

## Article

Symposium 2008 :  
Collecte des données : défis, réalisations et nouvelles orientations

### **Sélection de modèles pour l'évaluation des hypothèses qui soutiennent les méthodes permettant de compenser pour les valeurs manquantes dans les enquêtes par sondage**

par Philip J. Smith

2009



## **Sélection de modèles pour l'évaluation des hypothèses qui soutiennent les méthodes permettant de compenser pour les valeurs manquantes dans les enquêtes par sondage**

Philip J. Smith <sup>1</sup>

### **Résumé**

Les valeurs manquantes attribuables à la non-réponse constituent un type d'erreur non due à l'échantillonnage. Le rejet de cas comportant des valeurs manquantes dans des analyses statistiques peut entraîner des estimations biaisées en raison des différences entre les répondants pour lesquels des valeurs sont manquantes et les autres. De plus, lorsque des variables présentent différentes structures de valeurs manquantes parmi les cas échantillonnés et lorsque les cas comportant des valeurs manquantes sont rejetés des analyses statistiques, celles-ci peuvent produire des résultats incohérents puisqu'elles reposent sur des sous-ensembles différents de cas échantillonnés pouvant ne pas être comparables. Cependant, les analyses qui rejettent des cas avec valeurs manquantes peuvent être valides à la condition que ces valeurs soient manquantes entièrement au hasard. Or, est-ce bien le cas?

En guise de compensation, les valeurs manquantes sont souvent imputées ou les poids de l'enquête sont ajustés à l'aide de méthodes de catégories de poids. Les analyses subséquentes peuvent être valides à condition que les valeurs soient manquantes au hasard dans chaque catégorie de données que supposent les variables indépendantes des modèles qui reposent sur des techniques d'ajustement. Or, peut-on dire que les valeurs sont manquantes au hasard?

Comme les valeurs manquantes ne sont pas observées, les hypothèses sur les valeurs manquantes entièrement au hasard et les valeurs manquantes au hasard formulées dans les analyses ne sont pas souvent examinées. La communication décrit un modèle de sélection à partir duquel des tests d'hypothèse concernant les valeurs manquantes entièrement au hasard et les valeurs manquantes au hasard peuvent être réalisés même si les valeurs manquantes ne sont pas observées. Les données de la National Immunization Survey menée par le Department of Health and Human Services des États-Unis sont utilisées pour illustrer ces méthodes.

---

<sup>1</sup> Philip J. Smith, Centers for Disease Control and Prevention, Atlanta, GA É.-U.