

DISCUSSION

David L. Hubble, U.S. Bureau of the Census
Demographic Statistical Methods Division, Washington, DC 20233

Prior to reading these papers, I was quite illiterate in the subject of these NCES surveys, their associated components, and the non-sampling error issues addressed in these papers. Now, unfortunately, I dream of their issues and problems and how they might find solutions. I commend (or is it condemn) the authors for doing such a fine job of drawing me into their world.

As with a child, please forgive the silly questions I may ask, as I feel my way through this new maze, but with the same gentle manner please straighten me out if I stray off course and am missing the point.

I will discuss each paper in the order in which they were presented.

"Comparisons Across and Within NCES Surveys" by Salvucci, et al.

Unfortunately, Salvucci is dealing with one of the messiest problems in statistics. That is, collect the same information by two different methods and then try to reconcile the differences that arise in your basic count statistics.

I know of only one clean solution to this kind of problem and that is to ignore one of the data sources. Foregoing this rather simplistic solution, Salvucci attempts to identify the various differences between the Core of Common Data (CCD) surveys and the Schools and Staffing Surveys (SASS). If I understand it correctly, the CCD surveys are actually censuses and serve as the sample frame for the SASS.

After establishing the existence of many significant differences in the count of students, teachers, schools and local education agencies between the CCD and SASS data, Salvucci provides a good discussion enumerating the coverage differences between the CCD and SASS surveys that may explain the observed differences in the basic count statistics.

However, while the reconciliation process explained away the differences for many states, several states still had significant differences. And in fact, some states showed significant differences only after adjustments were made.

Personally, I felt like many of the differences were still unexplained.

Some questions come to mind:

- (1) Do the same people respond to the CCD and the SASS? I wasn't sure.
- (2) Are the CCD and SASS data collected at the same time of the year?

In the final section and in the conclusion, an attempt

is made to assess who's right and who's wrong between the CCD and SASS. SASS was assessed to be the major source of the differences.

While there may be merit in such work, I have some reservations with the method. The method compared the SASS to CCD percentage difference with CCD year-to-year percentage differences. If the SASS to CCD percentages was larger than the CCD year-to-year percentages then the discrepancy was attributed to the SASS estimate.

However, my guess is that the reporting error for CCD estimates by state are highly correlated from year to year, while the correlations between SASS and CCD reporting errors are not nearly as high.

The problem scenario I envision is this: the same CCD reporting error is made year after year while the SASS figure is virtually error free. However, under the decision criteria stated earlier, SASS would be labeled as the source of the discrepancy because the SASS to CCD difference would almost always be greater than the CCD year-to-year differences.

Whether SASS is the major source of the differences, or not, I don't know, but I do believe this assessment technique needs to be rethought. Possibly something along these lines could be done once SASS 93-94 data is available.

But beyond this issue of blame is a larger question. Does NCES want these differences in estimates to disappear or at least be substantially reduced in the future? If so, then plans for modifying the CCD or SASS to bring them in line need to be developed. And while being able to accomplish this completely is extremely difficult, if not impossible, at least those differences that are little more than arbitrary should be removed.

Also, I wonder if some of the issues presented and discussed at an earlier session apply to possibly bringing these CCD and SASS estimates into agreement.

"Documentation of Nonresponse Across NCES Surveys" by Saba, et al.

Saba has tackled the rather daunting task of reviewing the detailed documentation of 13 different NCES surveys.

Saba has amassed a great deal of information on how each of these surveys address 4 issues:

- The calculation of unit response rates.
- The calculation of item response rates.

- Methodological and analytical issues of addressing nonresponse in estimation, and
- Categorization of demographic variables.

Saba's work was quite comprehensive. The only possible omission was that the discussion of methods on how to deal with nonresponse only included unit nonresponse. There was no mention of how these surveys dealt with item nonresponse. Maybe the documentation was lacking, but still I was left wondering just the same.

However, even if this had been covered, my overall feeling after reading this paper would have been the same. And that is...

WHERE DO WE GO FROM HERE?

Saba's work clearly demonstrates that the methods are very different from survey to survey.

BUT WHAT HAPPENS NOW?

Certainly you want to avoid this document living a quiet life (or death) in a dozen or so filing cabinets. I have two thoughts here:

(1) Create a database with all this information. The paper stated that this information would be useful for users of the NCES data and especially to those making comparisons across surveys. Possibly this database could be updated and, therefore, not force others to read through tons of material in search of a few pieces of information.

