

THE ACTIVITIES AND FINDINGS OF THE PANEL ON ALTERNATIVE CENSUS METHODOLOGIES

Keith Rust, Westat, and Joint Program in Survey Methodology, University of Maryland
Keith Rust, Westat, 1650 Research Boulevard, Rockville, Maryland 20850

Key Words: Census coverage; master address file; dual-system estimation

1. Introduction

In April 1995 the Bureau of the Census asked the National Research Council's Committee on National Statistics (CNSTAT) to form a study panel to review plans and research and make recommendations regarding the design of the 2000 Census. The resulting Panel on Alternative Census Methodologies consisted of 12 members, several of whom had participated as members of one of the two predecessor CNSTAT panels on the 2000 Census. I was asked to chair the panel.

The panel's charge was to review the Census Bureau's plans for the 2000 Census, and to make recommendations regarding the census design. Specifically, we were asked to review the results of the 1995 and 1996 census tests, particularly with respect to the sample design for the nonresponse follow-up and the planned integrated coverage measurement sample design, to evaluate the statistical estimation procedures for the 2000 census, to recommend additional field tests and research to carry out before finalizing plans for the 2000 census, and to review the potential use of administrative records in the 2000 census.

Originally the panel was to have completed its work early in 1998. The panel's duration and scope were extended, however, to include consideration of planned experimentation and evaluation to be included in the 2000 census itself, to guide development of future censuses. The panel last met in June 1998, and released its final report in February 1999, following two interim reports and a letter report. The report is titled *Measuring a Changing Nation: Modern Methods for the 2000 Census*, and is available from National Academy Press (www.nap.edu; 800-624-6242).

In this paper I will discuss the scope of the panel's work, and its findings and recommendations, particularly those included in the final report. The conclusions that I will discuss are the results of the panel's deliberations, and have been subject to the NRC's formal review process. I will include some personal observations, however, on the processes that led to the panel's conclusions, and other panel members and staff may have different perspectives on some of these aspects.

In considering the panel's findings and approach, I think it is important to consider the charge to the panel, and findings of its predecessor panels. Both the earlier Panel to Evaluate Alternative Census Methods (Steffey and Bradburn, 1994), and the Panel on Census

Requirements in the Year 2000 and Beyond (Edmonston and Schultze, 1995), concluded that the methods of sampling for nonresponse follow-up, and integrated coverage measurement using sample data, appeared to be methods that could substantially improve the quality of census data, absolutely and especially per unit of cost. The subsequent Panel on Alternative Census Methodologies contained many members who had served on one of these earlier panels, including myself. Thus, one would not have expected this panel to conclude that these proposed uses of sampling and statistical techniques were inherently bad ideas that should not be pursued. Rather, the panel was, I believe, expected to use its expertise to judge whether the Census Bureau's approach to developing and implementing these methods was likely to be successful, and to comment on whether specific technical features could be improved. The panel's reports reflect this background.

2. The Scope of the Final Report

The panel's final report includes a number of components. The report gives findings and recommendations with regard to the methods to be used for the 2000 census, especially the novel ones. Over the period of the panel's life, the Bureau changed the details of the plans for many of these procedures. This was of course to be expected, since the 1995 and 1996 census tests provided vital information as to the success of various components, and pointed the way to improvements. We review the current status of plans, taking into account their history. In addition, plans may change as a result of considerations that arise when it comes to implementing them full scale for the census itself. This did occur in the cases of the plans for developing the Master Address File (MAF), and the mailing of replacement census forms to nonrespondents, and the report addresses the consequences and issues of these changes.

The timing of the panel's work did not permit us to evaluate the results of the important 1998 dress rehearsal, conducted in sites in California, South Carolina, and Wisconsin. However, we did provide suggestions as to what kinds of evaluations of the dress rehearsal data should be carried out in finalizing plans for the 2000 census.

A chapter in the report is devoted to reviewing the technical issues involved in implementing and evaluating statistical procedures for incorporating data on errors in coverage into the results of the census. In its interim reports, and again in the final report, the panel has endorsed the concept of using these

procedures as part of the census, and many of the plans for developing their implementation. However, we recognize that the value of this approach is far from self-evident, and there are serious technical concerns as to whether the use of these techniques is justified in practice, where data will be flawed and technical assumptions unverifiable. Thus, the panel thought that the final report was an appropriate point to review at least some of the technical concerns about “statistical adjustment” or Integrated Coverage Measurement (ICM), as it has been referred to by the Census Bureau. The report endeavors to explain why we conclude that, on balance, incorporating the sample data on under- and over-coverage, via statistical estimation, will improve census results.

