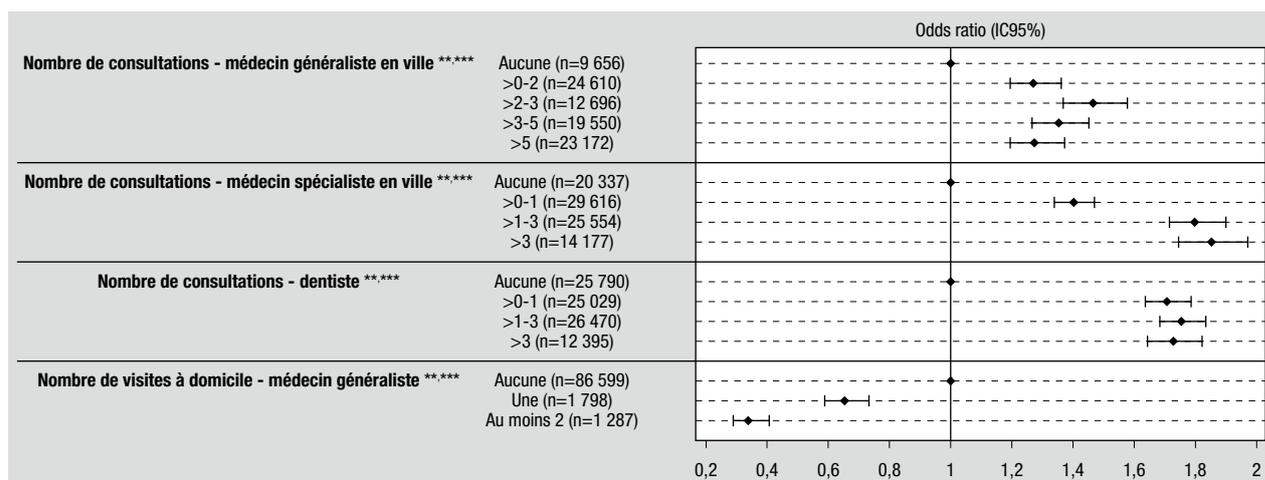
NB : odds ratios (OR) issus du modèle final. IC95% : intervalle de confiance à 95%.

* Variable issue de la Cnav.

ALD : affection de longue durée ; CMU : couverture médicale universelle.

Figure 2

Association de la participation dans Constances avec des variables socioéconomiques et le recours aux soins (soins de ville, hospitalisation, ALD). Partie 2 : variables de consommation de soins de ville



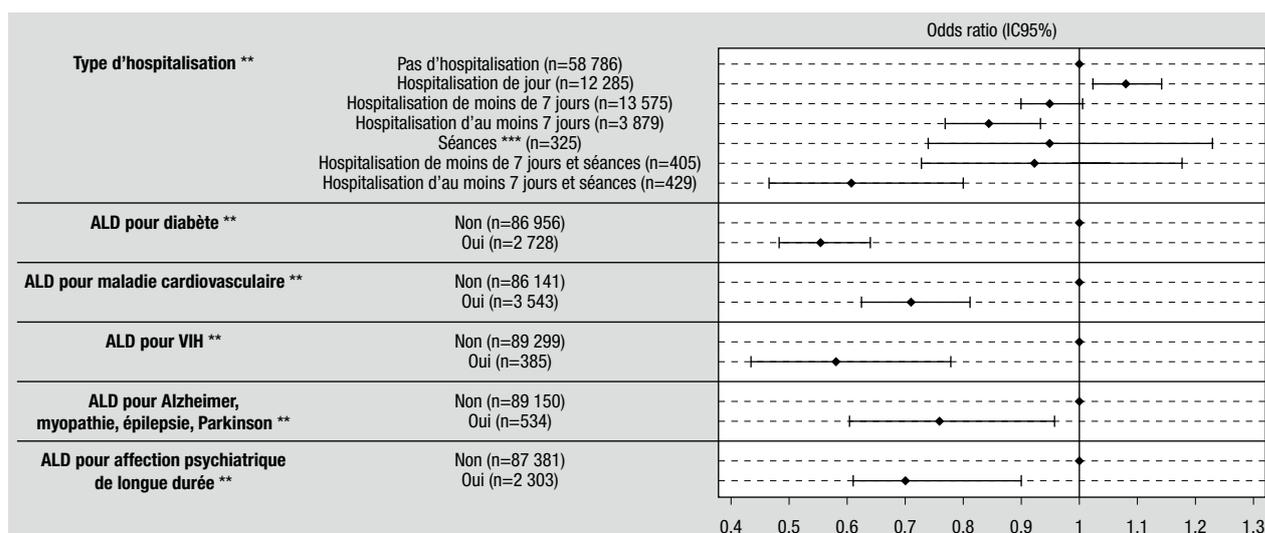
NB : odds ratios (OR) issus du modèle final. IC95% : intervalle de confiance à 95%.

