

DISCUSSION

Joseph Waksberg, Westat, Inc.
1650 Research Blvd., Rockville, MD 20850

Bill Butz and Tom Plewes' brief historical summary of important developments in the 50 year old history of CPS provides a useful background for a session on its plans for the next decade. In the 50 years since 1940, which is generally treated as the start of CPS in its present form, there have been innumerable ASA sessions and other conferences devoted to discussions of various aspects of CPS methodology. These have been of great importance, not only to social scientists interested in labor force statistics, but more broadly to anyone concerned with the methodology of household and population surveys. The CPS is a cornerstone of both labor market information and post-censal data on a wide variety of demographic, social and economic statistics about the U.S. population and households. Furthermore, as the authors state, CPS has been considered the international standard for household surveys. Its methods and procedures have been adopted all over the world. As a result, any new developments in CPS and results of research are of interest to a very broad audience.

As a former participant in the development of CPS methodology and research and a current user of its data, I found the three reports useful and informative. The wealth and diversity of research in planning revisions is truly impressive. The introduction and enhancement of new technology such as CATI, CAPI, and possibly TIGER, and the modernization of the data processing should improve the efficiency of the system and permit better data analysis.

However, some of the decisions on plans for the redesign CPS appear to me to be based on rather shaky evidence. In addition, I have some concerns on the direction the research seems to be taking. I would like to concentrate on these aspects of the paper rather than review components which are likely to be successful. Also, I do not feel competent to discuss the need for new data content or definitional changes, and will not comment on them.

Bill Butz and Tom Plewes' paper updates the information on plans for the CPS reported by the same authors in the Census Bureau's 1989 Annual Research Conference. The papers by Jay Waite and colleagues and by Copeland and Rothgeb expand on some of the issues mentioned both in 1989 and the current paper by Butz and Plewes. Since I was a discussant at the 1989 Conference session on the CPS, it seems appropriate for me to start by updating my previous remarks.

I. Research Program

My first comment at the 1989 presentation was that I was concerned that the CPS research program described had a rather narrow focus. It only addressed problems that needed to be resolved for the redesign of the sampling and survey procedures planned for 1994. There was no mention of research on fundamental data quality issues that are not involved in the changes in procedures considered for 1994. The three papers presented today do not indicate that there has been any change in the approach to research. It is still viewed as a set of projects that are started in response to the need for revision to be made in the next redesign and will

be completed in time to meet that schedule. I don't believe this is an effective way to conduct research, which should be planned as a continual operation and budgeted accordingly. A major reason for the reputation of the CPS, perhaps the major reason, is the innovations it introduced into so many aspects of survey methodology. Most of these innovations resulted from a tradition of research embedded in CPS that assumed a certain amount of freedom in exploring key aspects of methodology. Obviously, there need to be priorities that address critical problems for which early answers are necessary. However, I don't think it's wise to concentrate exclusively on these problems or to assume that just because a time schedule has been established, the problem will be solved when the specific research project is completed. The specific research projects should be viewed in the context of a broader and continuing analysis of household survey methods.

It is not difficult to list CPS problems that have been of concern for a long time and that should be investigated. Recent changes in methodology proposed for CPS create the need for additional research. Brief discussions with several colleagues produced the following suggestions:

- Effect of proxy respondents, including the possibility of reducing proxy reports by having sample household members keep some kind of diary during the reference week.
- Methods of implementing dependent interviews; CATI and CAPI make dependent interviewing much more practical. There is ongoing research on dependent interviewing, but it appears to be restricted to work that can be completed before the redesign rather than taking a long term view.
- Rotation group bias, and more generally panel conditioning. This was studied intensively about 20 years ago with inconclusive results. Further work would be useful.
- Industry and occupation data, which has problems in both reporting and coding.
- Differences between CATI and non-CATI reporting. I understand that there are huge differences between CATI and non-CATI reporting on some items in the Crime Victimization Survey, running as high as 30 percent. It may take longer to analyze the effects in CPS than the redesign schedule permits.
- Effect of different imputation strategies on bias and variance.
- Techniques to encourage better coverage in CPS, which has been deteriorating steadily over the years. Undercoverage may be at point of having a serious impact on some of the statistics, particularly for some of the CPS supplements.
- The possibility of improving response rates for RDD surveys. The papers indicate that problems of interviewer recruitment and retention are serious impediments to the expansion of CPS for state data, and that these problems are the impetus to considering a two week recall for the expansion. Using RDD for the expansion would make it much easier to get state data,

but has been rejected because of high nonresponse rates.