The final activity of the panel was to consider a number of proposals by the Census Bureau for experiments to be imbedded in the 2000 census, with the purpose of collecting information to guide the development of methods for future censuses. The panel took this opportunity to suggest ancillary data gathering, and parallel activities, that we suggest can provide valuable data for future planning, and evaluation of the 2000 census, without the risk of interfering with operations in 2000.

As I mentioned, the panel completed its work in the summer and fall of 1998, and the final report was published in February 1999. On January 25, 1999, the U.S. Supreme Court ruled that sampling cannot be used to collect census data that will be used for purposes of congressional apportionment. As I understand it, the court made no determination as to whether or not the use of such an approach is allowed by the U.S. constitution. Rather, the court determined that an existing law, that governs the conduct of the census, prohibits such a practice, and that this law is valid.

Since, there is evidently no prospect that the law governing the conduct of the census will change before April 2000, this finding means that the planned new approach referred to as sampling for nonresponse follow-up cannot be used in 2000. The panel has endorsed this plan (at least in concept and general approach) in its interim reports. This is repeated in the final report, since, due to timing, there was no opportunity for the panel to review the Census Bureau’s changed plans in light of this decision, or even to propose a strategy. Since, it may be that implementation of sampling for nonresponse follow-up requires only a change in law, and not necessarily a change in the constitution, to become a legal practice, my view is that the panel’s findings, and especially the research that the Census Bureau has conducted, remain valuable for considering plans for census methods beyond 2000.

3. Plans for New Methods in the 2000 Census

A. Address List Procedures

The procedures for developing a national address list, as complete and of high quality as possible (and substantially improved from past censuses), is a cornerstone of the success (or otherwise) of the 2000 census, in the panel’s view. This topic was not highlighted in the charge to the panel, not has it attracted the outside attention and scrutiny that has surrounded the use of sampling and statistical procedures. Nevertheless, the panel closely followed the plans and developments of the Census Bureau in relation to building an address file, and creating maps. An accurate address file and mapping system will both act to reduce the kind of errors that the statistical procedures will be attempting to correct, but will also be key to the successful implementation of those procedures. For example, errors in matching households and persons in the census with those selected in the Post Enumeration Survey (PES), represent a significant threat to the quality of the results of coverage improvement via statistical techniques. Having a high quality address list for the census will help to keep such matching error rates low. A high quality address list is also needed for producing high quality data for small areas.

The quality of the address list is also an important measure for many stakeholders, who are both important users and key partners in conducting the census. If local officials are not convinced that the address list, and the process by which it is developed, are of uniformly high quality, it will be virtually impossible to conduct a successful census for which the results are widely accepted.

The MAF must be referenced to the correct geographic location in computerized census maps - the Topologically Integrated Geographic Encoding and Referencing (TIGER) system. A complete and accurate MAF-TIGER system is used to support key census activities:

- Postal delivery of census forms to households;
- Delivery of census forms in rural areas;
- Unduplication of multiple questionnaire responses from the same household; and
- Enumerator field follow-up of nonresponding households.

For the 2000 Census, the Bureau has adopted an approach of working to develop the MAF throughout the decade, rather than relying on an intensive effort

leading up to the census period. This process started with the file developed for the 1990 Census, updating it with data from the U.S. Postal Service (USPS) and input from local officials. At the national level, the Bureau has partnered with the USPS to make regular updates to the MAF, based on the USPS Delivery Sequence File (DSF). At the local level the Bureau has partnered with local governments through the TIGER Improvement Program and the Program for Address List Supplementation (PALS), to identify the street location of new addresses added to the MAF, and to supplement MAF improvements.

Through spring of 1997, the Bureau anticipated that these efforts, combined with some targeted fieldwork in particular areas, would be sufficient to build a high quality MAF-TIGER. However, at that time the Panel was concerned that these efforts would not necessarily be sufficient (White and Rust, 1997). The Bureau's own evaluation in mid-1997 concluded that the steps in place would not be adequate. Briefly, the DSF missed too many addresses resulting from new construction, and was not updated at the same rate across the country. The PALS failed to meet expectations because of insufficient response from many local governments, which lacked resources to participate, and because the Bureau could not handle the wide variety of formats in which submissions were received.

Thus, in September 1997 the Bureau added to the MAF-TIGER plan, by calling for greatly expanded field canvassing operations in 1998 and 1999, in similar manner to the blanket canvassing operations used in prior censuses. This effort is being combined with an approach of giving local governments the opportunity to review the Bureau's address list, in a program known as Local Update of Census Addresses (LUCA).

The success of this combined canvass and LUCA effort depends upon:

- The recruitment and supervision of a high quality field staff to carry the field canvassing in the limited time available;
- The ability to manage the field data and their incorporation into the files; and
- The ability to work in partnership with local governments to make the most effective use of the resources that they will have available for local review.