(2) A second thought is to establish even greater uniformity between surveys in the future (a goal Saba pointed out in the paper). In terms of nonresponse, it may be helpful to establish guidelines for defining and classifying nonresponse for different types of surveys (such as mail, RDD, or personal interview). This would probably also lead to greater uniformity in methods used to adjust for nonresponse.

In terms of categorization of demographic variables it may be helpful to develop forms (or at least specific question and answer categories) that are used across several surveys. An example of this is the Census Bureau's attempt to create what is generally referred to as a "Uniform Control Card" for all its major household-based demographic surveys (CPS, NCVS, SIPP, CE). In doing so analysts can be fairly sure when comparing statistics on, let's say, race that the question was worded the same way and that the answer categories were presented in the same order.

I believe doing this would increase the utility of all the NCES surveys.

"Multivariate Modeling of Unit Nonresponse for SASS 1990-1991" by Gruber, et al.

Gruber's paper is another good example of using modeling techniques to gain a more complete understanding of nonresponse in surveys and, in this

case, unit nonresponse. The surveys included SASS's Public School Surveys and Private School Surveys.

Assessments of the available variables of state, urbanicity, school level, and school size for public school surveys and the variables of urbanicity, school level, school size, and affiliation for private school surveys have been made before through the modeling work of Shen, Parmer, and Tan in 1992.

The new twist in the paper is the combining of final models across the surveys that is school district, school and teacher surveys. This was done separately for public and private schools.

These combined models were referred to as "Cross Component Models". While the idea is intriguing, I believe the paper needs to provide more information in terms of the motivation for combining the models. It wasn't clear to me.

The conclusion from the paper was that fewer variables appeared to be significant in modeling unit nonresponse than from the Shen, Parmer, and Tan modeling research.

I wondered why this outcome. Does it relate to the fact that the "unit" is different in the various surveys with the cross component models? That is, the unit ranges from school administrator to school to teacher.

Or is it because states were collapsed into strata in this paper, while no state collapsing took place in the Shen, Parmer, and Tan research?

One final side note: while the collapsing of states may improve the model, I get a little nervous when one of the defined state strata consists only of Alaska and Massachusetts. Maybe regional constraints are needed.

"Evaluation of Imputation Methods for the CCD Finance Data" by Johnson, et al.

Johnson's paper deals with the problem that for certain reported aggregate amounts the corresponding components are not reported by many states. This was addressed both in terms of revenues and expenditures.

Specifically, the paper evaluated different approaches to impute the missing internal values. Before discussing the specific techniques used for imputation, I would like to raise a few issues.

An example in the paper describes a situation in which the components require imputation because one of the components is missing or zero. Having worked on a similar problem, I know we grappled with the proper treatment of reported zeros. I wonder did you always impute for a reported zero? Or were states ever contacted about missing values or reported zeros for clarification?

Also, out of curiosity, I wondered if states ever report values for some components that do not add to the reported total value? And, if so, how are these situations handled?

In the year investigated (91-92) only 14 states reported the desired detail of the revenue variable examined. One concern I have is; are the 14 states representative of the 36 who did not report the detailed information. Generally, it is problematic if correlations exist between whether a state reports the detailed distribution and the distribution itself. And when response rates are low the likelihood of this occurring increases. Specifically, of the 14 states there appears to be some geographic clustering.

This is less of a concern for the expenditure variable which had 38 states report the components.

As I was reading the paper I had another "just wondering" question. Is longitudinal imputation a possibility? Even for just some of the states needing imputation? If previous years distributions were available it would make sense to use them. I get the impression though, that the details of the reported finances are fairly constant from year to year.

In terms of the possible techniques evaluated in the paper, I also have a few comments.

One is that, the ability to interpret the plots is increased if the X- axis is always in terms of the sum that was used to compute the ratio on the Y-axis. This was not always done.

Of all the specific techniques compared, I agree that the NCES III Alternate Herriot is probably the best. Also, the number of states evaluated was quite small, 6 or 7 states at times.

However, one potential drawback to all these techniques is that the distributions for all the imputed states will be exactly the same and probably looking different than any of the reporting states. This is a problem if one needs to estimate variances associated with these distributions. Though, I am not sure if such needs arise.

Finally, I was interested in knowing more about why the regression method with its promising "highly significant linear relationship" failed in warranting further consideration.

CONCLUSION

As I said in the beginning, I have learned a lot about the NCES surveys and some of their current issues. The authors are commended in their fine work. My hope is that you may find some of my comments to be of some use. And finally, I appreciate the opportunity to have participated in this session.