** Variable issue du Sniiram ; *** Nombre moyen de consultations durant les trois années glissantes précédant l'invitation.

ALD : affection de longue durée.

Figure 3

Association de la participation dans Constances avec des variables socioéconomiques et le recours aux soins (soins de ville, hospitalisation, ALD). Partie 3 : variables d'hospitalisation et d'ALD



NB : odds ratios (OR) issus du modèle final. IC95% : intervalle de confiance à 95%.

** Variable issue du Sniiram ; *** Une séance est une hospitalisation d'une journée ou d'une nuit répétée régulièrement pendant une certaine durée (par exemple pour une chimiothérapie).

ALD : affection de longue durée.

pour les hommes (-2,1%) et modérés pour les femmes (-7,5%).

Variables issues des questionnaires

Pour la variable « respect que je mérite au travail », les changements relatifs sont faibles (3,4% pour les hommes ; 0,5% pour les femmes).

Pour la symptomatologie dépressive, ils sont importants et différents (22,4% pour les hommes, 10,3% pour les femmes) ; la prévalence de symptomatologie dépressive est estimée à 13,8% chez les hommes et à 25,1% chez

les femmes, sans correction de la non-participation, et à 17,8% chez les hommes et à 28,4% chez les femmes, avec correction de la non-participation.

Si on s'intéresse au statut tabagique au moment de l'enquête, les changements relatifs sont importants : avec correction de la non-participation, le pourcentage de fumeurs estimé est plus élevé et le pourcentage d'ex-fumeurs plus faible.

Prévalences selon le sexe estimées à partir des données Constances sans et avec correction de la non-participation parmi les 30-69 ans affiliés au régime général de l'Assurance maladie (sens large) de 16 départements

	Sans correction de la non-participation				Avec correction de la non-participation				Changement relatif	
	Homme (n=13 486)		Femme (n=15 428)		Homme (n=13 486)		Femme (n=15 428)			
	%	IC95%	%	IC95%	%	IC95%	%	IC95%	Homme	Femme
LDL-cholestérol (seuil)										
≤4,2 mmol/L (n=21 473)	78,4	(77,4-79,4)	80,5	(79,6-81,4)	78,9	(77,7-80,0)	81,9	(80,8-83,0)	0,6	1,6
>4,2 mmol/L (n=5 440)	21,6	(20,6-22,6)	19,5	(18,6-20,4)	21,1	(20,0-22,3)	18,1	(17,0-19,2)	- 2,1	- 7,5
Statut pondéral										
Dénutrition (n=532)	0,9	(0,7-1,2)	3,1	(2,7-3,5)	1,2	(0,9-1,7)	3,3	(2,9-3,9)	25,0	6,1
Valeurs de référence (n=14 055)	42,3	(41,1-43,5)	59,0	(57,9-60,1)	41,8	(40,4-43,2)	55,6	(54,2-59,1)	- 1,2	- 6,1
Surpoids (n=9 039)	43,3	(42,1-44,5)	25,2	(24,2-26,1)	41,0	(39,6-42,5)	25,4	(24,2-26,7)	- 5,6	0,8
Obésité (n=3 519)	13,5	(12,7-14,4)	12,8	(12,1-13,5)	16,0	(14,8-17,1)	15,7	(14,5-16,8)	15,6	18,5
Symptomatologie dépressive										
Non (n=21 601)	86,2	(85,3-87,0)	74,9	(73,9-75,8)	82,2	(80,9-83,4)	71,6	(70,2-73,0)	- 4,9	- 4,5
Oui (n=5 267)	13,8	(13,0-14,7)	25,1	(24,2-26,1)	17,8	(16,6-19,1)	28,4	(27,0-29,8)	22,4	11,4
Statut tabagique										
Non-fumeur (n=12 185)	40,6	(39,4-41,9)	55,6	(54,4-56,8)	40,0	(38,5-41,5)	55,2	(53,7-56,7)	- 1,7	- 0,7
Fumeur (n=3 882)	15,7	(14,9-16,6)	14,3	(13,5-15,1)	20,3	(19,0-21,6)	15,9	(14,9-17,0)	22,3	10,3
Ex-fumeur (n=9 271)	43,7	(42,4-44,9)	30,1	(29,1-31,2)	39,8	(38,3-41,2)	28,9	(27,6-30,2)	- 9,7	- 4,4
Respect que je mérite au travail*										
Non (n=4 856)	29,2	(27,8-30,7)	31,3	(29,9-32,6)	31,5	(29,8-33,3)	31,6	(29,9-33,3)	7,4	1,1
Oui (n=11 093)	70,8	(69,3-72,2)	68,8	(67,4-70,1)	68,5	(66,7-70,2)	68,4	(66,7-70,1)	- 3,4	- 0,5

