NINE YEARS OF COVERAGE EVALUATION RESEARCH: WHAT HAVE WE LEARNED?

Howard Hogan, Bureau of the Census
Washington, D.C. 20233*

Keywords: Census, Coverage, Undercount, Post-Enumeration Surveys

1. Introduction

In 1980 the Bureau of the Census began an intensive program of coverage evaluation research. Its aim was to develop methods to evaluate census coverage that were both timely and accurate. That program is now nearly over. We have chosen our methods for 1990 and will be implementing them soon.

This paper reviews what we did and what we have learned. It starts with coverage measurement research done before 1980, but concentrates on the research program for the 1980's. It ends with a brief discussion of what we still have to learn and where we go from here.

2. Early Research

2.1 A Research Proposal for a Study of Methods for 1990 Decennial Census Coverage Evaluation

Explicit planning for 1990 had begun by May 1980 with the development of research proposals and experiments into alternative methods. By 1981, the Bureau had developed a Research Proposal for a Study of Methods for 1990 Decennial Census Coverage Evaluation. (Mulry et al, 1982). The report reviewed the theory, and developed some standard nominclature. After reviewing the available methods, the report proposed eight studies. Several of these studies were funded and implemented.

2.2 Reverse Record Check Research

There was considerable disenchantment with the Post-Enumeration Survey methodology based on the Bureau's experience in the 1980 PEP, specifically with respect to the problem of correlation bias The Reverse Record Check approach seemed to hold promise. The RRC samples from a nearly complete frame, i.e. a combination of the previous census, the previous PES or RRC, the list of intercensal births, and immigrates. The method had been used in Canada for many years with good results. The research question was whether we could successfully trace such a sample in the United States and match it to the Census. Two projects were implemented to test these ideas.

The CPS/Census Retrospective Match matched people listed on the 1977 CPS to the 1980 Census. Its purpose was to see how well we could match a sample gathered some years before, and whether we could trace and interview those who had moved. The other study was the Forward Trace Study. We felt we could keep more people in sample if we did not wait until the end of the period to begin tracing. By selecting a sample at the beginning and then following the people over the decade, perhaps we could maintain more people in sample.

The result of these studies was to conclude that the reverse record check approach was not workable. Given the mobility of our population and the ten years between sample, we lost too many sample cases.

2.3 Administrative Record Checks

The Bureau again looked at administrative record checks. A match of Medicare record to the Census had been conducted in 1970. The Research Proposal suggested a match of records from the 1979 IRS Individual Master File to the 1980 Census. Unmatched cases were followed-up. The results were interesting, indicating for example a higher census omissions rate for those who file single returns vs those filing jointly. However, we decided that the IRS files excluded too many types of people to be used by themselves.

2.4 Computerized Matching

The Research Proposal recognized the need for computerized matching for any future match studies and to the formation of the Record Linkage Staff. The staff proceeded to develop computer software and statistical analysis needed to implement the Fellegi/Sunter model of record linkage to the problem of census evaluation. (Jaro 1989). These programs were implemented successfully in later PES tests.

3. Improvements in Post-Enumeration Survey Methods

3.1 1984 Research Plan on Adjustment

At the beginning of 1984, the Census Bureau set up the Undercount Research Staff in order to conduct, coordinate and monitor research into undercount measurement and undercount correction. One of its first official acts was to issue The Research Plan on Adjustment. The plan laid out plans for developing improved methods of census coverage measurement and undercount correction.

3.2 1985 Test PES in Tampa Florida

The 85 Test reflected several basic design desisions. It was a sample of whole blocks. We felt that the computer matcher would work better in this situation. The P and the E sample were drawn from the same blocks.

The 1985 Test PES was designed principally to test the new computer matcher and to test a newly designed PES questionnaire. The test was successful in its goals. The computer

matcher was able to link 85.9 percent of the matches eventually found. All matches assigned by the computer were reviewed by clerks. Only 0.174 of matches assigned by the computer were actual errors, (Jaro, 1989) We developed a computer generated match form which listed all members of linked households on the same form. This form presents the information in a neat and easily accessible way.

