

Traitement de la non-réponse par imputation multiple dans une enquête en population générale

Noémie SOULLIER¹

Les enquêtes sont très souvent confrontées à des cas de non-réponse totale ou partielle. Les analyses statistiques doivent tenir compte de cette non-réponse, afin de réduire le biais qu'elle pourrait engendrer.

Dans l'enquête Handicap-Santé en ménages ordinaires (INSEE, 2008-2009), les personnes interrogées devaient préciser le nombre d'heures d'aide apportée, que ce soit par les aidants professionnels ou ceux de l'entourage. S'agissant de l'aide apportée par l'entourage, le taux de non-réponse s'est toutefois avéré important, en particulier lorsque l'aidant cohabitait avec la personne aidée. Les nombre d'heures d'aide des répondants à la question étant possiblement différents de ceux des non-répondants, l'analyse ne peut pas être effectuée uniquement sur les répondants (analyse en cas complets) et une imputation multiple des nombres d'heures manquants est mise en place.

L'imputation multiple consiste à remplacer chaque valeur manquante par plusieurs valeurs plausibles, en se basant sur l'information collectée par ailleurs. Elle permet de tirer profit de toute l'information renseignée pour estimer le nombre d'heures d'aide informelle, tout en tenant compte de l'incertitude de la valeur imputée résultant du fait que celle-ci n'est pas observée. Le modèle d'imputation inclut à la fois les variables expliquant la non-réponse et celles expliquant la variable à imputer, soit ici une vingtaine de variables caractérisant la personne aidée et son degré de dépendance. Des modèles différents ont été ajustés pour les aidants cohabitants et les non-cohabitants. L'imputation multiple est réalisée via l'algorithme MCMC avec chaînes multiples, en incluant les variables qualitatives sous forme d'indicateurs. La variable « nombre d'heures » est log-transformée avant l'imputation, afin de s'approcher d'une distribution normale (la transformation inverse est appliquée à la sortie). Vingt imputations ont été effectuées.

Compte tenu du taux de non-réponse plus faible lorsqu'il s'agit d'aidants non-cohabitants et de la présence de l'information sur la fréquence de l'aide (nombre de visites par jour/semaine/mois) pour ces aidants, l'estimation obtenue est plus précise pour cette catégorie d'aidants. Les résultats présentent séparément les volumes horaires d'aide apportée aux personnes âgées selon que leurs aidants résident avec elles ou non, ainsi qu'un volume horaire total pour l'aide apportée par l'entourage.

La moitié des personnes âgées aidées par leur entourage pour les tâches de la vie quotidienne reçoit au moins l'équivalent d'1h40 par jour d'aide de ces proches. Le volume de cette aide dépend du degré de dépendance de la personne âgée : la moitié des personnes âgées très dépendantes (GIR 1-2) aidées par un ou plusieurs proches reçoit au moins l'équivalent de 5h15 d'aide de ces proches par jour, alors que la moitié des personnes âgées

¹Ministère des Solidarités et de la Cohésion sociale,

Direction de la Recherche, des Études, de l'Évaluation et des Statistiques,

Sous-direction « *Observation de la Solidarité* », Bureau Famille Handicap Dépendance

les moins dépendantes (GIR 5-6) aidées par un ou plusieurs proches en reçoit au moins l'équivalent de 1h10 par jour. Par ailleurs, le volume médian de l'aide apportée par les aidants cohabitants est deux fois supérieur à celui de l'aide apportée par les aidants non-cohabitants (respectivement l'équivalent de 2h10 par jour et 1h par jour).

La suite du travail devra s'attacher à développer une analyse de sensibilité, afin d'observer les variations de ces résultats lorsque l'on modifie des paramètres du modèle.