* Question posée uniquement aux personnes en activité professionnelle.

Discussion – conclusion

La participation dans Constances est associée à des variables sociodémographiques, socioprofessionnelles et à des variables relatives à la santé. Leur prise en compte par repondération dans l'estimation des prévalences montre des différences faibles à importantes lorsqu'on les compare aux prévalences brutes. Ces différences reflètent le lien entre certaines variables explicatives de la participation et les variables d'intérêt.

La cohorte Constances, de par sa taille et ses nombreuses données collectées, permettra de compléter les informations de santé publique déjà disponibles et issues d'enquêtes transversales répétées, comme l'Enquête santé et protection sociale (ESPS) de l'Irdes (Institut de recherche et documentation en économie de la santé) et le Baromètre santé de Santé publique France. Le faible taux de participation observé dans Constances est contrebalancé par deux atouts majeurs : la disponibilité de données mesurées lors de l'examen de santé et l'appariement aux grandes bases de données médico-administratives. Cet appariement permet non seulement d'enrichir les données de la cohorte de participants, mais également de disposer de données pertinentes pour corriger la non-participation.

Les facteurs associés à la participation dans Constances sont concordants avec ceux rapportés dans la littérature épidémiologique^{7,12,13} : les personnes qui prennent soin de leur santé participent plus, alors que les jeunes, les personnes en moins bon état de santé ou plus défavorisées socialement ou financièrement participent moins aux enquêtes de santé. L'odds ratio de participation associé au sexe suggère une participation plus élevée chez les hommes que chez les femmes ; ce résultat est inhabituel⁶. Néanmoins, en univarié, on trouve le résultat inverse ; le sens de l'association s'inverse après ajustement sur les variables relatives à la santé. Un résultat similaire avait été obtenu dans l'enquête Coset-MSA : en multivarié, l'association entre la participation et le sexe devenait non-significative après ajustement sur les variables de santé⁷. Ainsi, les variables de santé semblent être des facteurs intermédiaires importants expliquant l'association entre le sexe et la participation. L'étude de cette association après ajustement sur les variables de santé devra être approfondie, car le modèle utilisé ici n'avait pas pour objectif d'étudier cette question en particulier.

Par ailleurs, la prise en compte des facteurs associés à la participation dans l'estimation des prévalences entraîne des modifications attendues sauf, à première vue, pour le LDL-cholestérol. En effet, après correction de la non-participation,

la prévalence de LDL-cholestérol varie peu, alors que des variables qui lui sont associées comme la glycémie, qui est un marqueur du diabète, ou le statut tabagique¹⁴ voient leurs prévalences modifiées. Lorsqu'on étudie le LDL-cholestérol comme une variable continue, les changements sont également faibles (moyenne pour les hommes de 3,52 mmol/l sans correction vs 3,50 mmol/l après correction et, pour les femmes, 3,44 mmol/l sans correction vs 3,39 mmol/l après correction) ; néanmoins, les moyennes estimées dans Constances sont proches de celles estimées dans l'enquête ENNS 2006 (3,31 mmol/l pour les hommes et 3,28 mmol/l pour les femmes)¹⁵. L'absence de modification nécessite des analyses complémentaires, en étudiant par exemple séparément les personnes remboursées pour un traitement anti-cholestérol des personnes qui ne le sont pas.

