RESEARCH PLAN ON CENSUS ADJUSTMENT STANDARDS

Mary Mulry-Liggan and Howard Hogan

The Census Bureau expects undercoverage to be an issue in 1990 as it was in 1980 because federal funds are still distributed on the basis of census counts. Calls for census adjustment arise from an historical pattern of undercoverage. The Census Bureau's own evaluations of the past four censuses have indicated that minorities, particularly blacks, have been disproportionately undercounted.

For 1990, the Census Bureau has adopted a two-track approach to address the problem of undercoverage:

1. The goal is to count everyone.
2. The completeness of the coverage will be measured for the possible purpose of adjustment.

The Bureau has decided to establish in advance of the 1990 census standards or criteria that it will use to decide whether to adjust the census. These standards are necessary because we want to adjust the census only if we will be improving the data quality.

This paper discusses the research which the Census Bureau is currently conducting in preparation for setting the standards. First we will look at some background for the coverage error problem. Then we will examine the methods of measuring the quality of the census coverage. We will look at the standards that the Bureau used in the 1980 litigation over the Census. This will motivate the nature of the adjustment standards research. Then we will consider setting the standards, and the form these standards should take.

BACKGROUND

The undercount at the national level is defined by the average net undercount, that is the percentage difference between the estimated true population and the census figures. Net coverage has a different definition.

Net coverage
= true population - omissions + erroneous enumerations

Both misses and erroneous enumerations are important. At the national level, net undercount and net coverage error are equal. For subgroups they are not because misclassification errors exist.

Figure 1 contains estimates of the percent net undercount for the past four censuses. The net national undercount has shown a shady decline. The 1950 census had an undercount of 4.4 percent. In 1960, the undercount was 3.3 percent. The 1970 undercount was 2.9 percent, while the 1980 undercount was 1.4 percent. So we can see that over the past four decades, the net undercount has been cut from over 4 percent to 1.4 percent.

Figure 2 contains the estimates of the net undercount by race, for blacks and for whites and others. The net undercount for these groups has shown a steady decline. The undercount for whites and others has dropped from 3.8 percent in 1950, to 2.7 percent in 1960, to 2 percent in 1970 and to 0.7 percent. The undercount for blacks has dropped from 9.6 percent in 1950, to 8.3 percent in 1960, to 8 percent in 1970 and to 5.9 percent in 1980.

However, the estimates of differential undercount for these racial subgroups have remained about the same. Figure 3 contains estimates of the differential undercount by race. The undercount for whites and others has remained at about 0.5 percent less than the national average. The undercount for blacks has remained at about 5 percent above the national average. The Census Bureau has not made a dent in the differential undercount for blacks. The call for census adjustment arises not so much from a concern for overall census coverage, but a concern about this presistent differential undercount. In fact, the differential undercount for black males is about 7 percent.

The Census adjustment standards may have to address several dimensions of differential undercount. Differential undercount is complicated. The factors that are correlated with the differential undercount appear to be age, sex, ethnic group and geographic area. Other differentials in undercount may exist, although they are harder to measure. Central cities of large metropolitan areas seem to have a higher undercount. The undercount seems to fall as the central city and the metropolitan area become smaller. Rural areas also have high undercounts. The undercount seems to be higher for the poor, and for the single. Undercount seems to be higher among those who rent their homes than for those who own their homes. It may be that these social factors are more important in explaining the undercount than race itself.

METHODS

Four steps must be completed before the census can be adjusted:

- Measure the undercount,
- Estimate the undercount for local areas
- Adjust the data,
- Evaluate the results against our standards.

In 1990, the Census Bureau intends to use two methods to measure undercount:

- Demographic analysis,
- Case-by-case matching.

The demographic analysis technique uses aggregate sets of data. The estimates in Figures 1 through 3 are derived by demographic analysis. The sources used in demographic analysis include birth, death, and medicare records, estimates of immigration, and estimates of emigration.

These estimates are available for only the nation as a whole, not even for states. Other statistical techniques are needed to make undercount estimates for smaller areas.

The other technique using case-by-case matching can provide estimates of coverage error for subareas. Essentially, a sample of the population is selected independently of the census. Then each of these people is matched to the census. The reliability of this method depends upon getting a sample that is independent of the census and matching that sample to the census accurately.

Sometime near census day, the Census Bureau will conduct a household survey, a post-enumeration survey (PES). The objective is to list everyone living in the sample and find out where they were living on census day. In 1980 the Current Population Survey was used for evaluation and became part of the Post Enumeration Program.