Non-response was lowered over the 1980 rates. The whole household non-interview rate was only 3.7, for a sample concentrated in inner city area with large minority populations. The 1980 (national) whole household non-interview rate was 4.4 for the April sample and 5.3 for the August sample.

3.3 1986 Test PES in Mississippi

The Mississippi Post-Enumeration Survey was designed to test our new matching system in a rural area. Since the computer matcher had been only tested in an area with street names and house numbers, many doubted that it could work in an area of mainly rural routes and box numbers. Instead of relying principally on addresses, the computer matcher used telephone numbers in conjunction with names. The computer system was able to match 67.7 percent of the cases, or to express it another way 79.8 percent of all cases that were matched before follow-up. Of the cases matched by the computer 0.9 percent were found to be matched erroneously. (Anolik, 1988).

3.4 1986 Test of Adjustment Related Operations

The 1986 Test of Adjustment Related Operations was designed to test the operational feasibility of producing adjusted census counts by December 31, 1990. TARO was conducted in six small towns in East-Central Los Angeles County. The area was predominantly Hispanic with a large Asian immigrant community.

TARO began by drawing a stratified block sample in which all addresses were then listed. Interviewing began in late June and lasted, including clean-up, through early August. The P sample was all people living in the blocks at the time of the PES. The E sample was all enumerations coded to the blocks.

TARO did not meet its timing goal. The early operation of TARO went well. Initial interviewing was complete by August 8. Follow-up and After Follow-up Coding and Matching were complete by November 6. Two months were then needed to prepare the P and E sample files for estimation. The Final Census files were ready on January 5. The dual system estimates and the small area estimates were completed by February 22.

Although it did not finish by December 31, TARO represented a great improvement over any previous PES or PES test.

TARO utilized the small area estimation methods worked out by Isaki and colleagues. After dual-system estimation, adjustment factors were computed. A regression model was fitted to these, based on indicator variables such as age group, ethnic group, and whether the person rented or owned. Then the sample-estimated adjustment factors are shrunk toward this common pattern. The adjustment process added 8.2 percent to the census. This was lower than the direct estimate of 9.0 percent due to the way the smoothing process lowered the extreme values. (Diffendal, 1988).

In order to evaluate the accuracy of the PES, the Bureau had developed a generic list of eight errors. These are:

Matching error
Reporting census-day address
Fabrication in the PES interview
Missing data
Error in measuring the erroneous enumerations
Balancing gross overcounts and undercounts
Correlation Bias
Random Error

The TARO estimates were evaluated with respect to each of these errors (Hogan and Wolter, 1988). They conclude that the joint effect of these errors was to cause the 1986 TARO estimates of undercount (9.0 percent) to overstate the true undercount by about 1.2 percent. The corrected figure, 7.8 percent, may be viewed approximately as the means of a posterior error distribution for the TARO undercount. This number was 0.4 percent less than the 8.2 percent adjustment actually applied. Mulry and Spencer developed a complete posterior error distribution.

Jones (1988) also prepared an evaluation of the TARO results and arrived at a very different result. Jones's analysis led him to conclude that the true undercount was 3.6 percent or less than half of the 8.2 percent adjustment applied.

3.5 The 1987 PES of North Dakota

In 1987, the Census Bureau conducted a PES in conjunction with the Test Census in central North Dakota. The goals were to test methods of PES sampling in so-called List/Enumerate areas and to test test rural matching. The study did not include follow-up but still matched 94.8 percent of the in-scope cases.

As part of the North Dakota test, we tested the PES field quality control procedures. "Fictitious" cases were seeded into the QC workload to see if the procedures would detect them. All cases were detected.

3.6 1988 Dress Rehearsal Post-Enumeration Survey

The Dress Rehearsal Post-Enumeration Survey was conducted in conjunction with the 1988 Dress Rehearsal Census. In general, the

1988 Dress Rehearsal PES followed the steps laid out in earlier tests. Several changes in the PES process were instituted. The address listing was conducted in February. The computer matching process was streamlined. Matching for nonmovers was restricted to only the sample block and surrounding blocks. Since the census questionnaires were not sorted geographically, matching of movers was done using the microfilm of the census questionnaires rather than the original copies. The Census Bureau developed a computerized data entry and quality control system for matching that checks the match results for consistency and appropriateness. (See Childers and Hogan, 1989).

