DISCUSSION

John L. Czajka, Mathematica Policy Research, Inc.

In addition to dealing with nonresponse in the SIPP, all three of these papers represent continuations of research reported at last year's meetings. Kalton and Lepkowski, the second author on the final paper, combined then with another author in presenting a paper on adjustment for wave nonresponse. Apparently, this team has since split into two research teams. This being a session on SIPP, we have followed both researchers.

I will discuss these two papers first. The papers deal with different aspects of nonresponse adjustment for panel data. Kalton and Miller focus on wave nonresponse while Heeringa and Lepkowski concentrate on item nonresponse. Both papers examine a simple longitudinal imputation model: carryover or direct substitution of responses from other waves. In different settings they draw different conclusions about the performance of this model relative to an alternative. Kalton and Miller find the carryover model less effective than weighting when applied to wave nonresponse. Heeringa and Lepkowski find it more effective (when it can be used) than the current, cross-sectional hot deck in adjusting for item nonresponse.

By way of providing an appropriate context for this discussion, I think it useful to comment upon the differential needs for nonresponse adjustment and how these may bear upon the choice of method. Over the years the Census Bureau has refined the hot deck imputation method as used in the Current Population Survey (CPS) as an all-purpose, cross-sectional imputation tool for that particular survey. The method has held up well in test comparisons. One has to devote a lot of energy to developing an imputation model for a particular variable to do better. When cross-sectional imputations are used with panel data, however, the imputed records show more between-wave variation than those with complete responses. This presents a serious problem when the imputed data are used in longitudinal analyses.

The SIPP presents us with a dilemna. Getting the data to the users in a reasonably timely manner requires that the nonresponse adjustment (if any) be able to handle efficiently a very substantial array of missing values. Yet the richness of the data will attract enough investigators to almost insure that imperfections in the adjustments work their way into analytical findings. The need for a simple adjustment strategy provides the backdrop to Kalton and Miller's paper, while the improvement of imputation strategies is the subject of Heeringa and Lepkowski's work.

It is important to maintain the distinction between "mass production" uses and more specialized uses of nonresponse adjustments. For certain types of analyses, investigators will have to develop their own imputations; we cannot expect that the imputations provided with the SIPP public use files will be adequate for all purposes. This implies not only that better techniques are required to adequately deal with panel data, but that these tools need to be accessible to a broader community of researchers. In addition, the imputations on public use files need to be documented sufficiently to enable users to decide whether the mass produced imputations are adequate for their purposes. Both papers contribute to technique development and to assessing the performance of the adjustments.

As a prelude to their comparison of weighting and imputation, Kalton and Miller describe in detail their construction of a simulated data set, which will be used to evaluate alternative adjustment strategies. They also present data on the consistency over time of different types of items. This includes examination of month-to-month correlations between versus within waves, where they find significant differences. They observe that both types of correlations may reflect error. Their very careful analysis of the within wave responses during a period when we know independently that a benefit increase occurred is highly informative; surprisingly many households do not report a change during the wave.

The carryover method that Kalton and Miller test involves substituting the prior wave responses for a given missing wave. Such a model assumes that there occurs no change between waves. If there is change in a given case, the imputation will be in error. I must express some reservations about the tactic of testing an imputation model's assumptions by measuring the combined distribution of reported and imputed values against some standard rather than evaluating the no-change assumption directly. This procedure obscures error that may be of particular significance. This is not just a matter of the high response rate swamping the imputed responses. We conduct panel surveys to study change. It is critical to know whether our nonresponse adjustments adequately capture such change, even when change is infrequent.

In this light I find it difficult to view their carryover model as anything but a straw man. Even though I may ultimately accept the conclusion that weighting is the better method for wave nonresponse adjustment in the SIPP, I would be happier if the imputation method tested here included at least a simple depiction of change. I would ask the authors whether the addition of a simple change mechanism to the carryover model can be automated sufficiently to provide a realistic alternative. On another score, I found the analysis of the costs of weighting--namely, the loss of data--to be quite helpful in the assessment.

The Heeringa and Lepkowski paper addresses the use of data from other waves to improve imputations for item nonresponse. The authors outline several elaborations on the simple carryover model, and they provide a clear overview of alternative models with increasingly less restrictive assumptions about change between waves. The research results they report, however, extend only to a test of the carryover model. I suspect that they intended to have been farther along by now, and I anticipate future evaluations of models incorporating different representations of change.

