Planning for the census in the year 2000: an update 19 June 1997

Technical report No. 484, Department of Statistics, U.C. Berkeley

by Morris L. Eaton, David A. Freedman, Stephen P. Klein,
Richard A. Olshen, Kenneth W. Wachter and Donald Ylvisaker

Abstract

This paper is a discussion of Census 2000, focusing on planned use of sampling techniques for non-response followup and adjustment.

Introduction

Our object in this paper is to discuss plans for use of sampling in Census 2000. Sampling comes into the design in three ways.

1) Traditionally, long-form data are collected on a sample basis; the Bureau plans to do so in 2000. The long-form data are valuable to many users, and we hope these data can be insulated from the controversy about sampling as a basis for counting the population.

2) Non-response followup in the census has in the past been done on a 100% basis. In 2000, however, the Bureau plans to follow up only a sample of non-respondents. For reasons detailed below, we think that sampling for non-response may run into trouble in the census context.

3) The most salient—and problematic—use of sampling in Census 2000 is "Integrated Coverage Measurement," or ICM. Despite the name, the ICM program is not in our view an integral part of the census operation. Instead, it is an effort to adjust the census counts with the hope of reducing coverage errors. The Bureau plans to release only one set of numbers, based on the census and sampling for non-response followup as corrected by ICM. That concept is called the "one-number census."

Data for ICM will be collected in a special sample survey done after the census. The ICM is in many ways like programs that were offered for adjusting the census in 1980 and in 1990. Based on experience with these programs, we believe there is a substantial risk that ICM will degrade rather than improve the quality of census data. The ICM is a complex, multi-step process, intended to fix relatively small errors in the census proper; many somewhat arbitrary technical decisions are involved, which may turn out to have large effects. The ICM process depends on a number of statistical assumptions that have been tested in the past and found wanting. Moreover, there are many opportunities for error in ICM fieldwork, and there may be substantial difficulty in detecting or correcting such errors. As a result, ICM may add more error than it removes. (Details, and references, will be found below.)

The 1990 analog of ICM was the "Post Enumeration Survey," or PES. On 15 July 1991, the PES estimated a census undercount of 5 million people. However, 50% to 80% of this estimated undercount resulted from PES errors rather than census errors: thus, most of the estimated undercount reflected errors in the estimation process not errors in the census. The major weak points in current census adjustment techniques can be summarized as follows.

1) Many somewhat arbitrary technical decisions will have to be made. Some of these may have a substantial influence on the results.
2) Many of the ICM's statistical assumptions are rather shaky.
3) There is ample opportunity for error.
4) The errors are hard to detect.

One example serves to illustrate the last point. Some time after the Bureau recommended adjusting the 1990 census, it discovered a mistake in its analysis of the PES data; this error added about a million people to the estimated undercount, and had the effect (along with some other relatively minor factors) of shifting a congressional seat from Pennsylvania to Arizona.

The ICM seems to be just as vulnerable to error as the PES. Of course, the situation for 2000 will remain somewhat unclear until the data are collected. If past experience is any guide, however, it will be nearly impossible to demonstrate—to any reasonable degree of certainty—that ICM improves on the census for purposes of apportionment or redistricting. The basic issue, of course, is that the errors in the census are relatively small: in 1980 and 1990, net undercounts were in the range 1%–2%. To improve on the census, the ICM would need to achieve error rates well below 1%. That sort of accuracy does not seem to be within the realm of current survey techniques.

In the balance of this paper, we discuss (i) the ICM, (ii) sampling for non-response, (iii) the interaction of these two programs, and (iv) our conclusions. There are some notes on technical points, and then a bibliography of selected articles on census adjustment. The Bureau is still refining its plans for 2000; descriptions of the design are provisional for that reason among others.

Integrated Coverage Measurement

The ICM will be based on a sample of 25,000 blocks, containing 750,000 housing units and 1.7 million people. An independent listing is made of the housing units in the sample blocks, and persons in these units are interviewed after the census is complete. ICM records are then matched against the census.[1]

In most cases, a match validates both the census record and the ICM record. An ICM record that does not match to the census may correspond to a "gross omission," that is, a person who should have been counted in the census but was missed. Conversely, a census record that does not match to the ICM may correspond to an "erroneous enumeration," that is, a person who was counted in the census in error.

