

DISCUSSION

Ralph E. Folsom, Research Triangle Institute
P. O. Box 12194, Research Triangle Park, NC 27709

I will begin my comments by apologizing to the authors for the necessary brevity of my remarks. The likely effect of this brevity is to lend an overly critical tone to my discussion. While I have tended to emphasize areas where improvements could be made, I want to assure the session contributors that my overall view of their papers is very positive.

Turning to specific comments, I will discuss the papers in the order presented. The first paper in this session on nonsampling errors in the Survey of Income and Program Participation (SIPP) was presented by Dan Kasprzyk. The Kalton, Lepkowski, and Kasprzyk paper describes weight adjustment procedures used to cope with missing wave data in the 1984 SIPP Panel File. It is useful to note that the reported weight adjustment methods were initially viewed as an interim solution designed to speed up the production of a longitudinal research file. Observing that the method used, namely reweighting the eight wave complete data cases, discards 24 percent of the eligible wave 1 participants, many of us would look for an imputation solution hoping to save more of the partial data cases.

On the other hand, the simple reweighting of complete data cases has some appealing properties, namely

- Reweighting complete data cases avoids the creation of patched together artificial data records. Across wave associations are therefore preserved.
- The development of a general purpose longitudinal imputation strategy for wave nonresponse that preserves associations for all important variables and subpopulations is a very difficult if not impossible task.
- One possible extension of the weighting procedure that would preserve more of the partial data cases involves the creation of weights for other sets of wave combinations. For example, to maximize the use of partial data cases in the analysis of between wave transitions, one could produce weights for all persons with data for pairs of adjacent waves. This solution has the down side of fostering user confusion over which weight to use when.

Finally, I want to encourage the authors use of analytical methods like Search and logistic regression to help identify good weighting class variables. I also applaud their proposed direct use of logistic regression based response probability predictors as nonresponse adjustment factors. Judicious use of such response probability predictors has the potential to improve the stability of the resulting statistics with little or no loss in bias reduction.

The Corder and Manton paper reports a demonstration of Grade of Membership (GOM) model fitting applied to the aged subsample drawn from the 1984 SIPP panel. Fitting such a model in this context is an impressive computational accomplishment considering the massive numbers of parameter estimates that are required for a

data set as large as the 65 and older SIPP sample (6,355 persons and 57,000 person-wave records). I was particularly intrigued by the composition of the nine latent groups or pure types defined in terms of the first set of 36 variables. The two pure types (VI and VII) that have 100% proxy reporting are particularly interesting. All other pure types had zero percent proxy responses. Both types VI and VII were 100 percent female. Type VI was 80 percent Asian and 20 percent White while Type VII was 100 percent White. It is curious that no blacks were included in these proxy reporting pure types. Type VI was composed exclusively of females with less than a seventh grade education, with 55 percent aged 65 through 69, and 45 percent aged 80 through 84. The type VII females were 60 percent aged 70 through 74 and 40 percent aged 75 through 79.

Aside from the curiously restrictive nature of the proxy reporting pure types, I had a theoretical question regarding the parameter estimation. For this event history application, the assumption that an individual's responses are independent across questions is clearly unrealistic. While I acknowledge that consistent or asymptotically unbiased parameter estimates can typically be produced employing an estimation algorithm based on this unrealistic independence assumption, I was interested in knowing if there had been any work on GOM model estimation methods that make explicit use of the correlations among individual response. Finally, I want to encourage the authors to pursue their plans for follow-up work regarding the effect of proxy reporting on the accuracy of Medicare and Medicaid reports. I look forward to their demonstration of GOM model utility in this critical area of research.

The third paper presented in this session dealt with an analysis of time-in-sample bias in SIPP. The analysis performed by Chakrabarty and Williams suggests that SIPP time-in-sample effects are typically nonsignificant. The authors acknowledge that their estimates of adjacent time-in-sample differentials are potentially confounded with length-of-recall effects. They argue that SIPP length-of-recall effects may be small noting that significantly fewer transitions are reported within waves than between. While this rationale is certainly plausible, I would encourage the authors to consider fitting a general linear model to their data so that separate effects of time-in-sample and length-of-recall can be estimated.

I was pleased to see that Chakrabarty and Williams' statistical tests were based on direct variance calculations. I wondered, in this regard, whether the covariance between the two estimates in the difference statistic was estimated. For my final comment, I encourage the authors to pursue this analysis further when the availability of more panels make it possible to compare wider time-in-sample differentials.

Turning to the Vick and Weidman paper concerning the effect of self and proxy response status on the reporting of income reciprocity, I

want to complement the authors for their excellent presentation. I particularly liked the way they set out four specific research questions and proceeded to answer their questions with a clearly focused analysis. I also appreciated their clearly stated rationale for limiting inference to the observed sample and therefore using unweighted normal theory or Gaussian test statistics. Since all these tests tended to be statistically significant, it would be interesting to see how the analogous population level tests based on the probability sampling distribution would have performed. Finally, I would recommend that the authors consider a logistic regression analysis of their data. This is clearly a situation where proxy versus self-response effects are confounded with other person characteristics like age and other survey operations effects. Since this type of observational analysis is always subject to potential selection biases, a regression based covariance analysis is appropriate.

The Gbur and Petroni paper documents a preliminary evaluation of maximum telephone interviewing for SIPP. Considering the length and complexity of the SIPP interview, one could legitimately anticipate an increase in nonresponse and a deterioration in data quality associated with telephone interviewing. Surprisingly, the Gbur and Petroni analysis shows very little difference between the response performance and the reciprocity rates observed for the maximum telephone sample and the personal interview sample. The household, person, and item nonresponse rates for the two groups were all very similar. While there were

some reductions in reciprocity rates and rates of low income households in the maximum telephone group, the magnitudes of these reductions were not generally dramatic. I do have one slight bone to pick with the authors. The significance levels they quote are based on variance approximations derived from generalized variance charts. For their future analyses, I would strongly recommend direct variance calculations.

The final paper in this session was Dr. Haber's report of findings from the SIPP fringe benefits feasibility study. The results of this study are clearly of the bad news and good news variety. The bad news was the dismally low collection rate for signed waivers to contact employers (42 percent). As the author points out, for such a survey of employers to be feasible, follow-up procedures would have to be developed to substantially increase the waiver signing rate. The good news was the 96 percent response rate of employers given the waiver is collected. My concluding remark relates to a comment Dr. Haber makes in the footnote section of his paper. He comments that in his review of the survey methodology literature, he sees very little attention being given to the characteristics of persons who refuse to participate or who refuse to answer particular questions. While I do not deny that Dr. Haber's assessment is generally correct, I would point out that the great majority of surveys are one-time cross-sectional events, and short of conducting a nonrespondent followup armed with compelling monetary incentives, one is typically left with very little person specific information regarding nonrespondent characteristics.