- The sample design implications of planned changes in data collection procedures. In particular, the movement to more centralized telephone interviewing and employment of interviews for two weeks on CPS rather than one, could create a significant change in the cost function for data collection. The implication for sample design should be examined.

The above list is not meant to be exhaustive, but is illustrative of the research that I believe is needed. The Copeland-Rothgeb paper does describe ongoing research on some of these subjects, and the other two papers also briefly allude to research on some of these topics. However, the research is described as solely directed to the 1994 redesign. I think that Copeland, Rothgeb and the other talented researchers at the Bureaus of the Census and Labor Statistics should be encouraged to develop a much broader research program with long term rather than short term goals in mind.

Let me now move from these general remarks to specific items of the redesign plans.

II. Two-Week Recall

The most uncertain and controversial change in methodology is the proposal to use a two-week recall for the part of the sample to be used for state supplementation. It is not clear to me how firm these plans are. The Butz-Plewes paper seems to imply that the only problem is in the development of methods to implement a two-week recall method, with no fallback position given. The paper presented by Waite refers to the possibility of extending the recall period if tests indicate its feasibility, and states that if it does not work then data will be collected for the week following the CPS reference week for the supplementary sample used for state data.

I think there are serious problems both with the extended recall plan and the fallback position described by Jay Waite. I would be surprised if ways can be found to collect information based on extended recall without serious bias. Getting respondents to recall exact placement in time has not been particularly successful in most studies. The Census Bureau has reported on telescoping in a variety of surveys. What has been referred to as the seam problem in SIPP, which is the much greater frequency of changes in labor force status between two interview periods rather than within a period also reflects poor ability to recall the exact time an event occurred.

Given the uncertainty in whether two-week recall will work, the plans for research seem to me inadequate. First of all, the laboratory work is built on a rather shaky foundation. In order to compare errors in one week and two week recall a self-validation procedure will be used. Self-validation did seem to be effective in the study cited, but the sample sizes were extremely small and the occupations studied were highly restrictive. The laboratory work will be followed by a field test to measure whether important differences exist. This is scheduled for May 1994 through August 1995. The final decision and preparatory work for the State expansion are also planned for 1995. This kind of tight timing and lack of work on alternatives puts great

psychological pressure on staff to find positive results and rationalize why the experiment was successful.

The sample to be used for the field test is described as RDD supplemented by a list of nontelephone households obtained from some outside source. The Bureau of the the Census has previously reported very high nonresponse rates on attempts to use RDD, and I think the mixture of the two frames is a rather inadequate simulation of a CPS sample. Considering the crucial nature of the decision, it seems to be a rather cavalier approach to sample selection.

There are bound to be some differences between one and two week recall data. The best one can hope for is that the differences will be small. Each state's statistics will combine data for the part of the state sample included in the national sample and the part in the state supplement, which will be the two week recall. Since the 50 states will have variable fractions of their total sample in the supplement, the biases from two week recall will also vary among the states. I suppose no one knows what the major uses of the state statistics will be 8 or 10 years in the future but if they are involved in the allocation of federal money to states and localities, possible biases, especially ones having different effects in the 50 states, will be carefully examined. New Jersey sued about 10 years ago because they did not like the unemployment data for the state. It's not hard to visualize this happening on a much larger scale, if the allocation of billions of dollars are at stake.

The fallback positions described by Jay Waite, to collect data for the week following the CPS reference week for the supplemental sample, has some of the same problems as use of two-week recall. The mix of national sample and supplement will vary among the states, and the states will thus unevenly reflect the different labor force characteristics of the two weeks. Months in which such events as school openings or closing, major holidays, or unusual weather occur will be particularly affected.

I am dubious about any plan that treats the national sample and the supplement differently. I would urge consideration of an alternative which spreads the sample consistently over the reference weeks. If multiple reference weeks are chosen, I think four or five weeks should be used rather than two or three. Reflecting the full month in the CPS has a conceptual and economic meaning; use of a two or three week period seems like an arbitrary compromise between a week and a month. Moving to a different period of time for the labor force measurements would create some problems in seasonal adjustments during a transition period. I suspect these problems would not be much greater than those faced by the current CPS where normal seasonal factors are affected by such features as which week Easter occurs, or exactly when most schools close in May or June and open in the fall. It would be useful to study such effects.