An erroneous enumeration may be a person who was counted twice in the census—perhaps because he sent in two forms. Another person may be counted correctly but assigned to the wrong unit of geography: she would be a gross omission in one place and an erroneous enumeration in the other. Of course, some persons are missed both by the census and by ICM; their number is estimated by statistical modeling. However, these models are generally thought to be systematically in error. This error is called "correlation bias." The impact of correlation bias seems to vary from place to place in the country.[2]

Fieldwork may be done to resolve the "status" of unmatched cases, that is, to decide whether the error should be charged against the census or the ICM. However, even after fieldwork is complete, some cases remain unresolved. Such cases are handled by statistical models that "impute" (or estimate) the missing data. The number of unresolved cases may be relatively small, but it is likely to be large enough to have an appreciable influence on the final results. Statistical models used to make these sorts of imputations have many somewhat arbitrary elements, and should therefore be scrutinized with great care; experience from the past is not encouraging.[3]

Movers—people who change address between the time of the census and the time of the ICM interviews—represent a major complication to the ICM. Unless persons are correctly identified as movers or non-movers, they cannot be matched; and the identification depends on getting accurate information from respondents on where they were living at census time. The number of movers is relatively small, but they are a large factor in the adjustment equation. More generally, matching records between the ICM and the census becomes problematic if respondents give inaccurate information to the census, or the ICM, or both. Thus, even cases that are resolved though ICM fieldwork may be resolved incorrectly.[4]

We turn now to estimation. The Bureau divides the population into "post strata" defined by demographic and geographic characteristics. One post stratum might be Hispanic male renters age 30–49 in California. Persons in the ICM sample are assigned to post strata on the basis of the fieldwork. Moreover, each person in the ICM sample is assigned a "sample weight." If the Bureau sampled 1 person in 100, each sample person would stand for 100 in the population, and have a sample weight of 100. The actual sampling plan is more complex, so different people have different weights.[5]

To estimate the total number of gross omissions in a post stratum, one simply adds the weights of all ICM respondents who were identified as (i) gross omissions and (ii) in the relevant post stratum. To a first approximation, the estimated undercount in a post stratum is the difference between the estimated numbers of gross omissions and erroneous enumerations. One then computes an "adjustment factor"; when multiplied by this factor, the census count for a post stratum equals the estimated true count. Typically, adjustment factors exceed 1; most post strata are estimated to have undercounts. However, many adjustment factors are less than 1; these post strata are estimated to have been overcounted by the census.[6]

We now consider the process for adjusting small areas such as blocks, cities, or states. Take any particular block by way of example. Each post stratum has some number of persons counted by the census as living in that block. (The number may be zero.) The census number is multiplied by the adjustment factor for the post stratum; the process is repeated for all post strata; and the adjusted count for the block is obtained by adding the products. Finally, the count for any larger area is obtained by adding the counts for the blocks within the area.[7]

The adjustment process assumes that undercount rates are constant within each post stratum across all geographical units. This "homogeneity assumption" is quite implausible, and was strongly contradicted by data from the 1990 census. Ordinarily, samples are used to extrapolate upwards, from the part to the whole. In census adjustment, samples are used to extrapolate sideways, from 25,000 sample blocks to each and every one of 5 million inhabited blocks in the United States. That is where the homogeneity assumption comes into play.[8]

Sampling for non-response

Non-response followup in the census has in the past been done on a 100% basis. In the bulk of the country, forms are mailed out to all identified housing units. If there is no response from a housing unit, interviewers come knocking on the door. In 2000, however, the Bureau plans to follow up only a sample of non-respondents. As we understand the current idea, non-respondents will be sampled within each tract. If, for instance, a tract has 2,000 housing units and 1,200 return their census form by mail, there are 800 non-responding units. The Bureau would then sample 600 out of these 800 units and send interviewers only to the sample units. Additional housing units would be imputed into the census using statistical models based on sample responses. In particular, block-level data must be somewhat problematic in Census 2000.[9]

There is some body of opinion that sampling improves accuracy since interviewers can be better trained and supervised. Given the proposed sampling rates, this advantage cannot be substantial. Sampling seems inherently more complex than a census, and past experience shows that sample surveys have worse coverage than the census, and worse differential coverage. Of course, sampling for non-response in Census 2000 may not be directly comparable to past surveys.[10]

How does sampling for non-response interact with the ICM?