Thus, to a considerable extent, the Bureau has been forced to reintroduce expensive and intensive procedures for the completion of a high quality MAF-TIGER. While this is unfortunate, it is fortunate that the Bureau recognized this need in time to introduce the necessary intensive procedures at the end of the decade. Progress has been made in improving the MAF-TIGER, but not as great as was hoped for early in the decade

(and more expensively). The panel's final report contains the following recommendation 3.1:

The panel endorses the Census Bureau's plan to conduct a full canvass of the areas covered by the MAF-TIGER, which began in the fall of 1998 and will continue through 1999. In addition, the panel recommends that the Bureau investigate the usefulness of other data sources for updating MAF-TIGER during the coming decade, including address lists and maps from private companies and residential housing data from property tax records and maps.

B. The Date of Census Day

In preparing for the 2000 census, the Bureau considered pursuing legislation to move the date of the census from April 1 to mid-March, while retaining the mandated delivery dates for census data, specifically the December 31, 2000, deadline for delivery of state counts. This change would have provided two potential benefits.

One of the greatest difficulties in ensuring high quality census coverage are the instances of individuals and households moving close to census day. Such changes are more common right around the change of month, because leases are often written in terms of whole calendar months. No substantial evaluation of the potential change in improvement that might result in the U.S. if census day were to be moved to mid-month has been undertaken. Experience from the most recent Canadian census suggests that such benefits would likely be significant in the U.S.

The second reason for such a change would be to provide a small but critical amount of additional time to complete all census processes, including nonresponse follow-up and coverage improvement. The current schedule for the 2000 Census is very tight (of course this has been exacerbated by the introduction of the requirement that nonresponse follow-up be carried out on a 100 percent basis), and added time would reduce the chance that there will be insufficient time to resolve unforeseen problems, or to extend schedules to handle underanticipated demands.

The Bureau did not pursue this change, and this, in hindsight at least, was regrettable. Recommendation 3.2 reads:

The panel recommends that Congress enact legislation to move the date of the 2010 census to mid-March.

C. The Use of Blanket Replacement Census Forms

Early in the decade the Census Bureau conducted very promising research into ways to improve the return rate by mail of census forms. Much work was concentrated on the appearance and wording of the form and its accompanying material (a message on the

envelope indicating that response is required by law appears to be particularly effective). Considerable research was also directed to the use of prenotices, reminder notices, and the provision of replacement census forms. The research demonstrated that sending a replacement questionnaire to nonrespondents, after a certain period, was highly effective in improving the overall response by mail.

In the panel's view, all efforts are needed to achieve a high mail return rate. Apart from the obvious cost implications, when high rates of mail return are achieved, higher quality data is collected from the respondents (this was actually the motivation for introducing a mail census in 1970), and the reduced need for in-person follow-up means that the nonresponse follow-up operation can be completed more accurately.

In 1997, the Census Bureau determined that mailing replacement forms only to nonrespondents would not be possible in the time available. While a file of nonrespondents as of a particular cut-off date can be created very quickly, it requires several weeks to convert that into a targeted mailing to the tens of millions of nonresponding households. Only two options are available: send no replacement questionnaires; or send one to every household (blanket replacement).

The panel viewed this development with some alarm at the time. The concern centered on three issues:

- All of the testing done earlier in the decade had used targeted replacement, not blanket replacement;
- With blanket replacement the potential for duplicate returns seems likely to increase greatly (and perhaps overwhelmingly); and
- The perception that this process is wasteful and inefficient might cause a considerable public relations problem.

It seems clear, however, that the results from the earlier studies for the nonresponding portion of the population should hold, with blanket replacement (barring a public relations backlash). Thus, there are very important potential benefits that will be lost if no replacement form is sent. It thus became critical that the process of blanket replacement be fully scrutinized as part of the 1998 dress rehearsal, the first and only opportunity to evaluate this procedure. This led the panel to its recommendation 3.3:

If the 1988 dress rehearsal gives any indication that there are substantial problems (of extensive duplication of returned forms or public dissatisfaction) associated with the use of a blanket replacement form mailing, this procedure should be dropped and only a reminder postcard sent to each

household. Furthermore, the Census Bureau should explore all possible approaches to having available, for the 2010 census, technology that will permit targeted mailing of second forms only to households that did not return their first forms by a specific date.

D. Sampling Rates for Nonresponse Follow-up

This topic was an important one for the panel throughout its life, but became moot for 2000 once the Supreme Court decision was handed down. The plan was that as of a certain date, the Bureau would determine the population of households that had been sent a form (both original and replacement) by mail, and failed to return it. Within each census block a sample of these nonresponding households were to be scheduled for follow-up by field staff. This was a major departure from past practice (and, as it turned out, is illegal), when all such households have been scheduled for follow-up. This will now happen in 2000 also.

