

PRISE EN COMPTE DE LA NON-PARTICIPATION DANS CONSTANCES POUR L'ESTIMATION DE PREVALENCES : PREMIERES EXPLORATIONS

Gaëlle Santin¹, Eléonore Herquelot², Alice Guéguen³, Matthieu Carton⁴, Diane Cyr⁵, Marie Genreau⁶, Julie Gourmelen⁷, Iris Hourani⁸, Anna Ozguler⁹, Marcel Goldberg¹⁰, Marie Zins¹¹

¹UMS 011 Inserm UVSQ "Cohortes épidémiologiques en population" Hôpital Paul Brousse -
Bâtiment 15/16 16 avenue Paul Vaillant Couturier 94807 VILLEJUIF CEDEX

gaelle.santin@inserm.fr

²UMS 011 Inserm UVSQ "Cohortes épidémiologiques en population" Hôpital Paul Brousse -
Bâtiment 15/16 16 avenue Paul Vaillant Couturier 94807 VILLEJUIF CEDEX

eleonore.herquelot@inserm.fr

³UMS 011 Inserm UVSQ "Cohortes épidémiologiques en population" Hôpital Paul Brousse -
Bâtiment 15/16 16 avenue Paul Vaillant Couturier 94807 VILLEJUIF CEDEX

alice.gueguen@inserm.fr

⁴UMS 011 Inserm UVSQ "Cohortes épidémiologiques en population" Hôpital Paul Brousse -
Bâtiment 15/16 16 avenue Paul Vaillant Couturier 94807 VILLEJUIF CEDEX

matthieu.carton@inserm.fr

⁵UMS 011 Inserm UVSQ "Cohortes épidémiologiques en population" Hôpital Paul Brousse -
Bâtiment 15/16 16 avenue Paul Vaillant Couturier 94807 VILLEJUIF CEDEX

diane.cyr@inserm.fr

⁶UMS 011 Inserm UVSQ "Cohortes épidémiologiques en population" Hôpital Paul Brousse -
Bâtiment 15/16 16 avenue Paul Vaillant Couturier 94807 VILLEJUIF CEDEX

marie.genreau@inserm.fr

⁷UMS 011 Inserm UVSQ "Cohortes épidémiologiques en population" Hôpital Paul Brousse -
Bâtiment 15/16 16 avenue Paul Vaillant Couturier 94807 VILLEJUIF CEDEX

julie.gourmelen@inserm.fr

⁸UMS 011 Inserm UVSQ "Cohortes épidémiologiques en population" Hôpital Paul Brousse -
Bâtiment 15/16 16 avenue Paul Vaillant Couturier 94807 VILLEJUIF CEDEX

iris.hourani@inserm.fr

⁹UMS 011 Inserm UVSQ "Cohortes épidémiologiques en population" Hôpital Paul Brousse -
Bâtiment 15/16 16 avenue Paul Vaillant Couturier 94807 VILLEJUIF CEDEX

anna.ozguler@inserm.fr

¹⁰UMS 011 Inserm UVSQ "Cohortes épidémiologiques en population" Hôpital Paul Brousse -
Bâtiment 15/16 16 avenue Paul Vaillant Couturier 94807 VILLEJUIF CEDEX

marcel.goldberg@inserm.fr

¹¹UMS 011 Inserm UVSQ "Cohortes épidémiologiques en population" Hôpital Paul Brousse -
Bâtiment 15/16 16 avenue Paul Vaillant Couturier 94807 VILLEJUIF CEDEX

marie.zins@inserm.fr

Résumé.

Introduction : Constances est une cohorte épidémiologique généraliste avec des objectifs analytiques mais aussi descriptifs.

Objectif : Il est de prendre en compte la non-participation pour estimer des prévalences extrapolables aux personnes âgées entre 30 et 69 ans en 2013, résidant dans un des départements Constances, affiliées au Régime Général, à la CAMIEG, la MFPs ou la MGEN.

Méthodes : A partir des invités tirés au sort en 2013 âgés de 30 à 69 ans (n=245986), la participation est étudiée en fonction des données sociodémographiques, du SNIIRAM et de la CNAV. 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 7,0%. 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, que ce soit pour des variables mesurées au CES ou par autoquestionnaire.