The essential task of the ICM is to match records against the census. That conflicts with sampling for non-response because ICM respondents may be in households that did not return a census form and were not selected for followup. To solve this problem, the Bureau proposes to do 100% followup in the ICM sample blocks. That involves two kinds of sampling and three kinds of fieldwork:

The two kinds of sampling:

1) Sampling in the ICM.
2) Sample-based followup for census non-response.

The three kinds of fieldwork:

1) The ICM.
2) 100% followup for census non-response in the ICM sample blocks.
3) Sample followup for census non-response in the rest of the country.

Other assumptions must be made here: 1) census coverage will be the same whether non-response followup is done on a sample basis or a 100% basis; and, 2) residents of the ICM sample blocks do not change their behavior as a result of being interviewed more than once. Failure of these assumptions may be termed "contamination error." The magnitude of contamination error is unknown.[11]

Summary and conclusion

We see the following crucial difficulties in the ICM.

1) Many somewhat arbitrary technical decisions will be made. Some of these may have a substantial influence on the results.[12]
2) Many of the statistical assumptions are rather shaky. The homogeneity assumption is an example. The assumptions behind the models used to impute missing data can also be mentioned.
3) There is ample opportunity for error, especially when respondents give incomplete or inaccurate data to census or the ICM.
4) The errors are hard to detect.

Any effort to demonstrate that ICM improves on the census must reckon with the sources of error discussed in this paper.

Of course, the census has errors of its own. However, comparing the magnitude of census errors to ICM errors is fraught with difficulty. The ICM is quite a complicated operation, and there is a strong likelihood of significant—but undetected—error. In July 1991, proposed adjustments were predicated on the idea that the census missed (net) 5 million people out of 250 million. That figure of 5 million is only an estimate, derived from the "Post Enumeration Survey" or PES. Later research shows that the bulk of this estimate derived from errors in the PES rather than errors in the census. If the ICM—like the PES before it—puts in more error than it takes out, Census 2000 will be at considerable risk.[13]

Endnotes

1. More technically, the ICM sample involves 25,000 block "clusters," each cluster amounting to one or more blocks. (A "block" is the minimal unit of census geography; there are about 7 million blocks in U.S., of which about 5 million are inhabited.) In 1990, adjustment was to be based on the PES. Generally, the proposal for 2000 is very close to the proposal for 1990, with exceptions noted below. Among other things, the sample for 2000 is about five times larger than the sample for 1990. This will reduce sampling error, but may increase non-sampling error.

"Sampling error" is due to the luck of the draw in choosing the sample; "non-sampling error" is a catch-all term for other kinds of error, for instance, clerical errors in processing data. Generally, sampling error goes down with bigger samples; however, increasing the sample size generally does not reduce non-sampling error. Indeed, as sample size increases, it becomes more difficult to recruit, train, and manage personnel. Furthermore, complexity increases the likelihood of non-sampling error. Adjustment programs like the ICM are most vulnerable to non-sampling error; and it is non-sampling error that is hard to quantify.

2. In July 1991, results from the PES suggested there were about 19 million gross omissions and 13 million erroneous enumerations; 2 million persons were imputed into the census. A first approximation to the estimated undercount is $19 - 13 - 2 = 4$ million; another million persons are added by statistical modeling.

