

# ENQUÊTE HANDICAP-SANTÉ : APPLICATION D'UNE MÉTHODE SUR "PETITS DOMAINES" À DES ESTIMATIONS LOCALES

Josiane Le Guennec <sup>1</sup>

<sup>1</sup> *Insee, 36 place du Colombier, 35000 Rennes, josiane.le-guennec@insee.fr*

Une enquête sur le Handicap et la Santé a été réalisée en 2008 par l'Insee et la Drees sur l'ensemble du territoire, auprès d'un échantillon de ménages. Cette enquête comprenait deux phases. La première phase, dite "Vie quotidienne et santé" (VQS), réalisée avec un questionnaire auto-administré, visait à repérer les personnes susceptibles de se trouver en situation de handicap, afin de les réinterroger prioritairement en deuxième phase. La phase 2, dite "Handicap, santé des ménages" (HSM) a été réalisée par enquêteur auprès des personnes sélectionnées dans l'échantillon des répondants à l'enquête VQS, selon un plan de sondage surreprésentant les groupes à plus fortes déficiences. Les prévalences de handicap par type dans la population nationale sont fournies par cette deuxième phase.

Une modélisation des données recueillies dans l'échantillon national a été tentée afin de répondre à des demandes d'information locale. La méthode d'estimation retenue repose sur une approche prédictive, avec modélisation explicite des comportements individuels. Le modèle consiste à expliquer la situation de santé observée en deuxième phase par les réponses recueillies en première phase, ainsi que par quelques critères sociodémographiques de l'individu ou du ménage disponibles dans l'enquête ou dans la base de sondage. La valeur du paramètre est alors imputée, à l'aide du modèle, aux individus répondants en première phase et non réinterrogés en deuxième phase. L'extrapolation à la population totale du département se fait ensuite à l'aide des pondérations de l'échantillon de première phase.

$$\hat{Y}_D = \frac{1}{\tilde{N}_D} \left( \sum_{k \in s_{2D}} w_{1k} y_k + \sum_{i \in s_{1D} \cap \bar{s}_{2D}} w_{1i} \tilde{y}_i \right),$$

où

- $y_k = 1$  si l'individu  $k$  présente le handicap étudié, d'après sa réponse à l'enquête HSM, et  $y_k = 0$  sinon ;
- $\tilde{y}_i = \widehat{Prob}\{Y_i = 1\} = \frac{1}{1 + \exp(-[\hat{b}'x_i + \nu_d])}$  pour tout individu  $i$  répondant à VQS et non sélectionné en deuxième phase, où  $\hat{b}$  et  $\nu_d$  sont les coefficients obtenus dans la régression logistique entre le paramètre  $Y$  et un vecteur  $x_i$  de variables auxiliaires, admettant le département de résidence en effet aléatoire ;
- $w_{1k}, w_{1i}$  sont les poids redressés des individu  $k$  et  $i$  dans l'enquête de 1ère phase VQS ;
- $\tilde{N}_d = \sum_{i \in s_{1D} \cap U_D} w_{1i}$  est la population cible dans le département  $D$ , estimée à partir de l'enquête VQS.

L'estimateur final supporte une variance de sondage réduite par rapport à l'estimateur direct, la taille de l'échantillon de première phase étant beaucoup plus importante, mais la possibilité d'un biais dû au modèle. L'article présentera le modèle retenu, les difficultés rencontrées et les résultats obtenus.