En ce qui concerne la symptomatologie dépressive et l'obésité, l'ampleur et le sens de la modification de l'estimation de prévalence apportée par la correction de la non-participation étaient attendus ; en effet, il a déjà été montré dans la littérature que les personnes avec une affection psychologique ou obèses participent moins aux enquêtes¹³. Nous ne disposons malheureusement pas d'enquêtes de référence pour comparer les prévalences de symptomatologie dépressive. En revanche, pour l'obésité, les prévalences obtenues dans Constances sont comparables à celles estimées dans l'étude ObÉpi – enquête nationale sur l'obésité et le surpoids (14,3% pour les hommes et 15,7% pour les femmes)¹⁶ ; elles sont supérieures dans Constances, mais l'étude ObÉpi inclut les 18-29 ans et la prévalence d'obésité augmente avec l'âge.

Il en est de même pour le statut tabagique, puisqu'après correction de la non-participation, la proportion de fumeurs augmente tandis que celle des ex-fumeurs diminue. Il est néanmoins possible que le pourcentage de fumeurs soit sous-estimé ; il est en effet plus faible que celui obtenu dans le Baromètre santé 2010 (30,8% pour les hommes, 24,1% pour les femmes)¹⁷. Cependant les populations-cibles, les méthodologies d'enquête et les questions posées ne sont pas exactement les mêmes.

L'absence de variations pour le « respect que je mérite au travail » était également attendue et cohérente avec une étude antérieure⁷. Ceci peut s'expliquer par le fait que cette variable est issue d'une échelle mesurant les contraintes psychosociales ; il est donc possible que les données du Sniiram et de la Cnav ne soient pas suffisantes pour corriger la non-participation pour ce type de questions comportant une composante subjective. En ce sens, la cohérence de ces résultats ne garantit pas que tous les facteurs liés à la participation et susceptibles d'entraîner des biais soient pris en compte dans nos analyses. Ainsi, même si la correction de la non-participation permet de corriger de façon plus ou moins importante les biais de sélection, chaque prévalence obtenue doit faire l'objet d'un examen attentif et doit être interprétée autant que possible en la comparant à des valeurs de référence

ou, à tout le moins, en fonction de la littérature sur les effets de sélection. En ce sens, les corrections de non-participation peuvent être insuffisantes pour une variable d'intérêt donnée si on pense qu'un groupe de personnes ne participe pas pour des raisons liées à des facteurs non inclus dans les données du Sniiram ou de la Cnav.

La validité de la combinaison des deux années 2012 et 2013 repose sur des hypothèses supplémentaires que nous avons jugées plausibles. D'une part, la population source de 2012 doit avoir la même structure que celle de 2013 et les changements de strates d'une année à l'autre doivent être négligeables. D'autre part, la variable d'intérêt ne doit pas être affectée par un effet période entre 2012 et 2013 ; autrement dit, en adoptant une approche contrefactuelle¹⁸, un sujet invité en 2012 né une année donnée aurait répondu en moyenne comme un sujet invité en 2013 né un an plus tard et ayant exactement les mêmes caractéristiques influant sur la variable d'intérêt que le sujet précédent. La combinaison de deux ou plusieurs années conduit à des effectifs importants, ce qui peut être un avantage lorsqu'on s'intéresse à des pathologies rares, mais peut présenter dans les autres cas un risque pour un utilisateur non averti ; en effet, un effectif important conduit à des intervalles de confiance étroits qui pourraient être considérés au pied de la lettre sans tenir compte des autres incertitudes, en particulier celles dues à la non-participation et à sa modélisation.

Ces travaux préliminaires sont encourageants et devront être complétés par une étude plus fine des pondérations : modèles plus complexes, analyses de robustesse, de sensibilité et réflexion sur les analyses par domaine. La richesse des données du Sniiram et de la Cnav permettra de construire des modèles spécifiques pour une population ou une pathologie donnée.

Par ailleurs, les épidémiologistes spécialistes d'une thématique pourront extrapoler, grâce à un calage supplémentaire, les résultats à l'ensemble des affiliés du Régime général, voire à l'ensemble de la population française, s'ils considèrent que les biais engendrés par ces extrapolations sont négligeables. ■

Remerciements

Ce travail bénéficie du financement de l'Institut national du cancer (INCa) dans le cadre du projet Prevalest – *Prevalence estimation using data from individual surveys and administrative databases* (AAP 2013 « Prévention primaire »). Les auteurs remercient L. Bénézet, J. Chatelot, B. Geoffroy-Perez, J.L. Marchand et N. Soullier de Santé publique France, partenaire de ce projet.