Estimates of correlation bias are necessarily rather imprecise. Some observers feel that correlation bias only strengthens the case for adjustment. However, that is simplistic, because the size of the bias seems to vary from one area to another; and putting people in the wrong places can lead to rather perverse adjustments. *Jurimetrics*, vol. 34 (1993) pp. 75, 110–13. *Statistical Science*, vol. 9 (1994) pp. 514, 533–34

3. In 1990, there were about 4 million unresolved cases in the P-sample and a similar number in the E-sample. (The "P-sample" consists of persons found by the PES in the sample blocks; the "E-sample" consists of persons found by the census in these same blocks.) These are big numbers relative to a total estimated undercount of 5 million. For a discussion of the 1990 imputation model from various perspectives, see *Statistical Science*, vol. 9 (1994) pp. 469–70, 489, 531, 535–36. Errors in the 1990 adjustment are discussed in the same reference; also see note 13 below.

4. In 1990, there were 19 million movers, who contributed about 5 million to the estimate of gross omissions. Movers are a major part of the adjustment problem, not a minor part. In 2000, the ICM must identify both inmovers and outmovers, and try to match outmovers.

People who move into an ICM sample block would have to be matched to the census at their census-day address, which is likely to be in a non-ICM block—where census followup was done on a sample basis. Consequently, matching inmovers would be quite problematic. That is why outmovers must be considered. In this respect, the ICM design will be harder to implement than the design for 1990. On the difficulty of determining census-day addresses, see *Statistical Science*, vol. 9 (1994) p. 471–2.

One of the paradoxes of adjustment is that matching is hardest to do exactly in the geographical areas that are hardest to count. Therefore, errors introduced by adjustment are far from evenly distributed. *Statistical Science*, vol. 1 (1986) p.26.

5. Use of sample weights to make estimates is quite standard; of course, mistakes in fieldwork are magnified by the weights. The weights in the ICM will average around 150, but considerable variation may be expected.

In July 1991, the Bureau proposed to use 1,392 post strata. To adjust the census as a base for the post-censal estimates, the Bureau later proposed 357 post strata. *Statistical Science*, vol. 9 (1994) pp. 478–79. The Bureau has not finalized its proposals for 2000. However, post strata will not cross state lines; thus, each state would be adjusted only using data collected within that state. This is an improvement over 1990, because homogeneity of post strata that cross state lines is not assumed. The choice of post stratification is somewhat subjective and may have a substantial influence on the results: for data on 1990, see the memorandum by Henry Woltman, dated 15 July 1991, "Estimated state level adjusted counts based on revised state groups," Bureau of the Census, Washington, D.C.

To illustrate the idea of post stratification more specifically, Census 2000 might have

$$7 \times 2 \times 4 = 56$$

post strata in each state, based on 7 age-sex groups, 2 tenure groups, and 4 race-ethnicity groups.

The 7 age-sex groups:

Age 0–17, both sexes
Men 18–29
Men 30–49
Men 50+
Women 18–29
Women 30–49
Women 50+

The 2 tenure groups:

Owns or rents housing unit

The 4 race-ethnicity groups:

Asian
Hispanic
Non-Hispanic black
Non-Hispanic white

For example, one post stratum would consist of Asian children age 0–17 living in their own homes in Hawaii; another would consist of non-Hispanic white women age 50+ living in rental units in Maine; and so forth. This classification is illustrative only.

6. In the 1991 post stratification, about 70% of the adjustment factors exceeded 1 and 30% were below 1.

7. We are ignoring a number of complications. For instance, adjustment may lead to fractional numbers; the Bureau will undoubtedly use some form of controlled rounding to get whole numbers. In 2000, people may be added in family groups or households, which will necessitate further modeling. When post strata are overcounted, some real people need to be flagged out of the census, that is, not counted in the tabulations. This raises questions of equity.

8. *Statistical Science*, vol. 9 (1994) pp. 476ff, 514–15, and 529–30.

9. A "tract" is a unit of census geography, containing on average something like 100 blocks, 2,000 housing units, and 4,000 people. The idea is get responses from 90% of the housing units in each tract. In the hypothetical, 1,200 + 600 = 1,800, which is 90% of the assumed total of 2,000. Sample data may for instance suggest adding a certain number of households of a given type; some households would be selected at random from enumerated households of that type and added to the census.

Block-level data must be problematic in Census 2000, for the following reasons. If direct estimates are used (estimating the population of each block on the basis of its own data) sample sizes will be small for non-response followup; and the vast majority of blocks will have no ICM data at all. Thus, the Bureau will have to pool data for many blocks, assuming similarity of responses—within types of units—across blocks. In effect, this is a "homogeneity assumption" of the kind discussed above.