However, the panel's view that a more rational approach than was applied in planning for 2000 is needed to establish the rate of follow-up for nonresponse follow-up in different areas, remains valid if the law is changed to permit this in future censuses. The sampling rates that the Bureau proposed initially for 2000 would have resulted in very inequitable results for areas with differing mail return rates (with areas with the best mail return rates receiving results with the highest levels of sampling error). These rates were modified to remove this undesirable feature, but the rates that resulted from that process meant that the overall rate of nonresponse follow-up would have been about 75 percent. With such a high rate it would have been difficult to realize really significant savings in cost or time, or enhancements to data quality. This choice between two undesirable features was the result of a constraint on the sample design that the Bureau imposed on itself from the outset: that the sum of the mail return households and the sample nonresponding households must constitute at least 90 percent of the households in every census tract. Thus the panel concluded in its recommendation 3.6:

The Census Bureau should explore the advantages of sample designs for nonresponse follow-up that do not require a predetermined response rate and that can therefore achieve near equity in coefficients of variation across region, regardless of initial response rates.

E. Estimation for Integrated Coverage Measurement

The panel considered several features of the procedures by which the data from a PES should be incorporated into the census results. Chief among these were:

- Treatment of missing data;
- The use of demographic analysis;
- The use of dual-system estimation;
- Whether estimates should be completely independent from state to state;
- The use of raking procedures in estimation; and
- The construction of a transparent household file.

I will just cover the more major issues.

i. Demographic Analysis

Demographic analysis techniques are very useful for evaluating net census undercount, for the nation as a whole and for key demographic subgroups. The methods, however, are much less effective at establishing the geographic distribution of the undercount. The Census Bureau has considered whether it is feasible to incorporate the results of demographic methods into the process of determining final population counts from the census (Integrated Coverage Measurement). Several alternative approaches have been proposed, leading to quite different results, but difficult to distinguish in terms of quality. Furthermore, the methods rely heavily on the assumption that the results of the PES and dual-system estimation are unbiased for females, which will be violated to some, perhaps not ignorable, extent.

While the incorporation of demographic methods shows promise of future potential, and definitely warrants further research, the panel concurs with the Census Bureau's view that demographic analysis not be incorporated into the estimation procedures for the 2000 census. It will remain a valuable evaluation tool for 2000.

ii. Use of Dual-System Estimation

In planning for the 2000 census, the Bureau introduced plans for Integrated Coverage Measurement. This term covers the use of the PES data (and possibly other auxiliary data) to estimate the size of the net undercoverage, and to distribute it appropriately in census counts. Two alternative approaches were considered for incorporating PES data, referred to as Census Plus, and Dual-System Estimation. Dual-system estimation is the method that was used to estimate the net undercount following the 1990 census. The method recognizes the fact that both the census and the PES miss persons. It assumes that, within classes, the probability of being covered by the census is independent of the probability of being covered by the PES, and that for no class is the probability of both of these events equal to zero.

The Census Plus approach is actually a form of ratio estimation. In this method, it is assumed that, following the reconciliation of the PES and census

results, a complete roster of persons is available for each household selected for the PES. This is achieved by intensive probing of PES respondents about discrepancies between the PES and census rosters for the household, with a thorough canvass of all individuals who might conceivably have been missed, or incorrectly enumerated, based on census residency rules.

The Census Bureau hoped that, through the development of Computer-Assisted Personal Interviewing (CAPI) techniques in the PES, the Census Plus method could produce estimates as accurate or more so than the dual-system estimator, with rather less complexity of estimation.

However, the results from the 1995 census test, with confirmation in the 1996 test, convinced the Census Bureau and the panel that the Census Plus methodology is clearly inferior to the dual-system estimator. The positive development is that the CAPI methods developed with Census Plus in mind can be utilized to improve data quality in the PES, and hence enhance the dual-system estimator. This finding, of the superiority of the dual-system estimator over the Census Plus method, was reported in the panel's second interim report, and reiterated in the final report.

iii. Independence of Estimates Across States

An important issue in determining the exact form of the dual-system estimator is how one forms the poststrata, within which the dual-system estimator is applied, with results aggregated across poststrata. A key issue is whether the poststrata should, without exception, be formed within states.

The argument in favor of forming poststrata within states is largely one of face validity, or maintaining a system in which it is transparent that there is no favoritism of one state over another. When poststrata cross-state borders, then the results of the PES in one state affect the final census results in other states, when Integrated Coverage Measurement is applied. There is concern that this is inappropriate when (or, if) such adjusted census data are used to apportion congressional seats, or for allocating federal funds. If in fact, there is significant differential coverage between states within poststrata, this introduces bias into the estimates such that some states are "winners" and others "losers". By creating poststrata entirely within states, in fact the biases in the adjustments may be greater, but at least the data from one state does not affect the count in another state, although of course when a fixed quantity (such as the number of seats in the House of Representatives) is being apportioned among states, then in fact the data from one state do affect the outcome in other states.