Part of the Dress Rehearsal test was a test of using administrative records to supplement the PES in hard to count areas. This was seen as a way to reduce response correlation bias. This study is described in Wolfgang, (1989).

4. Other Methods

4.1 Pre-Enumeration Survey

In theory, a pre-enumeration survey is similar to a post-enumeration survey - only the interviewing is done before the census. The PrES. was of interest because it held the possibility of finishing the evaluation earlier than the PES. We were concerned, however, that the PrES might influence or contaminate the census mailback in the sample blocks.

The research concluded that there was some influence on the census blocks. This could be measured by comparing census mailback rates and edit-failure rates for a sample of paired blocks that were not part of the PrES. More important was the fact that the PrES held no timing advantage over a PES. The key constraint to when matching can begin is when the census files are ready. The availability of the PES files are not the constraint. Thus there were potential disadvantages to the PrES and no apparent advantage.

4.2 Ethnographic Research

There are some people who will never be picked up by any household survey. Another approach is needed. As early as 1971, the Bureau had used urban anthropologists to investigate the coverage for selected small areas. Interest in this type of research was renewed when the National Academy's panel recommended the Bureau initiate a research program using what it called resident or systematic observers. As a result, the Bureau of the Census formed agreements with three ethnographers to conduct research in a total of six Hispanic blocks in the Los Angeles test area.

The ethnographers counted 10 to 27 percent more people than did the census. A significant number of the uncounted Hispanics lived in illegally converted apartments that were missed in the census. Interestingly, both males and females were missed in roughly equal numbers, and many missed people were under age 15. (Hines, 1988).

The Census Bureau continued its research during the 1988 Dress Rehearsal. Study populations included urban Indians in St.Louis, poor blacks in St.Louis and migrant workers in Washington State. The Bureau is currently locating and recruiting qualitified researchers to conduct studies for the 1990 Census.

4.3 Progress on Demographic Analysis

The demographic analysis estimates published for the 1970 Census had made no allowance for "unrecorded alien immigration". (U.S. Bureau of the Census, 1974, p15) The initial coverage estimates made in 1981 also made no allowance for the resident population of undocumented immigrants.

In a series of articles, Passel and his colleagues (Warren and Woodrow) produced estimates of the number of illegal residents counted in the 1980 census. For 1990, the preliminary demographic analysis undercount estimates will include a component that reflects the increase in the population between 1980 and 1990 due to undocumented immigration. These estimates of net undocumented growth are already included in the Census Bureau's postcensal estimates of population.

4.4 Causes of Undercount Study

Coverage evaluation is, or should be, more than just trying to measure the census net undercount. It should include efforts to understand the causes of the undercount. As part of the 1986 TARO, the Census Bureau conducted a survey to determine why people were missed in the test census. (Fein and West, 1988) A sample of missed people were reinterviewed to determine their attitute toward government, the sources of income, and other characteristics that might make them hard to enumerate. The chief importance of this survey is not so much in the actual findings, which, after all, reflect only the experience of a test in one area. More importantly, the study served as a reminder that the purpose of census coverage evaluation is not just to measure the net undercount, but to learn about the processes so that the problem can be corrected.

5. Small Area Research

Much of the attention of the 1980 Conference on Census Undercount was directed at methods to estimate the undercount for small areas. Various models were proposed. This line of research has continued with much vigor.

5.1 Scherm and Preston

In an important article Schirm and Preston (1987) tackled the issue of using a synthetic estimate by race for states. They concluded, that "over a wide range of environments, nearly two out of every three simulated applications of

synthetic adjustment improve the state proportions for a majority of the national population." A valuable contribution of this article concerned the criteria for measuring improvement. Scherm and Preston carefully defined several statistical measures of improvement (or loss functions). Many of these measures were used in later research.

5.2 Isaki et al

The issue of producing small area estimates of population was tackled most directly by Isaki and his colleagues. Their research covered a wide range of issues, of which I will only mention a few.