The underlying assumption for this method is that

the census and PES are independent. Then the following two ratios are equal:

$$\frac{\text{sample matched}}{\text{sample total}} = \frac{\text{Total Counted}}{\text{Total Pop}} .$$

For this method to work it is not necessary for the PES to count more people than the census enumeration. It is only necessary that the survey picks up people that the census counted and people that the census missed in the correct proportions. This should be the case if a random sample of the national population is selected.

Algebra and the model that the proportions are equal provides an estimate of the total population:

$$\text{Total Pop} = (\text{Total Counted}) \frac{\text{Sample Total}}{\text{sample matched}}$$

This estimator is called the dual system estimator. In our application the two systems are the census and the PES. The PES and the census must be independent. Otherwise this equality is not true and a correlation bias is introduced into the estimate of the total population.

1980 STANDARDS

In 1980 the census undercount nationally was believed to be small - approximately 1 percent by the demograhic analysis estimate. However, there were those who were dissatisfied and filed suit requesting census adjustment. The decision was that the Census Bureau would not adjust except under a court order. The Census Bureau did not have an explicit set of standards to turn to for the decision.

The Detroit suit and the New York suit were the only two that made it to court. The decision in favor of Detroit was remanded on appeal, but the New York case is still pending.

The testimony in 1983, when the New York case was in court for the second time, is the most relevant to adjustment standards. The statistical standard on which the Bureau based its defense was that:

> Estimates of coverage error used in adjustment should have a smaller margin of error than the estimated census error.

The Bureau contended that estimates of coverage error from the 1980 Post Enumeration Program (PEP) were so flawed with nonsampling error that they were not suitable for census adjustment, particularly in light of the small undercount, 1 percent nationally (Wolter, 1984). Witnesses for the plaintiffs contended that regression analysis would smooth the effect of these errors.

However, the Bureau witnesses contended that the assumptions on which the regression analysis is based did not hold. This brings up another statistical standard.

> The assumptions on which the adjustment method is based should be stated explicitly, and there should be evidence that these assumptions hold.

The Bureau contended that the violation of assumptions was bad enough to cause bias in the regression estimates and make the estimates of the standard errors unreliable (Freedman, 1984).

If the assumptions underlying a method do not hold, then the results may not be valid. In fact, using the results to correct for undercount may give worse estimates than were avaialble from the census.

The Census Bureau contended that the uses of the census population estimates should be considered when judging the quality. The Bureau's statistical standard for judging the quality of the data is

> Slight variation in the assumptions underlying the adjustment methodology should not alter the conclusions drawn in using census population estimates.

However, this was not the case. Major uses of census counts, apportionment and revenue sharing, were shown to not be robust to variations in the assumptions (Wolter, 1984). The results were even counter-intuitive.

The plaintiff witnesses contended that the PEP data were sutiable for census adjustment. As a result of the testimony the Bureau decided that it needed census adjustment standards. The adjustment standards also must be of a form that we can tell when they are met.

ADJUSTMENT STANDARDS RESEARCH

The research on census adjustment standards has two major components. One is the development of a conceptual framework to measure improvement in population estimates. The other is the development of measures of the accuracy of the estimates of census error.

The nonsampling error in the estimates of census coverage error was an issue in the litigation over the 1980 Census. As a result the Bureau has designed research to investigate sources of nonsampling error in estimates of census coverage error. The Bureau plans to conduct a special-purpose coverage measurement survey. Estimates of coverage error will be based on case-by-case matching between the survey and the census.

The nonsampling errors that were of concern in 1980 were caused 1) by missing data, 2) by the clerical matching, 3) by evidence that PES and census were not independent, and 4) by the model for balancing the number missed with the number of erroneous enumerations. For 1990 these types of errors are being investigated in the context of a special-purpose coverage measurement survey, a PES.

The design of the PES for 1990 addresses nonsampling errors with a block sample, a survey independent of the census, and automated matching. Currently plans call for a block sample with a two-way match, from the PES to the census and from the census to the PES. This facilitates balancing the estimates of the gross number of erroneous enumerations with estimates of the number missed. The PES operation will be independent of the census operation which helps minimize the amount of correlation. In the evaluation of the test census the Post-Enumeration Surveys have been conducted in July. The scheduling is important so that the census and survey are independent but there aren't too many movers. The Bureau is developing an automated matching operation with clerical review. The uniformity that a computerized operation brings will control the matching error.

There are three types of nonsampling error research

in progress: theoretical, sensitivity analyses, and experiments. Correlation bias and the balancing are being addressed through theoretical work on the estimator. Sensitivity analyses are assessing the effects of different levels of nonresponse and matching error on the dual system estimator. We are designing an experiment aimed at measuring the matching and imputation errors during the evaluation of the test census.