The Census Bureau had decided, prior to the Supreme Court decision of last January, that poststrata should be defined within states. As is indicated in the

recommendation quoted below, the panel on balance supported this decision. This, however, was on the basis that counts adjusted for net undercoverage would be used to apportion seats in the House, along with all other uses, and also assumed that the PES sample size was to be 750,000 households. This view of the panel was based on the position that there is insufficient research available from which to make a case that, for a PES of that size, there would be likely to be appreciable gains from using strata that cross state boundaries, sufficient to outweigh the risks, both technical and perceptual.

It has now been determined that the adjusted results will not be used for apportionment, so that this alters the basis on which the panel made its recommendation. Also, a considerably smaller PES sample size is planned (300,000 households). The panel also suggested that the results using within-state poststrata might perhaps not be used directly for deriving final counts within each state. The panel's recommendation 3.8 reads:

The panel supports the decision of the Census Bureau to produce state total estimates using the 2000 census that are derived only from data collected within a given state. For the 2000 census, models across states should be examined for use in allocating populations within states. Both forms of the constraint on estimates that are based solely on data from a given state should be re-examined with respect to the 2010 census.

I believe that the circumstances for which the adjusted census data will be used have now changed so considerably that this particular recommendation (apart from the advice with respect to 2010) can no longer be viewed as necessarily representing the panel's position. That is, the recommendation was made within a context that has changed rather drastically.

iv. The Use of Raking Procedures in Estimation

The Census Bureau is considering using the statistical technique known as raking as part of its Integrated Coverage Measurement procedures (Farber et al, 1998). The aim of this technique is to preserve the stated goal that state total population estimates should be a function only of data collected within that state, but which permits the use of statistical models in which some terms are estimated using data across states (known as "borrowing strength"), in an effort to improve estimates at the substate level. Although, new to census methodology, the raking procedure has been routinely applied to survey data and other large data sets for over fifty years.

The panel appreciated the merits of the proposed raking approach, although it saw a need to investigate alternative methods to achieving the aims of the raking procedure, for possible use in future censuses. We were also concerned that the Bureau must work quickly to

determine exactly how to implement the raking procedure in 2000. The panel thus recommended (3.9):

The panel endorses the proposal to use raking ratio estimation to obtain substate estimates. Research should continue to define the poststrata and geographic regions as quickly as possible for the 2000 census and to examine alternative modeling options for use in 2010.

F. Other Topics Covered in the Report

The report also reviews several other innovations planned for 2000, particularly those for which the Census Bureau's plans have been modified over time. In the interest of economy of space these are simply listed here. In each case, the panel generally supported the Census Bureau's current plans and thinking on these topics:

- The use of multiple response opportunities (the "Be Counted" program), in which census forms will be available in a variety of public venues for those who believe that their household did not receive a form, or that they were omitted from the completed form;
- Sampling rates for the census-long form, which will vary by size of governmental unit, but with changes to the rates used in 1990;
- The sampling rate for follow-up of vacant units - this topic has also been rendered moot for 2000 by the Supreme Court decision in January;
- Hot deck imputation for nonresponse follow-up, as a means of estimation following sampling - again this is now moot for 2000;
- The use of CAPI for coverage measurement;
- The treatment of missing data in integrated coverage measurement; and
- The creation of a "transparent" household file, as a means of creating a household level data file that incorporates the results of integrated coverage measurement.

4. The Justification for Incorporating Coverage Adjustments in Census Counts

The question of whether or not estimates of net undercount, derived from data collected in a Post Enumeration Survey, should be included in census counts (or just used as an evaluation tool) has been a hotly debated statistical and political topic for the past two censuses, and even more so for 2000. The Swedish statisticians Lyberg and Lundström have described this issue as "the pinnacle of statistical methodological controversy" (Lyberg and Lundström, 1994). A number of articles in statistical journals have covered

this topic in detail, and a review of these indicates a divergence of opinion among statisticians who have reviewed the technical issues and the available data in great depth. The issues are covered in articles by Mulry and Spencer (1991, 1993), and three articles, and accompanying discussion, in the November 1994 issue of *Statistical Science* (Breiman (1994), Freedman and Wachter (1994), Belin and Rolph (1994)). These articles discuss the issue in the context of the 1990 census. More recently Brown *et al* (1998), have discussed this issue in relation to the 2000 census.

Our panel, in common with the two predecessor NRC panels on the 2000 census, has generally endorsed the Census Bureau's plan to incorporate coverage adjustment estimates derived from PES data into the census counts. Since, this viewpoint has previously attracted considerable opposition in the statistical literature, we have attempted to lay out our general views on the several technical issues involved. We have not, in our report, attempted to cover the detailed technical ground already described in the literature. We believe that the issues are clear, but complex to evaluate.