Mots-clés.biais de sélection, non-réponse, bases médico-administratives, repondération, prévalence

1 Introduction

La cohorte Constances est une cohorte épidémiologique généraliste[1]. 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 Constances [2].

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) en raison de contraintes logistiques diverses ; ces différences conduisent à des poids de sondage dispersés qui peuvent eux-mêmes engendrer des estimations de variance élevées [3]. 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'enquête [4] ; ceci peut entraîner des biais de non-participation importants qu'il faut minimiser autant que possible. Une étude a montré précédemment que le recours à des bases de données administratives telles que le SNIIRAM était très utile pour corriger la non-participation lorsque ces informations étaient disponibles pour les participants et les non-participants [5].

L'objectif de ce travail est de présenter les problèmes rencontrés pour estimer des prévalences dans Constances et les premiers traitements effectués. 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 Constances et affiliées au Régime Général, à la Camieg, la MFPs ou la MGEN. 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.

2 Matériel et méthodes

La population cible de Constances correspond aux personnes résidant dans un département Constances, affiliées au Régime Général de la Sécurité Sociale (au sens large) et âgées de 18 à 69 ans au moment de leur invitation.

2.1 Plan de sondage

La base de sondage correspond aux personnes du Répertoire national inter-régimes des bénéficiaires de l'assurance maladie (Rniam), âgées de 18 à 69 ans, appartenant au Régime général (au sens strict) ou à l'une des SLM ayant signé une convention avec Constances (Camieg, LMDE, MFPs et MGEN), affiliées à l'une des CPAM couvrant un département Constances et résidant dans l'un d'entre eux[1].

L'inclusion s'étalant sur 6 ans (2012 à 2017), les clés NIR ont été partitionnées en 6 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 (SNGC) : dernière PCS connue et activité/inactivité de la personne (ces deux dernières informations combinées seront appelées par la suite typologie d'activité professionnelle).

2.1.1 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.

2.1.2 Tirage au sort de non-participants

A l'issue de la collecte, trois motifs de non-participation sont rencontrés :

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

Après la collecte, on constitue un échantillon de non-participants (cohorte témoin), apparié avec des données du SNIIRAM et de la CNAV. 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.

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

Dans ce travail, la population d'intérêt correspond aux personnes âgées de 30 à 69 ans en 2013, vivantes au 31 janvier 2014, résidant dans un département Constances, et affiliées au Régime Général, à la Camieg, la MFPs ou la MGEN.

2.2.1 Échantillon et pondérations

L'approche classique pour estimer des prévalences concernant la population d'intérêt consiste 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.

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 deux étapes :

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 (affectés de leur poids issu du tirage au sort de la cohorte témoin).
Les variables explicatives sont les variables de stratification, les variables provenant du SNIIRAM (consommation de soins, hospitalisation, ALD ; les données concernent les 3 années précédant l'invitation) et de la CNAV (emploi, revenus, CMU-C...)
- À partir des probabilités prédites par le modèle précédent, constitution de 20 groupes homogènes de réponse (GHR) par la méthode des scores [6].

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 courriers non distribuables et des refus*

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 2013. On obtient ainsi 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.

2.2.2 Variables d'intérêt

Elles sont issues :

- des données mesurées au CES : glycémie, statut pondéral
- des données de questionnaire : symptomatologie dépressive mesurée par le CES-D (seuil anglo-saxon), statut tabagique, et pour les personnes salariées en emploi, le respect mérité au travail.).

2.2.3 Prévalences estimées

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

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 :

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

3 Résultats

3.1 Taux de participation

La proportion de courriers non distribuables est de 6,5% parmi les 245986 invités en 2013 et âgés d'au moins 30 ans au 1^{er} janvier 2013. Le taux de refus est de 1,3% 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 7,0% (n=15890).

3.2 Facteurs associés à la participation

Les données SNIIRAM et CNAV sont disponibles pour 88% des participants, soit 13983 personnes, en grande partie à cause de procédures d'échange de fichiers non consolidées.

La probabilité de participer à Constances est associée, dans le modèle multivarié, à des variables sociodémographiques, socioprofessionnelles, relatives au recours aux soins, aux hospitalisations et aux affections de longue durée (ALD). Les principales sont présentées Figures 1, 2 et 3.

Concernant les variables sociodémographiques, la probabilité de participer est associée aux CES et augmente avec l'âge.