A core focus of their research has been to determine the level of error introduced in the small area estimation process. One result was that the estimated population was better than the census for areas the size of counties. For very small areas such as "enumeration districts," the estimated populations were no closer to the truth than were the census numbers.

In another line of research, Isaki and Diffendal investigated methods of smoothing the estimates produced by a PES. (See Isaki). Using indicator variables for age, sex, race and Hispanic origin, plus the number of census substitutions, they formed estimated undercount rates for each post-strata. They combined these with the direct estimates from the PES inversely proportional to the variances. This method reduced the variance by as much as one third. (See Diffendal 1988). Cressie (1988), also working with the Bureau, has pursued related research (Cressie 1988).

5.3 Causey and Wolter

Causey and Wolter (1988) investigated across-the-board ratio estimation and synthetic estimation that might be used for producing population estimates for small areas. The methods emphasize determining the breakeven accuracy of the estimate of the population totals that marks the point at which improvement occurs. They assume that either the direct estimates are unbiased, or that the bias can be estimated. In the latter case, the bias is corrected for and the variance of estimate of bias can be incorporated into the overall estimated variance.

They found that for counts, synthetic correction always beats across-the-board correction and, except at the enumeration district level, beats the original enumeration. For proportions, synthetic correction always beats the original enumeration and across-the-board estimates, even for enumeration districts. In general, the improvement for proportions is somewhat less than the improvement for counts. The improvement for both the across-the-board and synthetic is greatest for states and least at the enumeration district level. In general, we do not need to know the factors as accurately (in terms of CV) at the strata level in order to make synthetic estimation worthwhile, as we need to know them in order to make across-the-board estimation worthwhile.

6. Missing Data Research

6.1 Logistic Regression Models

The high levels of non-response in 1980 had highlighted the need for a better understanding of the missing data mechanism and better missing data models. Much of this research was done in a cooperative arrangement between the Bureau of the Census and the Department of Statistics at Harvard University.

In 1980, much of the missing data was handled using an imputation procedure based on the cases status before followup, its race/ethnicity and other demographic characteristics and its geographic location. We saw this as having two disadvantages. First, it was hard to define the model explicitly. It was difficult to assess the variability due to imputing these statuses.

Schenker (1988) developed logistic regression models for handling cases where the enumeration status could not be determined. The model predicted a probablility of being correctly enumerated. Rather than using this to impute, the method summed the probablilities to get an overall estimate. This had two advantages. It was more efficient than imputing a zero/one enumeration status. Second, the imputed probablilities represented the uncertainty about the missing enumeration status. (See Schenker, 1988 and Rubin *et al*, 1988).

6.2 Ignorable Non-Response Models

A particularly important question with respect to the 1980 PEP was how to handle cases that became non-response only during follow-up. These were cases, initially considered to be complete interviews, that did not match, either because they were not enumerated or the matching was incorrectly done. These were sent to follow-up, and many returned without a complete follow-up interview. Much of the range of the 1980 estimates involved the question of cases with missing data during follow-up. For example, all estimators imputing based on only follow-up cases showed a net undercount, all estimators based on weighting showed a net overcount.

Research has been going on to resolve this issue, as will be reported by Joe Schaffer (1989).

7. Adjustment Research

In order to be able to incorporate the adjustment into the actual census tabulations, the Bureau undertook research into integerization. For ease of computation and for user acceptability, the Bureau wanted to add whole people to the counts. The methods yielded fractions. That is, within any block, the best estimate of the number of people to be added by age, race, sex and Hispanic origin

would normally not be a integer. A simple block by block integerization might lead the marginal totals far from the best estimate. A method was developed and tested to integerizate the cells (blocks by age, race, and sex) while controling the marginals by block and by age, race and sex. This methodology was tested in TARO. The procedure went smoothly, but of course the number of blocks to be adjusted represented only a small proportion of what would be necessary for a national census.