The literature highlights three areas of concern related to the methods of census adjustment via dual-system estimation:

- Matching error, when reconciling PES data with census data, and imputation for unresolved matches;
- Unmodeled heterogeneity in census undercoverage for lower levels of geographic aggregation; and
- Lack of independence of the probabilities of the events of being enumerated in the PES and enumerated in the census, and correlated heterogeneity of these probabilities, within poststrata in each case.

When one attempts to evaluate the extent to which any of these phenomena result in unacceptable problems with adjusted census counts, one must face the difficulty of establishing evaluation criteria as to which of two (or more) sets of estimates of the population size, broken down along geographic and demographic lines, is "better". There are essentially two parts to any solution to this problem - a decision whether to use population totals or population shares, and a loss function to summarize the fact that, for two given sets of estimates, one will each be closer to the "true" count or share in some cases but not others. It can be very difficult to arrive at such loss function that all can agree upon, and it may be necessary to examine several alternatives. One also faces the considerable problem that the true count is never known, making it hard to evaluate the loss function.

To illustrate, consider the following comparison. Consider three sets of hypothetical numbers. In each case, both the total population of a state, and the proportion of the U.S. population that is within the state, are given. The three sets of numbers are the true population, and two sets of estimates, A and B. One could think of set A as the unadjusted census counts, and set B as the adjusted estimates.

State	California	Wyoming	U.S. Total
True population	30,000,000	600,000	280,000,000
True percentage	10.7%	0.214%	
Population A	28,000,000 (-6.7%)	580,000 (-3.3%)	250,000,000
Percentage A	11.2% (+4.5%)	0.232% (+8.3%)	
Population B	29,000,000 (-3.3%)	640,000 (+6.7%)	290,000,000
Percentage B	10.0% (-6.7%)	0.221% (+3.0%)	

How does one decide which set of numbers is better, A or B? If you are interested in the population of California, the answer is clearly set B. But if you are interested in whether the relative shares in California and Wyoming are correct, the set A is evidently the better. Now add to this the fact that the true population of each state is not known.

Consequently, it is very difficult to judge whether the three sources of bias, listed above, combined with sampling error, render adjusted estimates that are not as good as the initial census count over which they are supposed to be improvements. The panel's view is that the evidence from past research indicates that the bias and sampling error from incorporating PES data through dual-systems estimation are likely to be smaller in total than the unadjusted estimates. This is more clearly the case for population counts but also most likely for population shares.

Data exist from the 1990 census that aim to quantify the amount of matching error. The data are problematic because the information used to determine that a given match was made in error, or incorrectly imputed, is itself subject to considerable error. It is not possible here to go through the panel's analysis of the information about matching errors (let alone to try to consider all of the technical arguments in the literature, so in summary, I will just say that the panel concluded that matching error, although a constant threat to introduce bias into adjusted estimates did not appear to be sufficiently great in 1990 (or the 1995 test census) to render the adjustment as less accurate than the unadjusted counts on average (and especially at higher levels of geography).

The second source of bias in adjusted estimates is the unaccounted for heterogeneity in census

undercoverage for smaller geographic areas. Direct estimates of undercoverage will exist at the level of poststrata, which represent relatively large levels of geographic and demographic aggregation, of the order of several hundred thousand persons. The results of these direct estimates are projected to smaller geographic units via a synthetic estimation approach. This approach essentially assumes that, within a geographic poststratum, undercoverage rates are constant within each demographic poststratum. This assumption is not true in general, which introduces a bias in the adjusted estimates. The panel reviewed some of the available evidence about the size of this bias in 1990 for geographic units of varying size, especially Wolter and Causey (1991). In an interim report the panel considered the argument that counts at the block level are important because they are used to construct larger units, such as congressional districts, and pointed out that it is logical to consider the level of error in counts for the resulting aggregates, rather than for each of the small components. Sampling error in particular, but also heterogeneity bias, will be mitigated for the larger units. Considering the planned size of the PES, in the panel's view it is unlikely that heterogeneity bias will cause adjusted counts to have, *on balance*, more error than the unadjusted.

The third source of bias arises from the fact that, within a poststratum, the variation in individual probabilities of being enumerated in the PES (given that the block is in the sample) are correlated with the probabilities of being enumerated in the census. Given that the methods of the PES and census enumeration are somewhat similar, this correlation is very likely to be positive in most cases. This results in biased estimates of the population. However, the resulting estimates will tend to fall between the census count and the true population. This means that for population counts the adjustments are clearly improvements, even though biased, but when one considers population shares the effect of the bias is less clear. The panel's perspective on this is summarized in the report: "...stated broadly, one should measure what one can and that, for what cannot be directly measured, it is appropriate to act consistently with the assumption that the part that cannot be directly measured is, at worst, uncorrelated with the part that can be measured. In addition, it seems unreasonable to ignore information about the distribution of a major part of the undercount because there is a hypothetical, unmeasurable, but very likely smaller component that, only if it had a particular (empirically unsupported) distribution, would cause adjusted shares to have greater loss than unadjusted shares."