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 IC 95%(1,1 ; 1,4) au moins 5 fois versus aucune) ou spécialiste (OR=1,8 IC 95%(1,7 ; 2,0) au moins 3 fois versus aucune) ou à un dentiste (OR=1,8 IC 95%(1,6 ; 1,9) au moins 3 fois versus aucune) ; elle est en revanche moins élevée chez les personnes ayant eu des visites à domicile de médecin généraliste (OR=0,3 IC 95%(0,2 ; 0,4) au moins 2 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=0,8 IC 95%(0,7 ; 0,9)). Plusieurs ALD sont associées à une probabilité de participer moins élevée : diabète (OR=0,6 IC 95%(0,5 ; 0,7)), maladie cardio-vasculaire (OR=0,7 IC 95%(0,6 ; 0,9)), affection psychiatrique de longue durée (OR=0,8 IC 95%(0,6 ; 0,9)).

La probabilité de participation est par ailleurs associée à la catégorie sociale (OR=0,4 IC 95%(0,3 ; 0,5) chez les ouvriers versus les cadres). Elle est plus élevée chez les personnes ayant un revenu lié à l'emploi ou à la retraite important (OR=1,4 IC 95%(1,1 ; 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-C (OR=0,4 IC 95%(0,3 ; 0,5)).

Figure 1 : Variables socio-économiques associées à la participation

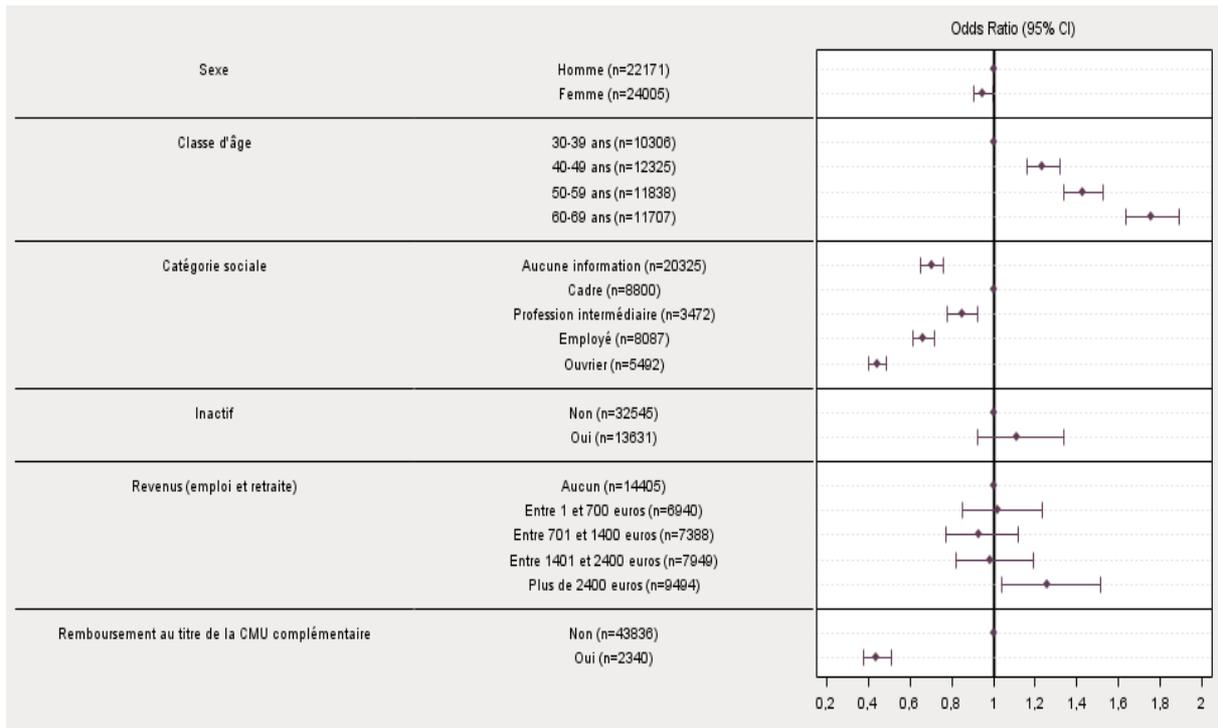
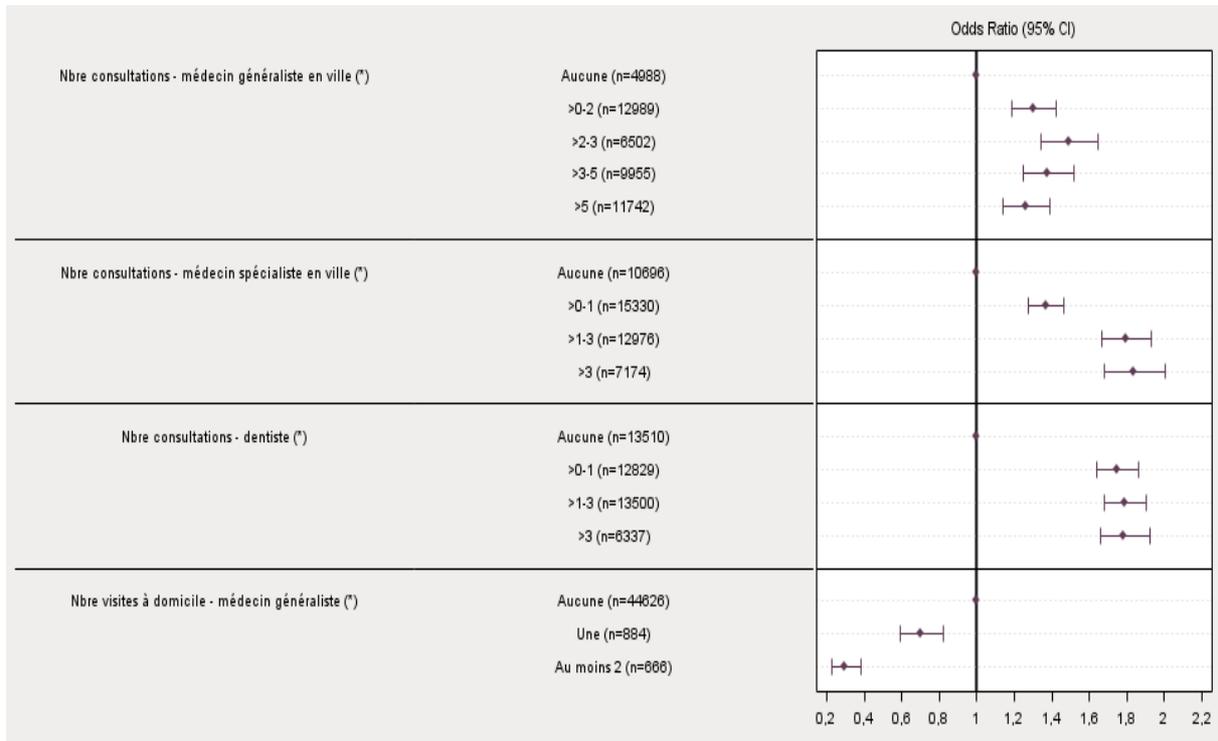
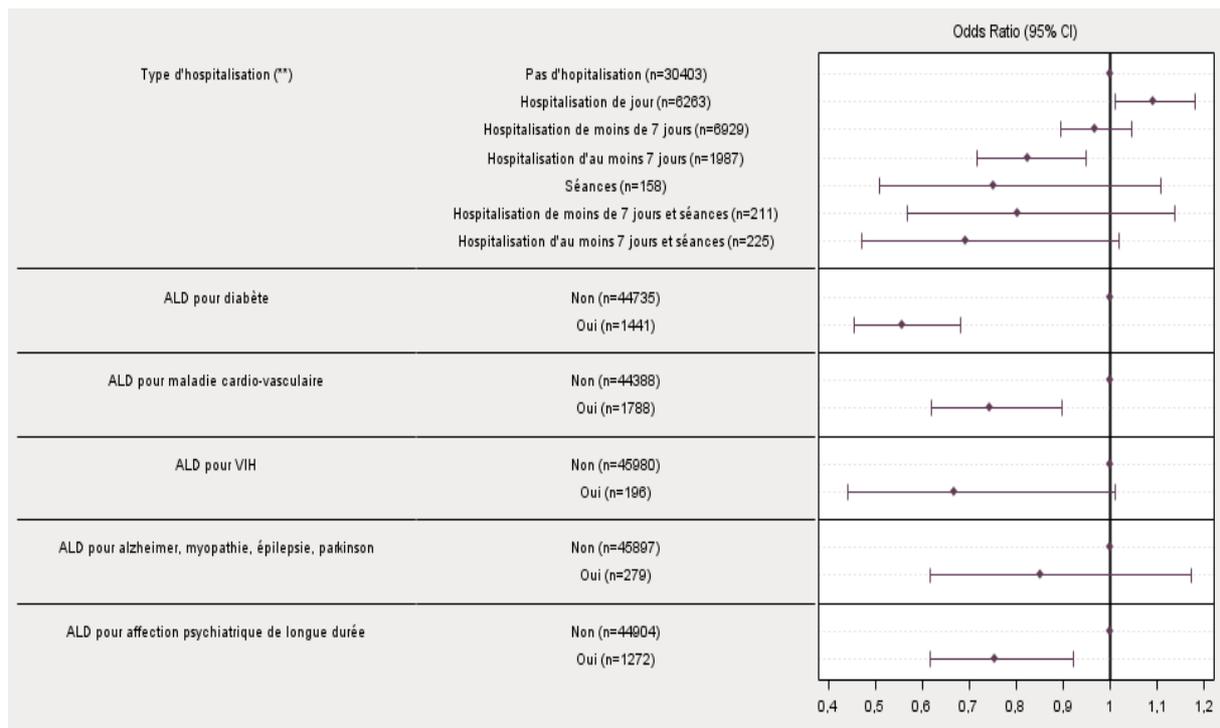