Additional complications are created by broadcasting census questionnaires, or allowing people to be counted based on telephone calls they initiate. Such measures will increase coverage, but are also likely to increase duplicate enumerations. Based on experience in 1990, we are not convinced that duplicates can be reliably detected by computer software.

In 2000, non-response followup may be truncated in order to complete the ICM on time. If so, census coverage and data quality will be adversely affected; the effect could be quite significant.

10. As noted above, each person in a sample gets a "sample weight"; this may, for instance, be the inverse of the probability of drawing that person into the sample. Based on sample data, one can then estimate the total number of persons in the population by adding up the weights. "Coverage" is the ratio of this estimated number to the true number. Likewise, the coverage of the census is the ratio of the census population to its (estimated) true value.

Generally, the Bureau will sample people (for instance) by first drawing a sample of blocks. All people living in the selected blocks are then interviewed. Blocks are relatively easy to sample because their number is fixed and known, and sample weights can be determined from the sample design. This process is standard, objective, and not easy to manipulate.

The Current Population Survey is a well-established, well-run sample survey; still, it reaches only about 95% of the population—as measured by the census. Black and Hispanic sub-populations have noticeably lower coverage ratios. See, for instance, *Money Income of Households, Families and Persons in the United States*, Series P-60, Bureau of the Census.

In 1990, as noted above, the analog of the ICM was the PES. That sample survey was done following the census, and it had about 98% of the coverage of the census. Of course, the census data were in some cases problematic; so were the PES data. (The calculation weights the P-sample and E-sample to national totals; see note 3.)

11. On "coverage," see note 10. The Oakland test census provides some evidence showing that contamination error is modest. The real census, however, may be somewhat different from a test census. Also see p. 115 of Steffey and Bradburn (1994), *Counting People in the Information Age*, National Academy Press.

12. In 1990, rather esoteric technical decisions made fairly late in the day almost doubled the estimated undercount, from 1.2% to 2.1%. These decisions involve what is called "pre-smoothing" and "benchmarking"; see *Statistical Science*, vol. 9 (1994) pp. 531-32; also see *Journal of the American Statistical Association*, vol. 88 (1993) p. 1051; for details, see *Evaluation Review* vol. 17 (1993) pp. 371–443, especially pp. 387–90.

Here is another example: a decision to change the weights given to two of the 5,000 block clusters in the Post-Enumeration Survey affected some 500,000 people in the counts. Memorandum from Howard Hogan to John Thompson, dated 18 June 1991, "Downweighting outlier small blocks," Bureau of the Census, Washington, D.C. Also see *Journal of the American Statistical Association*, vol. 88 (1993) p. 1050.

13. For a discussion of the magnitude of the errors in the census and the adjustment, see *Statistical Science*, vol. 9 (1994) pp. 458–537; there is a brief summary in *Evaluation Review*, vol. 20 (1996) p. 367. Basically, measured errors in the PES amount to 57% of the total estimated undercount on the Bureau's own figures, compared to 80% as computed by outside observers. In 1990, there was a fairly elaborate taxonomy for what we are calling "matching error"; see *Statistical Science*, vol. 9 (1994) p. 471.

We give one example of a mistake that was found only after the Bureau recommended adjustment to the Secretary of Commerce. As it turns out, the PES picked up some number of respondents who moved into sample blocks after census day but matched to the census. Such persons should have been classified as erroneous enumerations, but were—in parts of the country—classified as correct enumerations. This "computer coding error" added about a million people to the estimated undercount, with the effect—noted above—of shifting a congressional seat from Pennsylvania to Arizona. *Statistical Science*, vol. 9 (1994) pp. 471, 498. More exactly, the original adjustment shifts two seats, while the 357 post stratification shifts one.

In 1990, the Bureau made a major effort to assess the quality of the PES data, and the evaluation data were made public. In 2000, the planned scale of evaluation work is much reduced, making assessments of the ICM even more difficult.

