Joseph Waksberg, Westat, Inc.

There is an interesting contrast between an implicit assumption in the paper by Mr. Bailey and in the one prepared by Messrs. Presser and Schuman. Mr. Bailey starts off with the assumption that there is a "true" value for each respondent of the statistic being studied and that deviations from this response represent "errors" in reporting. There is no such notion in the Presser-Schuman paper which instead talks about the "illusion of absolute proportions for or against specific social objects or positions." The difference in approach, of course, at least partially arises from differences in the subjects studied by the two authors. Mr. Bailey is primarily concerned with items generally thought of as hard data -- age, educational attainment, employment status, etc. where there is a presumption of true value. The other paper discusses opinions and the notion of an abstract true value is not necessary.

In practice, the line of division between Census-type inquiries and attitude studies is not as sharp as this discussion would imply. Some Census inquiries have elements which veer toward attitude studies, for example unemployment,or mother tongue. I suspect that there are subjects for which opinions are so strong that it would be possible to conceive of a true value, although I would be hard pressed to provide an example.

I'd like to discuss the implication of the existence of a true value on the Hansen, Hurwitz, Bershad response error model which is basically the one used by Mr. Bailey. Ignoring for the time being the issue of bias, the model decomposes the total pattern of responses into response variance (both simple and correlated), sampling variance, and covariance between the two. For simplicity, let me restrict the discussion to simple response variance which can be measured by repeated observations on the same sample unit (assuming independence in the observations.) By now there have been emperical measurements of response variance for a variety of items, by a number of statistical organizations led by the Census Bureau. The ratio of response to sampling variance varies widely. In the examples given by Mr. Bailey, it goes from a low of .01 to a high of .61. Other studies show even higher values of this ratio. It should be noted that when the ratio is equal to 1.0, we have the curious situation that we can learn as much by interviewing the same person n times, as by interviewing n different persons.

What inference can we make about survey results or survey methodology when we are faced with high response variance? Mr. Bailey, and some of the published Census reports imply that when this happens the quality of the data are suspect, and the methodology needs improvement. This is certainly an appropriate statement for some types of statistics. I think it may be an oversimplification to generalize it to all situations.

It seems to me that three views are possible of high levels of response variance. One is the inference just made, that the instrument used to collect data is unreliable and the results are uncertain. A second is that response variance is simply a component of total variance, and like sampling variance, it is a nuisance that can be overcome by taking a large enough sample size. A third view is that response variance provides some analytic information on the subject being studied. This is particularly appropriate in attitude studies where it is not surprising to find inconsistencies in response. The level of response variance can be thought of as an index of how strongly and therefore consistently people feel about their responses, with response variance indicating the degree of uncertainty.

There is a parallel between this interpretation of response variance and intraclass correlation. Response variance can be thought of in the same way as intraclass correlation. It reflects a within-sampling unit variance, postulating a distribution of responses within each unit in the population. From the point of view of sampling theory, a large intraclass correlation is a nuisance, requiring an increase in the sample size for a given degree of reliability. From the point of view of the social scientist, a large intraclass correlation tells something about how people live, or other information on the population being studied. The response variance can be similarly interpreted.

Obviously, it is difficult to make this interpretation when a "true value" exists, for example in measuring age, educational attainment, income, etc. and in these circumstances a high level of response variance does imply inadequacy in the measuring instruments being used. When this occurs, there is certainly a likelihood that biases also occur and there is justification for being skeptical that a large sample size would compensate for errors in measurement.

In attitude surveys the other two views are more appropriate. I am aware of only a limited number of attitude studies in which reinterviews with the same questions have been performed, but in those the level of response error has generally been high. I suppose that some of this could be caused by the wrong questions being asked, or poorly trained interviewers. However, I suspect that more often, it indicates that individuals do not have firm and fixed positions on the questions being asked. This probably at least partially explains the reason why apparently minor changes in question wording can have large effects on the marginals. I think it would be interesting to combine reinterviews, or a randomized interviewer assignment experiment with the kind of study reported by Mr. Presser. I would expect that there would be high levels of response variance for the kinds of items that show large differences with minor changes in question wording.

The distinction between hard data and attitude questions is not always very clear. If one looks at the Census data, the high values of response variance appear for such items as unemployment and mother tongue. There appear to be some segments of the population for which these items veer away from facts and towards attitudinal questions. The answers to questions on these subjects sometimes depend on an individual's perception of himself, and these perceptions are not necessarily consistent. I doubt that the Census Bureau, or other social researchers will always be effective in refining question wording or improving interview training to significantly reduce response variance. Obviously, they have the responsibility to attempt this, and often they will succeed. However, in many cases, the real issue is not poor question wording or failure to understand the intent, but honest uncertainty about the reply.

Even for so-called hard data, this is not necessarily a reason to reject the data or even for undue concern. Many examples exist of useful statistics that are subject to relatively high response error. Unemployment statistics in CPS have almost always showed high levels of response error, but when looked at as a time series it has shown remarkable stability and there has been very little reason to doubt its intrinsic validity. There is another, rather curious example in the Census records. In the 1950 and 1960 Census, each housing unit was classified as to whether or not it was "dilapidated." The classification of "dilapidated" was one of the most inconsistently reported items ever studied in the Census. A unit classified as dilapidated by one interviewer was infrequently classified the same way by another. However, when counts of dilapidated units were prepared for areas such as blocks or Census tracts, there turned out to be a very high correlation between the two classifications. Even though the classification for a single unit was subject to a high response variance, this was not the case for statistics for blocks or tracts.

In social science research, it is different to make sweeping generalizations. Sometimes response variance can be thought of as a measure of error. In other cases, it would be more appropriate to consider it another tool of analysis of social phenomenon.