8. Measuring Error In Census Coverage Estimates

An important part of the adjustment research was developing explicit standards for deciding when the "adjusted" numbers were better than the "unadjusted." Sets of preliminary standards were drafted and circulated. (Hogan and Mulry 1987a,b). Although this specific activity was temporarily stopped when the Department of Commerce decided not to adjust the 1990 Census, much of the statistical work was incorportated into work on measuring and reporting the error in census coverage estimates.

The Bureau of the Census has been conducting research aimed at measuring the level of error in the PES. We have attempted to measure each of the eight components of PES error. The effects of P-sample fabrication will be examined by using the quality assurance results to estimate the residual error. We will also look for block or interviewer effects that might signal PES fabrications. (See Stokes and Jones, 1989). Some of the reinterview studies might also turn up evidence of PES curbstoning not previously detected.

Mulry and Spencer (1988) have investigated the total error in the dual system estimator. The model relates the observed indicator of data quality, such as the matching error rate, to the first two moments of the components of error. They investigate the propagation of error in the DSE including its bias and variances.

The demographic analysis estimates are widely accepted as the most accurate available at the national level. However, even at that level they are not without error. Knowledge of that error is essential if differences between the demographic analysis estimates and the post-enumeration survey are to be understood. (The demographic estimates can be thought of as a sum of many individual components: births, immigration, deaths, medicare population etc.) In order to construct uncertainty measures, there are several important considerations. First, few of these components have explicit variances attached. A strictly classical approach would not work. However, if one adopts a Baysian (or Baysian-like) approach, one can model the errors or uncertainty surrounding the preferred estimates. The Bureau is currently tackling these issues (See Passel 1988, Robinson and Das Gupta 1989).

9. Future Research

9.1 Role of Follow-up in Post-Enumeration Surveys

Perhaps the most interesting statistical controversy concerning the PES design is in defining the proper role for PES follow-up. In 1980, the rule was that all non-matched PEP cases had to be sent to follow-up. The rules used in the tests this decade allow for some of the non-matched cases to be called non-enumerated on the basis of the initial interview alone. These two approaches are based on two different philosophies about the PES. The correct resolution of this problem will require careful analysis of the role of unique-address matching, recall bias in both the P sample and E sample, the accuracy of the initial match, the accuracy of proxy reporting, the imputation model to be used and the level of PES fabrication in both the initial and followup interview.

9.2 PES and Demographic Analysis

In the past, demographic analysis and the PES were reported separately. There was no attempt to synthesize them into a single estimator. For 1990, we plan to produce a combined estimator from the results of both of these studies. The method currently under investigation concerns using the sex ratios obtained from the demographic analysis estimates to correct the results of the PES for differential correlation bias between the sexes. That is, traditionally the PES has underestimated the undercount of males relative to females. This tendency is especially pronounced for blacks. The demographic analysis sex ratios gives us the opportunity to model the correlation between the census and PES for males and thus to make a correction. Since the demographic analysis sex ratios are available only at the national level, such a correction must entail some synthetic modeling to the stratum level.

9.3 Alternative Estimates and Census

To produce the best population estimate for a given area, one needs to look at both the census result for that area and the results of the evaluation. Traditional statistics would combine the two inversely proportional to the errors in each. However, it is from the evaluation survey that we derive the estimate of the errors in the census. Depending on the level of the undercount and the level of error in the PES, there may be some, perhaps many strata, where the best estimate of the undercount is the unadjusted census counts. In other, harder-to-count areas, the PES may be better. Decision rules must be devised.

10. The 1990 Post-Enumeration Survey.

Research into coverage evaluation methodology is not an end in itself. The true goal is to produce estimates of the population, and to measure the accuracy of those

estimates. The goal of the 1990 PES is to produce corrected population counts usable for congressional and legislative reapportionment, redistricting and all other purposes for which the Bureau of the Census publishes data. If the Secretary of Commerce decides we have met that goal, adjusted figures must be published by July 15, 1991. The 1990 Post-Enumeration Survey will begin in January with advance listing of all addresses in the sample blocks. We will soon know if nine years of research have borne fruit.

11. References

Anolik, Irwin, 1988. "The 1986 Post Enumeration Survey of East Central Mississippi."

Causey, Beverley and Wolter, Kirk, 1988. "Evaluation of Procedures for Improving Population Estimates for Small Areas" Unpublished Manuscript.

