REJOINDER

Roderick J.A. Little, University of California, Los Angeles

I would like to thank the discussants who submitted written comments, and other discussants at the session for their interesting contributions.

Graham Kalton

Graham Kalton's discussion shares the lucidity and common sense of his other contributions in this area, and I agree with most of his remarks. With respect to his comments on the choice of conditioning, I agree that no single choice is appropriate for all analyses. I favor conditioning on n, n_R and y since a) I like to be as conditional as possible, ideally Bayesian, b) one can always become less conditional by averaging, but cannot go in the other direction and c) averaging as in Thomsen (1973) fails to capture differences between methods caused by imbalances in the distribution of n and n_R. This point is emphasized in a simpler setting in the important paper by Holt and Smith (1979), which I recommend to interested readers.

Graham states that unit nonresponse adjustments are directed at response propensity stratification, and item nonresponse adjustments are directed at predicted mean stratification, and provides an interesting rationale for this practice. It is not clear to me that practitioners necessarily have these theoretical objectives in mind, and my own uncertainty about the objectives of adjustment cell methods motivated the paper. I welcome Graham's clear statement, since it allows us to question whether this is in fact what we should be doing. In particular, if variables used for forming unit nonresponse adjustments are only weakly related to the survey variables, adjustment may be increasing variance, with negligible impact on nonresponse bias. Analyses such as those outlined in section 5 are directed at this question, but more work is needed on combining the propensity and prediction approaches to provide estimators with good mean squared error properties.

Susan Hinkins

The latter might not please Susan Hinkins, who notes that mean squared error is inappropriate for the highly skewed data encountered in IRS settings. I found a similar problem in a study of alternative methods for imputing wages and salary in the Current Population Survey (David, Little, Samuhel and Triest, 1985). Mean squared error was completely swamped by a few very large deviations, forcing us to adopt other measures, such as mean absolute error and mean absolute relative error. An obvious solution is to measure error on a transformed scale, such as the logarithm, but methods that work well on the transformed scale may not work well when converted to provide estimates on the raw scale (Rubin, 1983). I agree that there are problems worthy of attention here.

Susan's balanced discussion indicates the problems of implementing sound statistical theory in the setting of a large government agency. The view of cell mean imputation as a procedure requiring simplification is sobering to the theoretician, who might regard cell mean imputation as the naive starting point for some complex empirical Bayes modification! Susan and her colleagues should be commended for the fine applied work they have carried out in this environment.

Let me briefly comment on two specific points in Susan's discussion. She asks for the appropriate means of evaluating the results of alternative missing data methods. Simulations on artificially generated incomplete data are the stock solution to this thorny problem. However, simulations are limited by a priori knowledge of the statistician who sets them up. I am inclined to favor direct analysis of the incomplete data. Common missing data procedures usually involve certain modeling assumptions about the data; for example, imputing ratios of the form $\hat{y}_i = (\bar{y}_R/\bar{x}_R)x_i$, where $\bar{y}_R$ and x_R are respondent means, is optimal when y_i has mean βx_i and variance $\sigma^2 x_i$. Is this model realistic? In particular, are other observed variables predictive of y, and is the variance assumption reasonable? Such questions can be addressed using the respondent data. We need to assume that relationships found for respondents also apply to nonrespondents, but this assumption seems necessary for any evaluation, in the absence of information external to the data set.

Secondly, Susan requested my comments on nonresponse adjustments for panel surveys. Little and David (1983) presents preliminary work on extensions of propensity weighting to panel surveys with monotone missing data patterns, such as arise from attrition. Briefly, let r_j indicate response (r_j=1) or nonresponse (r_j=0) in wave j, for a survey with J waves. A sequence of propensity regressions $p(r_1=1)$, $p(r_2=1|r_1=1)$, $p(r_3=1|r_1=r_2=1)$, $p(r_J|r_1=\cdots=r_{J-1}=1)$ can be estimated, with predictors for r_j consisting of survey design variables and all survey variables collected on waves $1,2,\cdots j-1$. The crosssectional weight for wave j can then be computed as

$$w_j = [\hat{p}(r_1=1)\ \hat{p}(r_2=1|r_1=1)\ \cdots$$
$$\hat{p}(r_j=1|r_1=1\cdots r_{j-1}=1)]^{-1}\ ,$$

where the parentheses contain estimated response propensities from the regressions. Extensions of the response propensity method to non-monotone patterns appear less obvious.

David Chapman

Finally, I agree with David Chapman's remarks on substitution as a viable alternative to imputation. More theoretical work seems needed on appropriate methods for analyzing data that include substitutions.

Additional References

David, M., Little, R.J.A., Samuhel, M.E., and Triest, R.K. (1985). "Alternative methods for CPS income imputation," to appear in Journal of the American Statistical Association, 80.

Little, R.J.A. and David, M. (1983). "Weighting adjustments for nonresponse in panel surveys," working paper.

Rubin, D.B. (1983). "A case study of the robustness of Bayesian methods of inference: estimating the total of a finite population using transformations to normality," in Scientific Inference, Data Analysis, and Robustness, New York: Academic Press, 213-244.