Letters to the Editor

IMPLICATIONS OF MISSING INCOME DATA

We strongly support the recommendation of the recent article by Kim et al.[1] that health researchers pay heed to the strong social patterning of missing data often exhibited by key variables in epidemiologic studies. Yet, although we agree with the authors' advice that researchers should "at a minimum, carefully examine characteristics of respondents with missing income information," we must strongly caution readers against their recommendation to "routinely includ[e] a separate income category of respondents with missing income information in all analyses."

Contradicting this recommendation is extensive literature on missing data,[2–7] including two articles cited by Kim et al.[2,4] This research has shown that this "missing indicator" method will result in biased effect estimates under most conditions. In particular, bias occurs even when the missing data are missing completely at random (e.g., people missing data on income are a random sample from all income groups). If the data are missing at random (e.g., missingness on income depends only on variables that are observed), then multiple imputation or weighting techniques can be used to obtain valid effect estimates.[6] If, however, the data are not missing at random (e.g., there are additional unobserved predictors of missingness), more complex models for non-ignorable nonresponse are required,[7] and, ultimately, investigators would do well to adopt a sensitivity analysis framework. In all cases, researchers must give serious thought to data quality issues raised by the extent and patterning of missingness, the pathways leading to this missingness, and the implications for valid causal inference.

In summary, the problem of the social patterning of missing data that Kim et al. highlight is very real and troubling for epidemiologic research and underscores why epidemiologists cannot afford to ignore poverty and its impact on both health status and causal inference.[8] To do the research right, we must use appropriate methods. In 1995, Greenland and Finkle[2] were alarmed to find that "the indicator method [was] widely perceived as a formally correct method of handling missing values." Because this view continues to persist in some quarters, we must reiterate their recommendation that epidemiologists avoid the "potentially disastrous ad hoc" missing indicator approach.

Jarvis T. Chen, ScD[a]
Afamia Kaddour, MPH[a]
Nancy Krieger, PhD[a]

[a]Harvard School of Public Health, Department of Society, Human Development and Health, Boston, MA

REFERENCES

1. Kim S, Egerter S, Cubbin C, Takahashi ER, Braveman P. Potential implications of missing income data in population-based surveys: an example from a postpartum survey in California. Public Health Rep 2007;122:753-63.
2. Greenland S, Finkle WD. A critical look at methods for handling missing covariates in epidemiologic regression analyses. Am J Epidemiol 1995;142:1255-64.
3. Jones MP. Indicator and stratification methods for missing explanatory variables in multiple linear regression. J Am Stat Assn 1996;91:222-30.
4. Vach W, Blettner M. Biased estimation of the odds ratio in case-control studies due to use of ad hoc methods of correcting for missing values for confounding variables. Am J Epidemiol 1991;134:895-907.
5. Horton NJ, Kleinman KP. Much ado about nothing: a comparison of missing data methods and software to fit incomplete data regression models. Am Stat 2007;61:79-90.
6. Carpenter JR, Kenward MG, Vansteelandt S. A comparison of multiple imputation and doubly robust estimation for analyses with missing data. J Royal Stat Society Series A 2006;169:571-84.
7. Schafer JL. Analysis of incomplete multivariate data. Boca Raton (FL): Chapman and Hall; 1997.
8. Krieger N. Why epidemiologists cannot afford to ignore poverty. Epidemiology 2007;18:658-63.

KIM, EGERTER, AND BRAVEMAN RESPOND

As Chen et al. note, we are in agreement that health researchers should pay attention to the potential for strong social patterning of missing data in epidemiologic studies. We believe that the descriptive findings presented in our recent article provide additional evidence of the importance of considering how survey respondents with missing income data may differ systematically from those with known income information.[1] Based on our examination of data from a large population-based postpartum survey in California, we concluded that excluding respondents with missing income information from analyses would bias study findings—posing a potentially serious problem in public health research that informs resource allocations and policy decisions.