The other component of our research is to develop a conceptual framework to judge the seriousness of errors in population estimates. This framework will give us something the Census Bureau did not have in 1980, a way to answer the question:

> How should one decide when the adjusted census estimates are superior to the unadjusted census estimates when the actual population count is unknown?

In 1990 the Bureau will have a data set that is the result of the field enumeration and a data set that is the result of the adjustment. Even if the true population were known, it is unlikely that one or the other data set will be closer to the truth for all states or areas.

Even in theory, we need a measure or a "yardstick," to judge the total error in a data set. We need to weigh the improvement due to the reduction in errors in order to decide which is better. In statistics common measures of improvement are the square of the error or the mean square error. These measures are used because often describing all the consequences of a decision is not feasible, and they are believed to be reasonable substitutes for most problems.

Statistical inference provides a way of formalizing the decision of whether to adjust the census counts although the "true" population counts are unknown. The need to make a decision when the "truth" is not known motivates all statistical investigations, even when the decision is the estimation of a parameter.

In the census adjustment setting, the true population counts are unknown. The unadjusted census estimates can be viewed as the realization of a random variable whose distribution depends on the true situation. Likewise the adjusted census estimates can be viewed a value of another random variable whose distribution also depends on the true population. The observations from the coverage measurement studies provide information about the distribution of the unadjusted and adjusted census estimates. These observations can be used to increase the chance of making a good decision in selecting the unadjusted or adjusted census estimates. One major goal of the Census Bureau's standards research is to design a loss structure appropriate for error in population counts. The size of the improvement needs to be expressed as a nonnegative function of the error. When the measure of improvement is specified, then a decision rule also can be developed.

Qualities that a measure of improvement should have are simplicity, reflect data uses, and reflect error sensitivity. Ideally we would want to develop a measure of improvement that reflects the true gain to society from reduced misestimation of population. However, we must be realistic as to what is feasible. This includes considering mathematical simplicity when we consider the effect of errors on the uses of the data.

The evaluation criteria for the Census Bureau's Post Censal estimates for revenue sharing contain some example of measures of improvement for population estimates.

The evaluation criteria for the post-censal estimates are

1. Average absolute error,
2. Average realtive error,
3. Amount of extreme relative error,
4. Amount of bias in estimates for subareas.

The methods of estimation are selected so as to maximize the improvement in these categories. The loss is determined by comparison with census results, special census results and among alternative estimates.

The uses of approtionment and federal fund allocation have been shown to be sensitive to errors in census estimates although their robustness has not been investigated. We are not restricting ourselves to revenue sharing. Most other programs are also competitive.

We intend to quantify improvement in census estimates in the context of their key uses. This appears to be a reasonable way of judging the quality of the census estimates.

Several people have published research that contained measures of improvement for population estimates.

A form that keeps appearing in the literature is

$$k \sum_{i=1}^{I} t_i \left(\frac{y_i}{t_i} - \frac{Y}{T}\right)^2$$

The notation is

- There are I areas.
- t_i is the true population for area i.
- y_i is the census estimate for area i.
- T is the true population count and Y is the national census estimate.
- k is a constant of proportionality.

This measure is the sum of the square of the difference between relative magnitudes of the error for an area and the nation as a whole weighted by the population size. This form merits our consideration because it has been shown to be relevant to fund allocation (Fellegi, 1980) and apportionment (Spencer, 1985).

We have already mentioned the problem of aggregation. Depending on the form of the measure of improvement, the decision on whether to adjust may depend upon which geographic level the measure is applied. Thus the measure and error structure may be such that improvements may be made by adjusting states but not counties, or vice versa. Further the political geography of this nation is not strictly heirarchial. Counties, cities, school districts and Congressional districts are laid across each other in no particular pattern. Adjustment will affect all these. The final standards must address explicitly the level at which they will be applied.

When the National Academy of Science Committee on National Statistics Panel on Census Methodology (1985) looked at the technical problems of census adjustment, it considered measures of improvement for population estimates. One of their recommendations was that the Bureau choose an adjustment methodology that is robust to a range of measures of improvement. By robust they mean that the same decision on adjustment is indicated by a

range of measures of improvement under given conditions.

Current research is aimed at assessing the implications of several common measures of improvement. We are identifying theoretical conditions under which the data are moved closer to the truth when the true population total is known or when the true population total is based upon an estimator with known distributional properties. We are considering several adjustment methodologies.

SETTING STANDARDS

There is another issue we must consider. How specific or general the standards should be?

The standards we saw from the NY litigation were general.

An example of a specific standard is one which identifies the maximum level of nonresponse acceptable in the PES.

The Bureau has to strike a balance between specific and general standards. They need to be flexible enough to allow for unforeseen circumstances. However, if they are too general they fail to give guidance to the decision process.

