

DISCUSSION

Brian Greenberg, U.S. Bureau of the Census

The speakers are to be commended for their paper, "Imputing for Missing Survey Responses." It is broad in scope and usefully addresses a variety of item imputation methods.

In my discussion today, I would like to touch on a few issues that might be called "imputation strategy." In particular, I would like to discuss employing information embedded in editing criteria to aid in determining a value for an imputation. As noted in the presentation, item imputation is used either because of item nonresponse or because a reported value was deleted due to edit failures. In addition, one demands that a system imputed value pass all editing criteria.

When imputing for a deletion due to edit failures, one should endeavor (when possible) to utilize the reported value, although, of course it is incorrect. To the extent that there may be patterns in erroneously reported values, these may be employed to obtain a correct value, i.e., what the respondent intended to convey had there been no error. As a simple example, suppose some item was to be reported in tons, and based on other response items a valid response is expected to lie between 5 and 15. Suppose further that 14,000 was the reported value. The value of 14,000 will be targeted for adjustment and an alternate value will be imputed in its place. A reasonable impute is 7, the underlying assumption being that the respondent reported in pounds rather than tons. (It might be noted that an imputation based on the assumption that data are missing at random is unwarranted here.) Although this example may appear trivial, more complicated scenarios do exist where this approach can be fruitfully employed. If we do not obtain a valid value when attempting to adjust in the manner suggested above, we can then proceed to a more general technique suitable for item nonresponse.

As noted above, another connection between editing and item imputation is the requirement that the imputed value pass all edit constraints. One may be employing an imputation procedure that is quite good and yields a suitable imputation in the vast majority of cases. The question must arise, however: how does one recognize the records for which the imputed value is not appropriate? And when recognized, what does one do about it? An imputation can be inappropriate because the record is significantly atypical or because the imputation procedure contains a stochastic component that threw the impute out of bounds.

A simulation study was discussed today involving Income Survey Development Program data. Quarterly earnings were imputed using a regression model containing a stochastic component. Due (in part) to values for the stochastic component, in some cases a negative value would have been imputed for earnings. This, of course, was not acceptable, and such a value could be thought of as failing the non-negativity edit restriction. As a rule, an inappropriate (edit-failing) impute is not quite so glaring, but comparable considerations apply.

One can address the question of determining whether an imputation is appropriate for a particular record by asking the question "Does the imputed value pass the relevant edit constraints?" That is, does the imputed value conform to the strictures which are

imposed on respondent provided values?

In the processing of surveys, one has to consider imputation options, or better yet, a hierarchy of options. That is, suppose a technique to impute for some field will not yield an appropriate value. For instance, a necessary ancillary variable may be missing, or a suitable "match" cannot be found in the selected imputation cell. In complex surveys one has to anticipate these problems and provide for alternate actions. Basically, one has to try something else—a secondary or even tertiary option which one hopes will eventually provide a suitable impute.

Let us return now to the question of what to do if an imputed value fails some edit constraints. In such a case, these secondary options can be brought to bear. If the first choice of an impute is inappropriate (i.e., fails edit criteria) one can pass to a second or, if necessary, a third option. If none of the usual imputation procedures provides an acceptable impute, the record under consideration should be sent for analyst review.

Another approach in dealing with an edit-failing impute is to utilize the invalid candidate for imputation to obtain a suitable value. Reconsider the example above in which some imputes for quarterly earnings came out negative. If the negative values were adjusted so that they fell into the lowest earnings category, the distribution of quarterly earnings would have resembled, reasonably well, the distribution of the actual values on which the simulation study was based. Such an adjustment to an edit-failing impute corresponds to adjusting reported invalid responses in such a way that the revised value is based on the reported value. As a final case, when one imputes because a field was deleted due to edit failures, a third choice exists for dealing with edit-failing imputes. This was discussed by Gordon Sande earlier and consists of returning to the original record and selecting (when appropriate) an alternate field (or set of fields) for deletion and imputation.

The last issue that might be discussed involves the role of subject-matter specialists in the development of edit and imputation procedures. Their role is rarely mentioned in discussions of edit and imputation techniques, yet their contributions can be incorporated in almost every phase of the process. In the case of repeated surveys, it is frequently true that subject-matter analysts have examined copious respondent forms and have personally contacted numerous reporters.

Subject-matter specialists can often aid in the recognition of systematic response error and suggest how to use a reporter's erroneous response to derive an accurate value (as in the example I gave at the beginning). In some situations, they can recognize when selected instances of item non-response signal the respondent's intention to indicate a zero. In the paper presented today, deductive imputation (perhaps a misnomer) was discussed—a technique that draws upon relationships between variables in determining a value to be imputed. Here again subject-matter specialists can contribute to the understanding of such factors and to their incorporation in the imputation process.

Subject-matter specialists can be, and should be, valuable members of a team whose goal is to design,

revise or update an edit and imputation process. It would be highly desirable that their role be discussed more, channels for incorporating their expertise be explored further, and their potential contributions be evaluated.

In summary, let me repeat that the Kalton-Kasprzyk paper is worthy of careful study. It touches on many crucial issues, presents a battery of techniques, and contains an extensive list of references. This work will be an asset for those evaluating and reviewing their imputation procedures and those in a planning stage.