Figure 2 : Variables de consommation de soins de ville associées à la participation



(*) Nombre moyen de consultations entre 2010 et 2013

Figure 3 : Variables d'hospitalisation et d'ALD associées à la participation



(**) Une séance est une hospitalisation d'une journée ou d'une nuit habituellement pendant une certaine durée

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

Le tableau 1 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.

Pour les variables mesurées au CES, les changements relatifs sont importants (supérieurs à 10%) pour l'hyperglycémie et l'obésité, les prévalences sans prise en compte de la non-participation étant sous-estimées.

Pour la variable « respect que je mérite au travail », les changements relatifs sont faibles (-2,6% pour les hommes ; -1,9% pour les femmes).

Pour la symptomatologie dépressive, ils sont importants et différents (24,9% pour les hommes, 5,2% pour les femmes) ; la prévalence de symptomatologie dépressive est estimée à 14,2% chez les hommes et à 25,8% chez les femmes sans correction de la non-participation et à 18,9% 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.

Tableau 1 : Prévalences selon le sexe estimées sans et avec correction de la non-participation

	Sans correction de la non-participation				Avec correction de la non-participation				Changement relatif	
	Homme		Femme		Homme		Femme		Homme	Femme
	%	IC 95%	%	IC 95%	%	IC 95%	%	IC 95%		
Glycémie										
Hypoglycémie (n=46)	0,3	[0,1 ; 0,7]	0,5	[0,3 ; 1,0]	0,2	[0,1 ; 0,5]	0,5	[0,3 ; 0,8]	-50,0	0,0
Glycémie normale (n=13057)	94,0	[93,1 ; 94,8]	97,7	[97,2 ; 98,2]	92,9	[91,7 ; 93,9]	97,3	[96,6 ; 97,9]	-1,2	-0,4
Hyperglycémie (n=486)	5,7	[4,9 ; 6,6]	1,7	[1,4 ; 2,2]	6,9	[5,9 ; 8,0]	2,2	[1,7 ; 2,9]	17,4	22,7
Typologie de masse corporelle										
Maigre (n=246)	0,8	[0,5 ; 1,2]	3,0	[2,5 ; 3,6]	0,9	[0,5 ; 1,4]	3,3	[2,7 ; 4,0]	11,1	9,1
Normal (n=6709)	42,3	[40,6 ; 44,0]	59,9	[58,2 ; 61,5]	42,0	[40,0 ; 44,0]	56,1	[54,1 ; 58,1]	-0,7	-6,8
Surpoids (n=4282)	43,0	[41,3 ; 44,8]	24,1	[22,7 ; 25,5]	41,5	[39,5 ; 43,6]	24,6	[22,8 ; 26,3]	-3,6	2,0
Obèse (n=1715)	13,9	[12,8 ; 15,2]	13,0	[12,0 ; 14,2]	15,6	[14,1 ; 17,2]	16,0	[14,5 ; 17,7]	10,9	18,8
Symptomatologie dépressive										
Non (n=10457)	85,8	[84,5 ; 86,9]	74,2	[72,7 ; 75,7]	81,1	[79,2 ; 82,9]	71,6	[69,7 ; 73,4]	-5,8	-3,6
Oui (n=2594)	14,2	[13,1 ; 15,5]	25,8	[24,3 ; 27,3]	18,9	[17,1 ; 20,8]	28,4	[26,6 ; 30,3]	24,9	9,2
Statut tabagique										
Non fumeur (n=5820)	37,9	[36,2 ; 39,6]	50,1	[48,4 ; 51,8]	37,1	[35,1 ; 39,1]	49,5	[47,6 ; 51,5]	-2,2	-1,2
Fumeur (n=2312)	17,7	[16,4 ; 19,0]	17,1	[15,9 ; 18,3]	22,3	[20,6 ; 24,2]	19,2	[17,7 ; 20,7]	20,6	10,9
Ex-fumeur (n=5121)	44,4	[42,7 ; 46,2]	32,8	[31,3 ; 34,4]	40,6	[38,7 ; 42,6]	31,3	[29,5 ; 33,1]	-9,4	-4,8
Respect que je mérite au travail										
Non (n=2333)	28,5	[26,5 ; 30,6]	30,7	[28,7 ; 32,7]	30,3	[28,0 ; 32,7]	32,0	[29,6 ; 34,5]	5,9	4,1
Oui (n=5409)	71,5	[69,4 ; 73,5]	69,3	[67,3 ; 71,3]	69,7	[67,3 ; 72,0]	68,0	[65,5 ; 70,4]	-2,6	-1,9

4 Discussion - Conclusion