Références

[1] Zins M, Goldberg M, Carton M, Guéguen A, Henny J, Le Got S, *et al.* La cohorte Constances : une infrastructure pour la recherche et la santé publique. *Bull Epidemiol Hebd.* 2016; (35-36):612-6. http://invs.santepubliquefrance.fr/beh/2016/35-36/2016_35-36_1.html

[2] Équipe Constances. Projets acceptés par le Conseil scientifique international. [Internet]. <http://www.constances.fr/espace-scientifique/projets-acceptes.php>

- [3] Ardilly P. Présentation des plans de sondage classiques. In: Ardilly P. Les techniques de sondage. Paris: Editions Technip; 1994. p. 47-93.
- [4] Meffre C. Prévalence des hépatites B et C en France en 2004. Saint-Maurice: Institut de veille sanitaire; 2006. 176 p. http://opac.invs.sante.fr/index.php?lvl=notice_display&id=4362
- [5] Strandhagen E, Berg C, Lissner L, Nunez L, Rosengren A, Torén K, *et al.* Selection bias in a population survey with registry linkage: Potential effect on socioeconomic gradient in cardiovascular risk. *Eur J Epidemiol.* 2010;25(3):163-72.
- [6] Van Loon AJM, Tjihuis M, Picavet HSJ, Surtees PG, Ormel J. Survey non-response in the Netherlands: Effects on prevalence estimates and associations. *Ann Epidemiol.* 2003;13(2):105-10.
- [7] Santin G, Geoffroy B, Benezet L, Delezire P, Chatelot J, Sitta R, *et al.* In an occupational health surveillance study, auxiliary data from administrative health and occupational databases effectively corrected for nonresponse. *J Clin Epidemiol.* 2014;67(6):722-30.
- [8] Smith P, Lynn P, Elliot D. Sample design for longitudinal surveys. In: Lynn P, editor. *Methodology of longitudinal surveys.* Chichester: Wiley; 2009. p. 21-32.
- [9] Eltinge JL, Yansaneh IS. Diagnostics for formation of nonresponse adjustment cells, with an application to income nonresponse in the U.S. consumer expenditure survey. *Survey Methodol.* 1997;23:33-40.
- [10] Radloff LS. The CES-D Scale: A self-report depression scale for research in the general population. *Appl Psych Meas.* 1977;1(3):385-401.
- [11] Siegrist J, Wege N, Pühlhofer F, Wahrendorf M. A short generic measure of work stress in the era of globalization: effort-reward imbalance. *Int Arch Occup Environ Health.* 2009;82(8):1005-13.
- [12] Martikainen P, Laaksonen M, Piha K, Lallukka T. Does survey non-response bias the association between occupational social class and health? *Scand J Public Health.* 2007;35(2):212-5.
- [13] Vercambre MN, Gilbert F. Respondents in an epidemiologic survey had fewer psychotropic prescriptions than nonrespondents: An insight into health-related selection bias using routine health insurance data. *J Clin Epidemiol.* 2012;65(11):1181-9.
- [14] de Oliveira Fontes GL, Neuberger M, Tichy A, Moshammer H. Cross-sectional association between cigarette smoking and abdominal obesity among Austrian bank employees. *BMJ Open.* 2014;4(7):e004899. <http://bmjopen.bmj.com/content/4/7/e004899.long>
- [15] Castetbon K, Hercberg C; Usen. Étude nationale nutrition-santé, ENNS, 2006. Situation nutritionnelle en France selon les indicateurs d'objectif et les repères du Programme national nutrition-santé (PNNS). Saint-Maurice: Institut de veille sanitaire; 2007. 74 p; http://opac.invs.sante.fr/index.php?lvl=notice_display&id=3793
- [16] Inserm, Kantar Health, Roche. ObÉpi 2012 : Enquête épidémiologique nationale sur le surpoids et l'obésité. 2012. 60 p. http://www.roche.fr/content/dam/roche_france/fr_FR/doc/obepi_2012.pdf
- [17] Guignard R, Beck F, Richard JB, Peretti-Watel P. Le tabagisme en France : analyse de l'enquête Baromètre santé 2010. Saint-Denis: Inpes; 2013. 56 p. <http://inpes.santepubliquefrance.fr/CFESBases/catalogue/pdf/1513.pdf>
- [18] Hernan MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? *Epidemiology.* 2006;17(4):360-72.

Citer cet article

Santin G, Herquelot E, Guéguen A, Carton M, Cyr D, Genreau M, *et al.* Estimation de prévalences dans Constances : premières explorations. *Bull Epidémiol Hebd.* 2016;(35-36):622-9. http://invs.santepubliquefrance.fr/beh/2016/35-36/2016_35-36_3.html