DISCUSSION

John L. Czajka, Mathematica Policy Research, Inc.

Longitudinal data open up many new areas of research, but as you have just heard, they also create new types of nonresponse problems. All three of these papers address nonresponse in the Survey of Income and Program Participation (SIPP). The first two papers present preliminary data on the incidence of nonresponse in the early waves of SIPP; the third discusses methods of adjustment for wave nonresponse. I will discuss the papers in the order in which they were delivered.

McMillen and Kasprzyk present estimates of item nonresponse for both recipiency and amounts from several income sources in the first two waves of SIPP. As they indicate, the results are in some sense uninteresting; nonresponse is very low. The nonresponse is lower than has been the experience with the Current Population Survey (CPS) for many of these same kinds of items, and lower even than was recorded in the Income Survey Development Program (ISDP) panel. I would ask the authors what they believe accounts for this success. What is the Census Bureau doing differently with SIPP that has produced these improvements over the CPS? Is there any evidence over the first two or three waves of a deterioration in the item response rates?

Tables 3 and 4 present distributions of the wave I wage and salary imputations by race, sex and proxy status. I feel these tables would be more informative if they conveyed instead the frequency of imputation relative to the total size of each subgroup.

McArthur and Short provide a descriptive analysis of wave nonresponse over the first three waves of SIPP. With three waves of data they are able to observe some instances of respondents missing an entire wave but returning the following wave and to contrast these with cases of what appear to be complete attrition. A general question that I would direct to the authors is how do you assess these results? Does the volume of wave nonresponse differ from what you might have expected? What are the implications for the quality of the SIPP data over the remaining waves? Do these results lead you to any predictions regarding future nonresponse?

A related issue is whether the observed patterns of attrition suggest any long-run problems for particular kinds of dynamic analyses. Some of the classes of persons on whom longitudinal surveys provide critical data may be among those with the highest rates of attrition: those who experience transitions in their marital status, in the composition of their households, in their residence, and in their geographic location. For example, does SIPP often lose whole households when there is a marital break-up? Do the data even allow an answer to this question? What data on transitions might SIPP be missing as a result of attrition?

In examining the reported distribution of reasons for nonresponse among groups with different combinations of waves present or absent, I computed the probability that a unit which missed the second wave would return for the third wave, conditional on the reason for nonresponse in the second wave. Among persons who refused the second wave interview, only 14 percent returned in the third wave. Among persons who moved to an unknown address, 20 percent were interviewed in the third wave. Among other persons who could not be contacted, 50 percent returned in the third wave. These results suggest some ability to forecast nonresponse in the next wave from the distribution of reasons for nonresponse in the current wave.

The presentation of the results in the remaining tables could be improved if the authors provided a better indication as to how strongly different characteristics are related to attrition. For example, they might report the chi-square values to which they make reference in the text. I note that marital status is one of the characteristics most strongly related to attrition. From table 2 we see that married, spouse present households represent 58.1 percent of the wave 1 universe but only 49.9 percent of the attritors (those with one or two missing waves). Never married persons represent 25.0 percent of the universe but 31.0 percent of the attritors, and separated persons are 2.3 percent of the universe but 3.7 percent of the attritors.

Table 2 would benefit from one additional column, showing attritors for all reasons other than refusal. I found myself often comparing the three wave respondents with both the total attritors and the attritors by refusal in order to determine the distribution of a characteristic for those who dropped out for reasons other than refusal. This group often appears quite different from the refusals. For example, on the aforementioned marital status, the refusals are quite similar to the three wave respondents, implying that the other attritors account for most of the reported differences between the three-wave respondents and all attritors.

In their further research, I would encourage the authors to look carefully at factors associated with attrition which suggest ways to intervene to reduce such attrition. For example, they mention the exploration of interviewer effects as offering prospects of improving response rates. Other characteristics may hold out such promise as well. For example, I note that there are quite substantial differences among the regional offices in their retention of respondents over the three waves. The New York office's share of attritors is nearly double its share of respondents. The Chicago office, on the other hand, had very low attrition. Chicagoans are noted for casting ballots from the grave; perhaps they do SIPP interviews that way as well.

I would also encourage the authors to develop multivariate models of attrition. This is particularly important in sorting out survey effects from respondent effects. For example, they note in the paper that rather than being associated with high attrition, the length of the interview has tended to be inversely related to attrition. By way of explanation they suggest that the persons likely to have brief interviews tend to be those with other characteristics associated with a high risk of attrition. It will be necessary to examine these characteristics jointly in order to determine whether there is any basis for drawing conclusions about the impact of interview length on the probability of attrition.