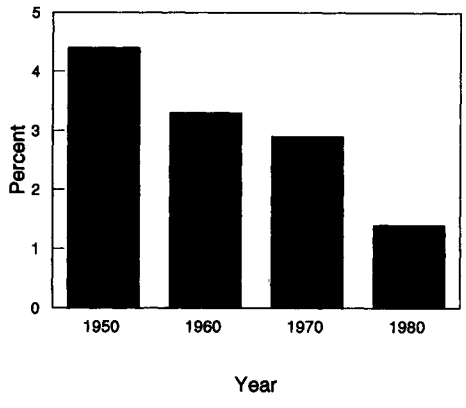
There are several issues to consider in forming the standards:

1. The standards need to be flexible enough to account for any type of systematic differential undercount.
2. How much of a guarantee of improvement is needed to justify an adjustment?
3. The geographic level that will be the focus of the decision making process needs to be decided.
4. The observability of the standards is also an important issue.

The observability of the standards is particularly important. It must be clear to all when the standards are and are not met.

The Census Bureau seeks a dialogue with the statistical community while we are performing the research and setting the standards. The interaction will lead us closer to a consensus on the criteria. Drafts of the standards will be available for external comment starting in May 1987.

Figure 1



Estimates of Net Undercount, 1950-1980

In the Spring of 1988, the Census Bureau must submit its census budget request to Congress. We must submit the adjustment standards for Congressional review at this time.

A more detailed description of the adjustment standards research plan is contained in a paper by Mulry-Liggan and Hogan (1986).

REFERENCES

Fellegi, I. (1980) "Should the Census Counts Be Adjusted for Allocation Purposes: Equity Considerations," Proceedings of the 1980 Conference on Census Undercount, U.S. Bureau of the Census.

Freedman, D. (1984) Affidavit in Cuomo vs Baldridge 80 civ 4550 (JES).

Mulry-Liggan, M. and Hogan, H. (1986) "Research Plan on Census Adjustment Standards," Proceedings of the Second Annual Research Conference, U.S. Bureau of the Census.

Panel on Decennial Census Methodology (1985) The Bicentennial Census, Citro and Cohen, eds., National Academy Press, Washington, D.C.

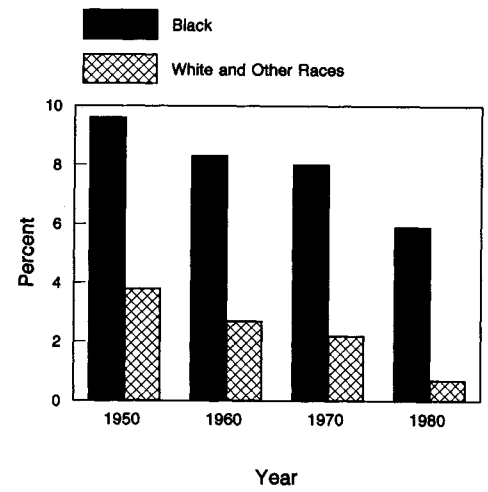
Passel, J.S. and Robinson, J.G. (1984) "Revised Estimates of the Coverage of the Population in the 1980 Census Based on Demographic Analysis: A Report on Work in Progress," presented at 1984 American Statistical Association meetings.

Passel, J.S. and Robinson, J.G. (1985) Memorandum for Roger Herriot, Chief, Population Division. Subject: Revised Demographic Estimates of Coverage of the Population by Age, Sex and Race in the 1980 Census," U.S. Bureau of the Census.

Spencer, B. (1985) "Statistical Aspects of Equitable Apportionment," Journal of the American Statistical Association, 10, 815-822.

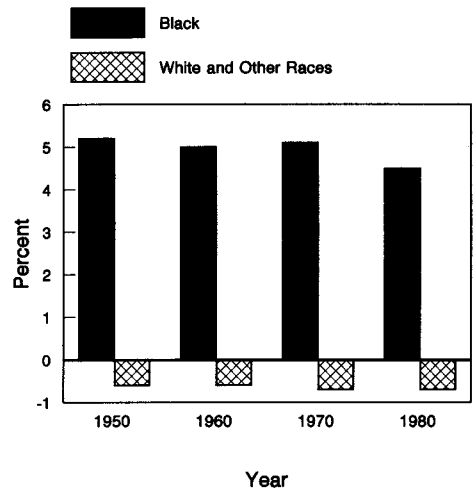
Wolter, K. (1984) Affidavit in Cuomo vs Baldridge 80 civ. 4550 (JES).

Figure 2



Estimates of Percent Net Undercount by Race, 1950-1980

Figure 3



Estimates of Percent Differential Undercount by Race, 1950-1980