It has not been possible to do justice to the technical issues of adjustment here. The research on this issue does, however, suggest that, if using dual-systems estimation to produce adjusted counts, it is vital that the Bureau develop high quality procedures for

carrying out the PES, reconciling match status, and implementing the dual-systems estimation. It is also critical that evaluation procedures, including especially demographic analysis, be developed for the 2000 census. And finally (a personal observation), one should consider the consequences very carefully before deciding to reduce the size of the PES sample.

5. Research and Experimentation in the 2000 Census

In 1998 the Census Bureau informed the panel about plans that were being developed for incorporating research and experimentation efforts into the 2000 census. The final activity of the panel was to consider these plans, comment on how valuable we perceived each of them to be, and pose alternatives that we felt the Bureau should consider as well.

The Bureau suggested a number of experiments to be imbedded in the census. Frankly, the panel was not overly enthusiastic about much of the proposed program, while recognizing that the issues that led to the proposals were important ones. There was a concern, strongly expressed by those with experience in the practical issues of census taking and other large scale data collection activities, that imposing experimental design features on a massive and time critical field operation, can place tremendous strains on the systems capability. It is difficult to organize a uniformly high quality effort from a large temporary work force with widely disparate skills under any circumstances, and asking these field staff to deal with the complexities of administering an experimental design invites problems. This is particularly so if several such studies are conducted at once.

So the panel felt that the proposed experiments should be subjected to several test criteria before being adopted:

- The information to be obtained could not reasonably be obtained by special test censuses during the coming decade--something about the circumstances requires a real census environment to give useful data;
- The impact on field operations activities must be minimal, and the success of the census not put at any risk;
- The experiment must provide measurable and interpretable results; and
- There should only be very few such experiments, at most.

We consider individual proposals below.

At the same time, the panel felt that too often opportunities are missed to collect auxiliary information during a census that can be used for evaluation and planning of future censuses. Thus, we consider that it is

important to consider carefully what data can be collected, perhaps on a sample basis, about census operations, without imposing experimental conditions on those operations - observe, not manipulate.

An example for the need for these kinds of data is provided by the case of cost and resource information about the household data collection by enumerators. Early in the panel's existence we were often given the impression that the last few percent of follow-up cases contributed very disproportionately to the total cost of the enumeration effort, and this certainly seems credible. However, when we were presented with research into the sample design for nonresponse follow-up, these generally involved assumptions that each personal visit household was equally expensive to enumerate. From this it became evident that no data are available from the 1990 census, in useable format, to quantify the extent to which a few percent of hard to enumerate households dominate the cost. We reached the view that if a truly efficient sample design for nonresponse follow-up were to be developed, detailed cost data were needed.

Such considerations led the panel to recommend that a master trace sample be collected in 2000. A master trace sample involves creating a data file that records detail on all census data collection for a sample of households--we suggest 100 tracts. For each household in these tracts, information would be stored on such characteristics as where the address came from, whether a long or short form was mailed, when it was returned, and, if the household did not respond by mail, how many visits the enumerator made in order to complete the enumeration, when they were made, how long they took, the reason for any lack of success, and so on.

Not coincidentally, the use of a master trace sample for the 1990 census was proposed by a CNSTAT panel in 1988 (Citro and Cohen (1998)). Due to budget and time constraints, this suggestion was not implemented. We feel that the Bureau has been suffering from the lack of these data in its efforts to improve the census process. So the panel's recommendation 5.1 reads:

The panel recommends that a trace sample be collected in roughly 100 tracts throughout the U.S. and saved for research purposes. The trace sample would collect detailed process data on individual enumerations. In addition, similar information on integrated coverage measurement should be collected, on a sample basis if needed. It would be very useful if information could be collected, again on a sample basis, to support complete analysis of the census cost model, all aspects of the amount of duplication and efforts to unduplicate, and information needed to support total error modeling of the 2000 census.

One of the earliest activities that the panel undertook was an evaluation of the potential for

administrative records to contribute substantially to the census process. Possibilities included enhancements to the MAF, provision of data for nonresponding households, provision of additional data for coverage checking, and even the complete provision of short form census data, at least in some areas. Once serious research began, it became evident to the Bureau and the panel that in fact none of these possibilities was likely to be a significant component in the 2000 census, due to the lack of administrative records that are up to date and contain sufficient reliable data to replace the collection of census data. Matching files from different sources also proved to be very difficult. However, certain files, in particular 1040 form files from the IRS, showed some promise as possible sources of data in the future.