In the third paper, Kalton, Lepkowski and Lin do an excellent job of explaining the complex weighting and imputation issues raised by wave nonresponse and the possible approaches to adjustment. However, while Kalton et al. note that the objective of their study is to "provide evidence on the choice between weighting and imputation for handling wave nonresponse in the SIPP," they stop short of making recommendations between the two approaches. Perhaps they come out where I do: that weighting makes more sense on a number of grounds, but imputation is far more interesting to apply in this context. Nevertheless, I would encourage them to expand their discussion of the relative merits of the two approaches vis a vis SIPP, and I will raise some additional considerations in the course of my review of their work.

A clarification may be needed as to when adjustment for wave nonresponse is intended to be carried out for SIPP, and how this ties into the planned data products. The discussion in Kalton et al. centers around a 3-wave, annual file, with the wave nonresponse adjustments to be made once the three waves of data have been assembled. The authors point out that 3 waves yield 7 different patterns of response (excluding the 3-wave nonrespondents), such that if weighting were to be utilized as the method of adjusting for nonresponse, 7 sets of weights would be required to provide maximum utilization of data from the responding households for any combination of waves. They observe that the full 8 waves of the first SIPP panel will yield 255 possible patterns of response, with comparable demands upon the number of sets of weights required to allow every household to be included in a weighted sample for any combination in which it appears. With imputation one need not face this plethora of patterns. In fact, longitudinal imputation could be carried out as each wave is completed, thereby requiring imputation models for only simple attrition patterns. Moreover, such wave by wave imputation could in principle replace the current cross-sectional imputation for item nonresponse. The point remains, though, that the volume of weighting adjustments is a function of the number of waves included on the file, and this is therefore not irrelevant to the relative merits of weighting and imputation.

What differentiates weighting in the longitudinal context from weighting as it is applied in cross-sectional settings is the availability of at least one wave of data for units that were nonrespondents to any given wave. The choice between weighting and imputation centers around how best to use this information that goes far beyond what we have in the cross-sectional application, where we are often limited to the sample stratifiers. Kalton et al. detail efforts to use such additional information to differentiate between responders and nonresponders to waves in the earlier ISDP panel. Surprisingly, they found very little additional explanatory power from the prior wave data. This suggests that the additional data available in a longitudinal survey such as SIPP may not make much more than a marginal contribution to the longitudinal weights. The McArthur and Short findings on wave nonresponse to SIPP seemed to me to imply greater differentiation than Kalton et al. found for the ISDP, and I wonder if Kalton et al. were able to look at these same kinds of variables. If so, there could still be differences between the response patterns to the two surveys. Aggregate unit and item nonresponse rates have indeed been somewhat different between the two surveys. If subsequent analysis replicates the ISDP findings on SIPP, however, Kalton et al. point out that this is in fact a plus. Covariation between the survey responses and the probability of nonresponse is undesirable. Moreover, stronger relationships between wave nonresponse and prior wave variables would increase the variation among the weights, thereby reducing the precision of the survey estimates.

The discussion of longitudinal imputation strategies builds on the key difference between longitudinal and cross-sectional imputation: namely, the availability of prior measures of the missing characteristics for the same individuals. The starting point is the recognition that there is substantial cross-wave stability in many of the characteristics to be imputed, and Kalton et al. present evidence from the ISDP for selected characteristics.

Kalton et al. suggest as their most basic model one in which the current wave imputation is simply equated with the prior wave value. They add a stochastic component which could be drawn from the observed distribution of deviations of current from prior wave values for respondents to both waves. This model, $y_i = x_i + e_i$, is not a regression model, and it avoids an obvious disadvantage of a simple regression model in this setting: namely that regression imputed values will virtually all differ by small amounts from their prior wave levels. For characteristics with high cross-wave stability, the e_i in this alternative model will be zero in the great majority of cases. Further development of this model focuses upon improving the assignment of e_i. Kalton et al. suggest techniques for imputing this residual term.

While the cross-wave stability of many characteristics supports an approach such as this, we must keep in mind that if the world were all that stable, there would be little need for longitudinal surveys such as SIPP. Ideally, imputation models need to acknowledge the fact of much stability while incorporating plausible representations of how change occurs. Continued development of longitudinal imputation models can and should take advantage of what we already know about the dynamics of particular characteristics.