A particularly valuable part of their analysis examines the extent to which items missing in one wave are indeed present in another. This is critical to our ability to take advantage of the panel nature of the data. Indeed, other wave responses are not always present, indicating a continuing need for some amount of strictly cross-sectional imputation. However, even on the items with the highest nonresponse rates, other wave information is present for at least 40 percent of the cases and generally considerably more.

Ideally, the comparison of imputation methods would be carried out on a data set constructed from real records with complete data for all observations. Nonresponse would be simulated by deleting individual data elements in accordance with the nonresponse patterns observed in the full data set. The artificially missing values would then be imputed by alternative methods, and the results compared with each other and with the true responses. As the authors relate, constructing such a data set for evaluating cross-wave imputations is a major undertaking, and replicating the hot deck imputation procedure is no simple task either.

What the authors chose to do instead was to compare the cross-sectionally imputed values provided on the SIPP public use files with their alternative longitudinal imputations, formed by direct substitution of values from another wave. Such a comparison provides evidence on the extent to which the two procedures yield different results. The comparison does not provide direct evidence on the relative accuracy of the two sets of imputations, but inferences based on indirect evidence may be possible.

The analysis presented in the paper focuses on a small set of job-related items, including both continuous and categorical variables. The two types of variables require different types of imputations, and they also exhibit sizable variation in response rates. The authors relate these differences quite nicely.

With respect to the categorical variables, which include occupation, industry, employer type, basis of pay and frequency of pay, the authors reason that cross-wave stability is high and proceed to evaluate the cross-sectional imputations against the assumption that the longitudinal imputations are correct. The logic behind this comparison is as follows. Cross-sectional imputations ignore longitudinal information. What if the missing values were in fact equal to earlier or later wave values; how would the cross-sectional imputations perform? Differences between the two series are interpretable as error in the cross-sectionally imputed values. Such a comparison seems reasonable as long as the results are negative--i.e., the two series are judged to be identical. What Heeringa and Lepkowski find, however, are substantial differences between the two series. In view of this, we must ask how reasonable is the assumption that the cross-wave imputations are correct. This is something that the authors could answer with their data set, and it puzzles me that they have not done so.

The comparison of the two sets of imputations of continuous variables takes a rather different form. Disagreement between the two imputations at the unit level is never calculated. Instead, a comparison is made, first, between distributions of imputed values and, second, among distributions of cross-wave changes, as measured between reported values and between reported values and imputed values.

The differences between the first distributions are not very substantial; nor would we expect them to be. In theory the cross-sectional imputations should reproduce the true cross-sectional distributions quite well. The longitudinal imputations should be upwardly biased in the first wave because substitutions were made from later waves, and downwardly biased in the third wave because the substitutions were made from earlier waves. The observed differences are not consistent with these biases, but with small samples of different records across waves this is not surprising.

The comparison of alternative estimates of change between waves addresses the major point of difference between the two imputation strategies. In theory the cross-sectional imputations should overestimate the amount of change between waves because imputation error will appear as between-wave change. It is not obvious to me that the mean change implied by the cross-sectional imputations should deviate from the true mean change, but the variance of this change should be considerably greater than the variance of estimated change based on reported values.

The longitudinal imputations, on the other hand, should underestimate change, and this should be reflected in both the mean and variance of the estimated changes. In fact, there would be no change at all if the comparisons always involved imputed values and the reported values from which they were imputed. The direction of comparison does not always match the direction of imputation, however (e.g., a reported value in wave 3 might be compared with an imputed value in wave 2 that was drawn from the reported value in wave 1; in such a case, change might be overestimated).

Another element complicating the comparisons is the possibility that a value may have been missing from only one month in a wave. I am not sure how the Census Bureau's imputation procedure handles this case (are the other monthly values in that wave used at all?), but the fact that the cross-wave comparisons use average monthly values implies that if any change occurred between waves, some change will be evident in the longitudinally imputed case even though the imputed month may show no change from the comparison wave.

The authors' discussion of their results would benefit from a more thorough review of these subtleties. Interpretation by the reader is made more difficult by inadequate familiarity with all of the details.