14. Some of the 1990 numbers in previous notes were computed from the "Advisory Use File" provided by the Bureau to outside observers.

According to present plans, the ICM will be based on the "Dual System Estimator"; CensusPlus will not be used. The formula for the Dual System Estimator changed from 1980 to 1990, and seems to have changed again for 2000, indicating some degree of arbitrariness. Details on the current formula are given in a memorandum dated 1 April 1996, by E. Ann Vacca, Mary Mulry, and Ruth Ann Killion, "1995 Census Test Results," Bureau of the Census, Washington, D.C. For the 1980 formula, see *Science*, vol. 252 (1991) p. 1235. For the 1990 formula, or definitions of the DSE and CensusPlus, see *Evaluation Review*, vol. 20 (1996) pp.361–63.

Affiliations. M.L. Eaton, Department of Theoretical Statistics, University of Minnesota, Minneapolis; D.A. Freedman, K. Wachter, Department of Statistics, U.C. Berkeley; S.P. Klein, RAND Corporation, Santa Monica; R.A. Olshen, Division of Biostatistics, Stanford; D. Ylvisaker, Department of Statistics, UCLA.

Selected bibliography

Breiman, L., T.R. Belin and J.E. Rolph, D. Freedman and K. Wachter, and others (1994). Three papers on the census adjustment. *Statistical Science* 9: 458–537. Reviews arguments for and against adjusting the 1990 census; both sides are represented.

Ericksen, E.P. and J.B. Kadane (1985). Estimating the population in a census year. *Journal of the American Statistical Association* 80: 98-131. Favors adjusting the 1980 census; with discussion.

Fay, R.E., J.S. Passel, J.G. Robinson, and C.D. Cowan (1988). *The Coverage of the Population in the 1980 Census.* U.S. Department of Commerce, Government Printing Office. Generally critical of adjusting the 1980 census; written by senior Census Bureau personnel.

Fay, R.E. and J.H. Thompson (1993). The 1990 Post Enumeration Survey: statistical lessons, in hindsight. *Proceedings, Bureau of the Census Annual Research Conference*. Written by senior Census Bureau personnel; acknowledges many of the difficulties in adjusting the 1990 census.

Freedman, D. (1991). Adjusting the 1990 census. *Science* 252: 1233–36. Critical of adjustment.

Freedman, D. and W. Navidi (1986). Regression models for adjusting the 1980 census. *Statistical Science* 1: 1–39. Critical of adjustment; with discussion.

Freedman, D. and K. Wachter (1996). Planning for the census in the year 2000. *Evaluation Review* 20: 355–77. Discusses plans for census 2000; reviews evidence from 1990; summarizes the Supreme Court decision in New York v. Commerce.

Freedman, D., K. Wachter, D. Coster, R. Cutler, and S. Klein (1993). Adjusting the census of 1990: the smoothing model. *Evaluation Review* 17: 371–443. The "smoothing model" was proposed to reduce sampling error in the 1990 census adjustment; this paper is a critique; gives background on "pre-smoothing" and "benchmarking."

Freedman, D., K. Wachter, R. Cutler, and S. Klein (1994). Adjusting the census of 1990: loss functions. *Evaluation Review* 18: 243–280. "Loss function analysis" was used in an effort to demonstrate that adjustment would improve the accuracy of the 1990 census; this paper is a critique.

Kaye, D., editor (1993). Papers on census adjustment. *Jurimetrics* 34: 5-9-115. Reviews arguments for and against adjusting the 1990 census; both sides are represented.

Schenker, N., editor (1993). Special section on the 1990 undercount. *Journal of the American Statistical Association* 88: 1044–1166. Generally supportive of adjustment.

Singh, M.P., editor (1992). Special section on census undercount measurement methods and issues. *Survey Methodology* 18: 1–154. The focus is on 1980, but there is some discussion of 1990; both sides are represented.

Steffey, D.L. and N.M. Bradburn, editors (1994). *Counting People in the Information Age*. National Academy Press. The case for adjusting census 2000.

Wolter, K. (1991). Accounting for America's uncounted and miscounted. *Science* 253: 12–15. Favors adjusting the census of 1990.