I must say I don't understand the basic need to move away from the one week reference period. The two papers that discuss this say that the expansion is well beyond a size that can be economically handled in a one-week collection period. I assume that the term "economically handled" implies that it is possible to do it but it would cost less per household to give interviewers larger workloads rather than hire additional interviews. This is probably true of the current CPS but there doesn't seem to be any drive to change the reference period for national statistics, so why is it being done for state data.

Furthermore, I'm not sure that using a second week for the state supplement would significantly lower the need to add interviewer staff. I assume the expansion for the state supplements would largely be accomplished by adding PSU's in small states that now have only one or two PSU's. Won't new interviewers be needed in these PSU's, and if so does it make any difference in which week they work? If there is a serious problem in recruiting and retaining interviewers, then another alternative is to contract out some of the data collection. I know of at least one company that would be interested, and would perform commendably.

III. Other Issues

Let me move on to brief comments on other subjects. In discussing sample replenishment, the Butz-Plewes paper states that this time, rather than merely reselecting a sample based on the 1980 design, improvements will be made to reflect the new population distribution and new sample design developments. The implication seems to be that this goes beyond what was normally done in the redesigns after each decennial census. The redesigns have always reflected the changed population distributions and new sample design developments, including such major modifications in sampling methods as introducing a new rotation system, replacing area samples by list samples, changes in cluster sizes, and improvements in estimation methods. So there is really nothing new about the approach this time. Incidentally, the papers did not discuss the new sample design developments. I hope these are described in a forthcoming ASA meeting.

The Butz-Plewes paper also refers to the still-high response rates that CPS is able to achieve. The term "still-high" probably understates the situation. The response rates are virtually at the same level they were 30 or 40 years, a remarkable achievement in light of the increased difficulty of contacting the U.S. population that Butz-Plewes describe. However, it should be noted that coverage, on the other hand, has been deteriorating slowly but steadily over the years. It is now at the point at which it could be introducing significant biases to the measurement of unemployment rates, particularly for blacks and Hispanics. One of the greatest challenges for CPS in the coming decade is to attempt to find ways of reducing this undercoverage and I regret that there does not appear to be ongoing research on this.

The planned introduction of new technology such as CATI, CAPI and information electronically paper should improve the overall efficiency of the system. As pointed out by Butz-Plewes, other organizations have moved faster and more aggressively in this area and it is time for the Census Bureau to follow. I don't think anyone should expect these developments to reduce CPS costs. Their main advantages are in improving quality and flexibility in changing questionnaires. However, there should be a fairly

comprehensive testing before these devices are introduced. As I mentioned earlier, in at least one survey there is a surprisingly large difference between CATI and non-CATI reporting.

Also in regard to increased use of technology, I was interested in hearing that industry and occupation coding will be automated. Studies of computer coding have been going on for 15 or 20 years, and it's nice to hear that it seems to be working satisfactorily and implementation is near. It illustrates the importance of long term research.

I'm not sure I understood the discussion in Waite's presentation of whether the optimization of the sample should be for the redesign alone or for the combined national and state supplement. Since what is referred to as the redesign will only last a few years and the larger sample will be used for most of the decade, I would have thought that the efficiency of the larger sample would be maximized. However, Waite reported the efficiency will be maximized for the redesign sample. Is this because there is uncertainty whether the state supplementation will ever be implemented?

One other part of the sample discussion also puzzled me. Waite pointed out that in the eleven largest states, it will not be necessary to add sample households to produce state data. The virtual self-weighting sample in CPS will supply enough cases in these states to satisfy the required precision. However, it is stated that after the two phase feature is implemented, some of the sample households in these states will be interviewed in the second week. Since only the first week will be used for national statistics, the reliability of the national data will be weakened after the state supplements are introduced. In a different part of his paper Waite says that the reliability of the national sample will be the same as the current CPS. I don't understand how this will be achieved if the first week's sample is reduced in the large states.

I don't have any major comments of the details of the work on questionnaire design described in the paper by Copeland and Rothgeb. The projects described should provide useful insight to the way respondents answer the questions in CPS. However, the paper seemed to me to illustrate the problems of trying to get answers to complex research questions in a time frame set by the needs for scheduled redesign. Early in their paper the authors point out that the alternative questionnaires tested in the late 1970's and 1980's were not implemented in CPS because of funding for the large overlap sample necessary to measure the change in the data series. Not tying the research to a time schedule would have permitted the accumulation of a sufficiently large sample size over time. Similarly, the last sentence of their paper says "Barring any unforeseen problems, the revised CPS questionnaire will be introduced into the CPS sample in January 1994." What if there are unforeseen problems? This is not infrequent in social research.