The results conform to the expectations with regard to the variances of the estimated changes. Estimated changes involving one cross-sectionally imputed value have somewhat more variance than changes based on reported values between waves 1 and 2. Between waves 2 and 3 the standard deviation of the changes involving imputed values is three times that of the changes involving strictly reported values. The standard deviations of changes based on longitudinally imputed values range from one-half to two-thirds those based on reported values.

The changes based on cross-section imputations show substantial, negative mean change between waves in comparison with the positive mean changes estimated from reported values. The authors highlight this finding without being able to explain it, but I question its generality and even its importance, unless it be indicative of a problem in implementing the hot deck imputation procedure on a data file considerably smaller than the CPS.

Turning now to the paper by Short and McArthur, I note that the objective of the paper is to examine the extent of sample attrition and its association with characteristics and events prior to attrition. The analysis focuses on determining whether there are any systematic patterns in the attrition. This addresses a very important concern in panel surveys--namely, that over time the sample loses its representativeness. In particular the fear is that the households most interesting to longitudinal analysis drop out, leaving us with a sample of, well, boring people. The analysis, which is principally descriptive, uses a linked file of data extracted from the core part of the questionnaire from the first five waves. The small size of the extracts, and the location of the extracted material on the full files, has made a five-wave link possible so soon.

While the paper is billed as descriptive, I would like to see more interpretation of the findings--particularly since the data raise a number of questions. My comments generally reflect this perspective.

With regard to attrition, what have we learned after five waves? Is there any evidence that the rate of attrition is diminishing? One of the charts presented by McMillen et al. suggested a nearly asymptotic curve, but is this reason to conclude that future attrition will be minimal? Do the patterns provide any confirmation or refutation of particular theories of attrition? For example, we might expect attrition related to the respondent burden to be highest at the outset, as those respondents only marginally willing to participate choose to leave. On the other hand, there is a cumulative burden; is there any evidence of attrition in response to this? Another type of attrition is related to life events which have higher incidence in certain population groups than others. We may speculate that such attrition would diminish over time, as the affected households leave the sample and are not replaced. What do the results say about this?

With regard to household characteristics, the authors report a number of significant associations with attrition. Are there any generalizations that follow? I note that stayers include disproportionate numbers of persons who tend to be in relatively stable residential circumstances. In addition to homeowners these include the elderly, persons related to other household members, persons married with spouse present, and persons with savings accounts (a relatively strong predictor of staying). This suggests a link between one set of personal characteristics and one particular reason for noninterview--namely, moving to an unknown address. Other links can be established as well. For example, certain demographic categories are at much greater risk than others with respect to the probability of leaving the SIPP universe. I would encourage the authors in their future work to investigate such links.

One aspect of survey attrition that is of particular importance is the cumulative impact after several waves. With attrition being higher among some types of persons than others, how do the sample characteristics at wave five compare with those at the outset? Is there any evidence that persons with particular combinations of characteristics have become substantially underrepresented? This question addresses a major concern about sample attrition in a panel survey. The answer may suggest possible revisions to the sample design as new panels are added.

The findings with regard to the relationship between attrition and the reason for noninterview provide information that could be useful in predicting nonresponse to future waves of the survey. For example, there is evidence that persons leaving the survey universe rarely return in the short run. On the other hand, those whose reason for noninterview was the inability of the interviewer to contact the household have a very high probability of returning. These two observations alone provide a basis for generating what ought to be a fairly good prediction of the wave six response rate from the wave five reasons for noninterview. This is just one application of the findings with obvious utility.

In the area of life events and attrition, where the authors do speculate at the outset about a causal link, there proves to be little evidence of a relationship. I would echo the authors' concern about the fact that a noninterview yields no data on the four prior months. Important life events occurring in this time period are thus lost from the data base, thereby reducing our ability to observe a relationship between attrition and changes in household circumstances during the preceding months. The problem is particularly acute with respect to movers.

As a final point, I want to raise the question of what implications these findings hold for the conduct and use of panel surveys in general and the SIPP in particular. I have already alluded to the possibility that groups experiencing heavy attrition might be oversampled in future panels. In view of the difficulty of devising efficient procedures to compensate for wave nonresponse, some new tactics for reducing nonresponse are more than welcome. I would encourage the authors in their future work to think about such implications of patterned attrition in framing their analyses and drawing their conclusions.