It is clear, however, that a part of the process of instituting an "administrative records census" will be to carry out such a census in parallel with a more conventional census, for several large geographic areas (e.g., a few states). This will permit detailed review of the degree of success, and areas of weakness, in an administrative records census.

We therefore believe that, of the seven research experiments proposed by the Bureau that the panel considered, the Administrative Records Census Experiment, the Use of Administrative Records for Nonresponse Follow-up, and the Targeted Enhancements to the Master Address File, using seven national administrative records lists, are three that should be given the highest priority. In addition to evaluating the use of administrative records to provide Census data for households, the Administrative Records Census experiment also affords an opportunity to test a triple system estimation system, involving the census, the PES, and the administrative records data.

One other proposed experiment is seen as deserving high priority. This is the Alternative Questionnaire and Mail Treatment experiment. This has three parts:

1. The single page format of the questionnaire would be replaced with a booklet;
2. A reduced set of residence rules would be used in some questionnaires; and
3. In a sample of households replacement forms would be sent only to nonrespondents, rather than all households.

The panel views parts (2) and (3) as particularly important, as they have the potential to greatly enhance the census process, and must be conducted in a census environment to give useful data. The first part is of lower priority, as this could be effectively tested outside a true census environment.

The experiments that the panel thought were of lower priority are the use of the employee reliability inventory file for hiring nonresponse follow-up

enumerators, the census calling card incentive experiment, and the social security numbers, privacy attitudes, and notification experiment (if needed, this last could better be conducted in a future test census).

6. Conclusion

The Census Bureau faces many challenges in mounting a successful census in 2000. The Panel on Alternative Census Methodologies has reviewed the Bureau's plans leading up to the 1998 Census dress rehearsal. While finding a number of areas where improvements will be needed, building upon research and the dress rehearsal results, in the main the panel has endorsed the decisions of the Bureau as its plans for 2000 have evolved.

7. References

- Belin, T.R., and Rolph, J.E. (1994). Can we reach consensus on Census adjustment? *Statistical Science*, 9(4):486-508.
- Breiman, L. (1994). The 1991 Census adjustment: Undercount or bad data? *Statistical Science*, 9(4):458-475.
- Brown, L.D., Eaton, M.L., Freedman, D.A., Klein, S.P., Olshen, R.A., Wachter, K.W., Wells, M.T., and Ylvisaker, D. (1998). Statistical Controversies in Census 2000. Technical Report N. 537, Department of Statistics, University of California, Berkeley.
- Citro, C.F., and Cohen, M.L. (eds.) (1988). Priorities for the 1990 Census: Research, Evaluation, and Experimentation (REX) Program. Panel on Decennial Census Methodology, Committee on National Statistics, National Research Council. Washington, D.C.: National Academy Press.
- Edmonston, B., and Schultze, C. (eds.) (1995). Modernizing the U.S. Census. Panel on Census Requirements in the Year 2000 and Beyond, Committee on National Statistics, National Research Council. Washington, D.C.: National Academy Press.
- Freedman, D., and Wachter, K. (1994). Heterogeneity and Census adjustment for the intercensal base. *Statistical Science*, 9(4):476-485.
- Lyberg, L., and Lundström, S. (1994). Comment on three papers on census adjustment. *Statistical Science*, 9(4):508-517.
- Mulry, M.M., and Spencer, B. (1991). Total error in PES estimation of population. *Journal of the American Statistical Association*, 86:839-854.
- Mulry, M.M., and Spencer, B. (1993). Accuracy of the 1990 Census and undercount adjustments. *Journal of the American Statistical Association*, 88:1080-1091.
- Steffey, D.L., and Bradburn, N.M. (eds.) (1994). Counting People in the Information Age. Panel to Evaluate Alternative Census Methods, Committee on National Statistics, National Research Council. Washington, D.C.: National Academy Press.
- White, A.A., and Rust, K.F. (eds.) (1997). Preparing for the 2000 Census. Panel to Evaluate Alternative Census Methodologies, Committee on National Statistics, National Research Council. Washington, D.C.: National Academy Press.
- Wolter, K.M., and Causey, B. (1991). Evaluation of procedures for improving population estimates for small areas. *Journal of the American Statistical Association*, 86:278-284.

Acknowledgements

I would like to thank my fellow panel members, and my report co-editors, Michael Cohen and Andrew White, for their dedicated and insightful work during the panel's deliberations and the production of the report. I would also like to thank Ruth Ann Killion and the many staff members at the U.S. Bureau of the Census who contributed so much information and time to the panel over several years.