Childers, Danny R., 1986. "The 1985 Post Enumeration Survey," Unpublished Manuscript.

Childers, Danny, and Hogan, Howard 1989. "The 1988 Dress Rehearsal Post-Enumeration Survey, Design and Preliminary Results," paper presented to the American Statistical Association Annual Meetings.

Cressie, Noel, 1988. "When are Census Counts Improved by Adjustment?," *Survey Methodology*, Vol. 14, No. 2.

Diffendal, Gregg, 1988. "The 1986 Test of Adjustment Related Operations in Central Los Angeles County" *Survey Methodology*, Vol. 14, No. 1, June 1988.

Fein, David, and West, Kirsten, 1988. *The Sources of Census Undercount: Findings from the 1986 Los Angeles Test Census. Survey Methodology*, Vol. 14, No. 1, December 1988.

Isaki, Cary T., Schultz, Linda K., and Diffendal Gregg J., 1987. "Report on Using Regression Models for Small Area Adjustment," Statistical Research Division Report Series, No. CENSUS/SRD/RR-87/01. Bureau of the Census, Washington, D.C.

Isaki, Cary T., Schultz, Linda K., Diffendal, Gregg J., and Huang, Elizabeth T., 1988. "On Estimating Census Undercount in Small Areas," Journal of Official Statistics Vol. 4, No. 2 (to appear).

Isaki, Cary T., Huang, Elizabeth T., and Schultz, Linda K., 1988. "Model Bias and Sampling Error Considerations in Small Area Coverage Error Estimation," Statistical Research Division Report Series, No. CENSUS/SRD/RR-88/17. Bureau of the Census, Washington, D.C.

Hines, Catherine, 1988. "The Role of Participant Observation Research in Understanding the Census Undercount," paper presented to the Population Association of America.

Hogan, Howard and Mulry, Mary, 1987. "Preliminary Census Adjustment Standards," U. S. Bureau of the Census, Unpublished.

Hogan, Howard and Wolter, Kirk, 1988. "Measuring Assuracy in a Post-Enumeration Survey," *Survey Methodology*, Vol. 14, No. 1, June 1988.

Jaro, Mathew, 1989. "Advances in Record-Linkage Methology as Applied to Matching the 1985 Census of Tampa, Florida," *JASA*, Vol 84, No. 406, pp 414-420.

Jones, Charles D. 1988. "An Evalutation of TARO", Unpublished Manuscript, June, 1988.

Mulry, Mary, et. al., 1981. "A Research Proposal for a Study of Methods for 1990 Decennial Census Coverage Evaluation," (unpublished).

Mulry, Mary and Spencer, Bruce 1988. *Survey Methodology*, Vol. 14, No. 1, December 1988.

Passel, Jeffrey, 1987. Presentation to the American Statistical Association Census Advisory Committee.

Robinson, J. Gregory and Das Gupta, Prituis, 1989. Presentation to the American Statistical Association Census Advisory Committee.

Rubin, Donald, et. al., 1988. "Imputation Strategies for Estimating the Undercount," Proceedings of the Fourth Annual Research Conference. Bureau of the Census.

Schaffer, Joseph, 1989. "Why the PGP Series 14 Estimates Cannot be Trashed," ASA Proceedings. (This issue)

Schenker, Nathanial, 1988. "Handling Missing Data in Coverage Estimation, with Application to the 1986 Test of Adjustment Related Operations," *Survey Methodology*, Vol. 14, No. 1, June 1988.

Schirm, Allen, and Preston, Samuel, 1987. "Census Undercount Adjustment and the Quality of Geographic Population Distributions." *Journal of the American Statistical Association*, Vol 82, No 400.

Stokes, Lynn and Jones, Patty, 1989 "Detection Methods for Interviewer Fabrication in the PES."

Wolfgang, Glenn, 1989. "The Adminstration List Supplement." ASA Proceedings. (This issue).

* This paper reports the general results of research undertaken by Census Bureau staff. The views expressed are attributable to the author and do not necessarily reflect those of the Bureau of the Census.