Mise en place d'un double jeu de pondérations pour la deuxième vague de l'enquête Santé et itinéraire professionnel (SIP)

Nicolas DE RICCARDIS[1]

1) Présentation de l'enquête SIP

1-Objectif

SIP est une enquête innovante qui a pour objet les interactions entre le travail, l'emploi et la construction ou l'altération de la santé. La seconde vague de l'enquête, réalisée en 2010, a été conçue dans une double perspective : longitudinale, avec la ré-interrogation des personnes interrogées en 2006 et transversale avec des questions sur l'exposition aux risques psychosociaux (RPS) pour les personnes ayant un emploi, en écho aux travaux du Collège de suivi statistique des facteurs de RPS au travail. Cette double utilisation possible des données de la seconde vague de SIP nécessite donc la mise en place de deux jeux de pondération.

2-Champ

L'enquête a été effectuée auprès d'individus vivant en ménage ordinaire en France métropolitaine. Les individus âgés de 20 à 74 ans en 2006 ont été interrogés, quelle que soit leur situation vis-à-vis du marché du travail (actifs occupés, chômeurs ou inactifs, retraités, etc.). Près de 14 000 entretiens ont été réalisés en 2006 ; 11 000 personnes ont pu être réinterrogées en 2010.

Pour satisfaire le second objectif concernant l'exposition aux RPS, un échantillon complémentaire de personnes actives occupées d'au moins 16 ans vivant en ménage ordinaire a été constitué à partir des données de la Taxe d'habitation (TH). Environ 2 500 personnes ont répondu, ce qui conduit au total à environ 8 800 personnes ayant un emploi en 2010 ayant été interrogées sur les RPS.

2) Pondération longitudinale

La partie longitudinale est composée des répondants aux deux vagues de l'enquête et des personnes décédés entre les deux vagues. La mise en place des pondérations se fait en deux étapes.

1-Traitement de la non-réponse :

Le comportement de réponse à la deuxième vague est modélisé à l'aide d'une régression logistique sur les variables région, âge, situation face à l'emploi, santé déclarée et niveau de diplôme.

La régression permet de définir 12 groupes de réponses homogènes (GRH).

[1] Tél. : 01 40 56 88 13
Mail : nicolas.dericcardis@sante.gouv.fr
Direction de la Recherche, des Études, de l'Évaluation et des Statistiques (Drees)
Sous-direction de l'Observation de la santé et de l'assurance maladie
Bureau État de santé de la population

2-Calage :
On procède ensuite au calage sur marge, à partir des variables suivantes de l'enquête Emploi : ZEAT, tranches d'unité urbaine, diplôme, sexe par tranche d'âge, nationalité, puis la taille et le type de ménage.

Il a été choisi de ne pas effectuer un calage sur les catégories socioprofessionnelles, qui pourront ainsi servir ainsi pour le contrôle *a posteriori* sur le calage.

3) Pondération transversale

Traitement de la non-réponse pour l'échantillon complémentaire TH d'actifs occupés
Cette fois encore, on part de l'hypothèse que la non-réponse n'est pas uniforme. On crée des GRH en modélisant la non-réponse à l'aide d'une régression logistique sur les variables suivantes issues de la TH : taille du ménage, revenu du ménage et Région.
La régression permet de définir 7 GRH.

Partage des poids
L'observation transversale des RPS concerne uniquement les personnes en emploi au moment de l'enquête, soit l'échantillon complémentaire TH et les actifs occupés de la partie panel. La méthode du partage des poids permet de tenir compte des liens existants entre les échantillons initial et complémentaire.

Calage
Il n'y a pas de recueil ni du diplôme ni de la nationalité pour l'échantillon complémentaire. Les variables de calage seront donc différentes de celles utilisées pour la partie longitudinale. On retient les variables suivantes : ZEAT, tranches d'unité urbaine, sexe par tranche d'âge, catégorie socioprofessionnelle, puis la taille et le type de ménage.