La participation dans Constances est associée à des variables sociodémographiques, socio-professionnelles et des variables relatives à la santé. Leur prise en compte par pondération dans l'estimation des prévalences montre des différences faibles à importantes lorsqu'on les compare aux prévalences brutes. Ces différences reflètent le lien entre 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 en France et issues d'enquêtes transversales répétées, comme l'enquête ESPS et le Baromètre Santé. 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 permet également de disposer de données pertinentes pour corriger la non-participation. Constances a reçu le label d'intérêt général et de qualité statistique du Conseil national de l'information statistique (N°18/Label/L201 du 13/01/2015).

Les facteurs associés à la participation dans Constances sont concordants avec ceux rapportés dans la littérature épidémiologique [5;7;8] ; les personnes qui prennent soin de leur santé participent plus alors que les personnes en moins bon état de santé ou étant plus défavorisées socialement ou financièrement participent moins aux enquêtes de santé.

Par ailleurs, la prise en compte de ces facteurs dans l'estimation des prévalences entraîne des modifications attendues, comme par exemple, l'augmentation de la prévalence de la symptomatologie dépressive ou de l'obésité (mesurée au CES). Il en est de même pour les comportements à risque tel que le statut tabagique, puisque après correction de la non-participation, la proportion de fumeurs augmente tandis que celle des ex-fumeurs diminue. L'absence de variations pour le « respect que je mérite au travail » était également attendu et cohérent avec une étude antérieure [5]. 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.

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. Par exemple, il est possible que le pourcentage de fumeurs soit sous-estimé ; il est en effet plus faible que celui obtenu dans le Baromètre santé 2010 [9]. Cependant les populations cibles, les méthodologies d'enquête et les questions posées ne sont pas exactement les mêmes. Par ailleurs si un groupe entier de personnes n'a pas été observé, les corrections de non-participation sont insuffisantes.

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.

5 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.
- (2) Equipe Constances. Projets validés par le Conseil scientifique international. 1-3-2016.
Ref Type: Online Source
- (3) Ardilly P. Présentation des plans de sondage classiques. 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; 2016.
- (5) 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 Jun;67(6):722-30.
- (6) 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.
- (7) 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.
- (8) 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.
- (9) Guignard R, Beck F, Richard J, Peretti-Watel P. Le tabagisme en France : analyse de l'Enquête Baromètre Santé 2010. Saint-Denis: Inpes; 2013.