SCHENKER, N., TREIMAN, D. J. and WEIDMAN, L. (1993). Analyses of public use decennial census data with multiply-imputed industry and occupation codes. *J. Roy. Statist. Soc. Ser. C* **42** 545–556.

SCHENKER, N. and WELSH, A. H. (1988). Asymptotic results for multiple imputation. *Ann. Statist.* **16** 1550–1566.

STIGLER, S. M. (1986). *The History of Statistics: The Measurement of Uncertainty before 1900.* Belknap, Cambridge, MA.

TANNER, M. A. and WONG, W. H. (1987). The calculation of posterior distributions by data augmentation (with discussion). *J. Amer. Statist. Assoc.* **82** 528–550.

TAYLOR, J. M. G., MUÑOZ, A., BASS, S. M., CHMIEL, J. S., KINGSLEY, L. A. and SAAB, A. J. (1990). Estimating the distribution of times from HIV seroconversion to AIDS using multiple imputation. *Statistics in Medicine* **9** 505–514.

TREIMAN, D. J., BIELBY, W. and CHENG, M. (1988). Evaluating a multiple imputation method for recalibrating 1970 U.S. Census detailed industry codes to the 1980 standard. *Sociological Methodology* **18** 309–345.

TSCHUPROW, A. A. (1923). On the mathematical expectation of the moments of frequency distributions in the case of correlated observations. *Metron* **2** 461–493, 646–680.

TU, X. M., MENG, X. L. and PAGANO, M. (1993a). The AIDS epidemic: estimating survival after AIDS diagnosis from surveillance data. *J. Amer. Statist. Assoc.* **88** 26–36.

TU, X. M., MENG, X. L. and PAGANO, M. (1993b). Survival differences and trends in patients with AIDS in the United States. *Journal of Acquired Immune Deficiency Syndromes* **6** 1150–1156.

WEI, G. C. G. and TANNER, M. A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *J. Amer. Statist. Assoc.* **85** 699–704.

WELD, L. H. (1987). Significance levels from public-use data with multiply-imputed industry codes. Ph.D. dissertation, Dept. Statistics, Harvard Univ.

ZASLAVSKY, A. M. (1989). Representing census undercount: a comparison of reweighting and multiple imputation methods. Ph.D. dissertation, Dept. Mathematics, MIT.

Comment

Robert E. Fay

Meng's paper usefully addresses one of the limitations of multiple imputation that I raised a few years ago. The author has introduced the term *congenial* to characterize a set of analyses for which the multiple imputation analysis is most appropriate and has discussed some of the implications of uncongenial analysis.

My own work on missing data has two primary objectives:

1. to identify and encourage analysis of the limitations of multiple imputation;
2. to develop better or more appropriate theory.

The papers I have written and those that I plan often attempt to address both objectives at once, although over time I anticipate a focus on the second goal. Meng's paper and Rubin (1995) serve the first purpose by acknowledging one of the difficulties that I pointed out.

Does Meng's complex argument lead us to a conclusion that, if multiple-imputation variances are inconsistent, consistent variance estimates are inappropriate? I do not think so. Subsequent analyses of the data, such as hierarchical Bayes models, meta-analysis and small-domain models, often depend on good variance estimates.

Robert E. Fay is Senior Mathematical Statistician, U.S. Bureau of the Census, Washington, D.C. 20233-40001. The views expressed are attributable to the author and do not necessarily reflect the views of the Census Bureau.

As I have attempted to indicate elsewhere, however, the problem addressed by the author is only one of the deficiencies of multiple imputation. Another arises in the context of complex samples, central to survey research generally and the Census Bureau specifically. Features of complex designs have effects on the validity of multiple imputation, generally of the opposite sort than addressed in the paper. In other words, the paper celebrates the finding that multiple imputation intervals are too long when the multiple imputation variance is inconsistent, but, in application to complex designs, many multiple imputation intervals are instead too short.

As an example of the current level of misunderstanding of the implications of complex design, in discussing their variance estimation for missing data in the 1990 Post Enumeration Survey (PES), Belin et al. (1993, page 1153) justify the omission of complex sample considerations from the highly clustered PES sample. Little's (1993) questioning of this argument did not shake the authors' conviction (Belin et al., 1993, page 1165). Yet simple Monte Carlo evaluation of the performance of multiple imputation shows the argument in Belin et al. (1993) to be wrong, except under special conditions not clearly stated nor validated by the authors.

I will continue to await a systematic treatment of the joint effect of uncongenial estimators and complex samples in the multiple imputation literature. (I will comment below on how these issues affect the analysis of public use data specifically.)