As an example, consider social security income. The two principal changes in the income flows are from "off" to "on" and the regular cost of living adjustments. Declines in benefits are not common; when they occur they are accompanied by other changes in household circumstances (generally with some lag). If we are imputing entire household records, we will often not observe such changes except when they occurred near the end of the prior wave. Therefore, an imputation of a reduction in benefits should occur only in the presence of the appropriate prior wave change or with a current wave imputation of the kind of change that could give rise to a benefit reduction. In the absence of these conditions there should be no imputation of a reduction in benefits. Similarly, the imputation of a start-up of benefits should be accompanied by evidence of eligibility in the prior wave or an imputation of eligibility in the current wave. Likewise, cost of living adjustments should be imputed only in the months in which such across the board increases were recorded by other survey households.

Earnings from employment provide another example. In the absence of a change in job or change in employer, a worker's earnings are often subject to only one change per year. A recent increase may actually reduce the likelihood of an increase in the current wave. This aspect of wage dynamics could be incorporated into a longitudinal imputation model.

It is easy to imagine carrying this line of development to great lengths, but I think we need to ask how pertinent this is to imputation in the public use files from SIPP, where there is a great volume of imputation to be accomplished within a fairly restricted time table. Sophisticated imputation models might best be left to researchers who plan extensive study of a particular kind of income flow and who may not be willing to use any outside party imputation. For the purposes of SIPP, there may be a lot to be said for simple but well documented imputations where that strategy is adopted over weighting. I stress the documentation. A user who contemplates use of the imputed values provided with the data set must be able to determine whether the imputed values are adequate for his or her purposes. Generally public use datasets are weak in this regard. It is encouraging that in the IDSP and SIPP efforts so much attention has been focused on methods of adjustment for nonresponse. I hope this is carried through to the documentation.

Having said that, I would like to conclude by discussing one complication of wave imputation in SIPP that Kalton et al. do not address. For many characteristics the desired end result is not a single value for the missing wave but four monthly values. The observed values for the four months may be equal most of the time, but they are not always so. Any proposed imputation procedure must include a strategy for generating monthly values.

To illustrate some of the issues and possible approaches, let me suggest four different models for translating a prior wave monthly pattern into a current wave imputation. Each of the four models is built around a plausible interpretation of a particular kind of monthly pattern. The first of these models, designated "replication," allows that there is unexplained variation in the monthly values, which may or may not follow a regular schedule. To capture this variation the model imposes upon the imputed monthly values the pattern observed in the prior wave. The next two models interpret the observed variation as measurement error and propose two different schemes for addressing this error in the imputations. The "error average" model specifies that the error be distributed across the four months by imputing a mean value each month. The "error edit" model specifies that deviant prior wave values be excluded in calculating a uniform value to be imputed to each current wave month. A fourth model interprets the observed variation as indicative of a particular kind of progression in the monthly values. This "ratchet" model bases the current wave imputation upon the last of the prior wave monthly values. These four models are not intended to be exhaustive. Rather, my point is to show, first, that there are a number of plausible ways to interpret within-wave variation in monthly values and to express this variation in the imputed values and, second, that different observed patterns may invite alternative interpretations with quite distinct implications for imputation.

Consider the following distribution of reported hourly earnings over a four month period:

10.00	10.00	10.00	10.80

The four models would suggest imputation of the following patterns.

(1)	10.00	10.00	10.00	10.80
(2)	10.20	10.20	10.20	10.20
(3)	10.00	10.00	10.00	10.00
(4)	10.80	10.80	10.80	10.80

The fact that the variable is hourly earnings and that the change occurs in the final month gives an edge to the ratchet model, which interprets the change as a raise. Such a model might be criticized as giving excessive weight to the final month, but it builds on what we know about the dynamics of wages. (Of course, we would feel more secure with this imputation if the 10.80 value occurred in both the third and fourth months.)

What if the 10.80 value occurred only in the third month? In that case the ratchet model would provide the least plausible interpretation (its imputations would coincide with those of the error edit model, however). The 10.80 value in the third month could reflect legitimately higher hourly earnings due to overtime pay (this would be more plausible for some occupations than others), or it might be a fluctuation resulting from the way in which earnings are measured (for example, a fixed month salary divided by a varying number of hours). In this view the pattern would best support models one or two.