# Sampling and Census 2000–Methodological Issues: Report to the Donner Foundation


Beth Osborne Daponte
University Center for Social and Urban Research
University of Pittsburgh
Pittsburgh, PA 15260

Stephen E. Fienberg
Department of Statistics
Carnegie Mellon University
Pittsburgh, PA 15213-3890

Joseph B. Kadane
Department of Statistics
Carnegie Mellon University
Pittsburgh, PA 15213-3890

Duane L. Steffey
Department of Mathematical and Computer Sciences
San Diego State University
San Diego, CA 92182-7720


July 22, 1999

# Contents

## 1 Background and Overview

A national census is an extraordinarily complex and difficult undertaking, particularly in a nation as populous and diverse as the United States. Recent American censuses have been the focus of controversy concerning the accuracy of their population estimates and the viability of using sampling and statistical estimation to improve the estimates obtained via traditional methods. With Census Day 2000 little more than a year away, this report attempts to provide a broad and useful perspective on current census-related activities and to propose a framework for evaluating alternative census methods in hopes of resolving ongoing technical disputes.

Our primary mission is to discuss key methodological issues in depth and to examine the implications of the 1990 census experience for Census 2000. Such evaluations, however, lead us to consider which criteria are appropriate in judging between more traditional and more innovative census designs. We also consider the relevant experience of countries with similar cultural traditions—namely, Canada, England, and Australia. First, however, we undertake a brief historical discussion of the purposes and recent criticisms of the decennial census, as well as an overview of the key elements of modern census operations, to provide some background for the remainder of the report.

## 2 Historical Background

The decennial census of the American population was originally mandated in the Constitution to provide a basis for reapportioning the House of Representatives and for levying direct taxes. Census population estimates are also used to revise election districts for federal, state, and local offices. Federal grant-in-aid programs often rely on census population as a key factor in determining how to allocate funds among states and smaller political jurisdictions.

Over the past half-century, the increasing political salience of census population estimates and the concurrent development and refinement of credible methods of measuring census coverage of the population have led to greater scrutiny of the census process and to strong criticisms of its accuracy and cost. It is now widely recognized that the traditional census process undercounts the true population and that members of certain groups defined by demographic and geographic characteristics—e.g., racial and ethnic minorities, renters, children and young adults, urban dwellers—are more likely to be missed than members of other groups. This phenomenon is referred to as the differential undercount. Concern has also been expressed about the dramatically increasing costs of census taking. To count a household cost about $11 in 1970 (in constant 1990 dollars); by 1990, unit costs had risen to $25 per household (Edmonston and Schultze, [24]). We elaborate on possible reasons for the cost increases in a subsequent section.

Concerns about rising costs and the persistent problem of differential undercounting prompted widespread calls by members of Congress and others for fundamental census reform. After the 1990 census, the Census Bureau and the Department of Commerce in the Bush administration initiated an ambitious census research and development program to investigate alternative census methods for use in the 2000 census. The proposed 2000 census design that emerged from this program includes many innovative features, of which the most prominent is the integration of sample-based methods for measuring census coverage with other census operations to produce official population estimates. This feature, known as integrated coverage measurement (ICM), won approval from three National Academy of Sciences panels formed to study census issues. One panel, in concluding that ICM represented a feasible and cost-effective solution to the dual problems of accuracy and cost, summarized its position as follows (see Steffey and Bradburn, [66], p. 4):

Recommendation 4.2: Differential undercount cannot be reduced to acceptable levels at acceptable costs without the use of integrated coverage measurement and the statistical methods associated with it. We endorse the use of integrated coverage measurement as an essential part of census-taking in the 2000 census.

Despite enjoying broad support from both the scientific community and census data users, however, the Census Bureau's 2000 census design has not met with approval in Congress. The House of Representatives filed suit to prevent implementation of the proposed uses of statistical sampling, and the Supreme Court recently decided that ambiguous clauses in existing federal legislation prohibited the use of sampling in producing estimates to be used in congressional reapportionment [72].[1]

Setting aside legal issues, we take up the question of the relative scientific merits of the so-called traditional census—the product of many previous innovations in census methods (see, e.g., Goldfield, [40]; Anderson, [2], and Anderson and Fienberg [5])—and the 2000 census design proposed by the Census Bureau.

## 3 Elements of the Census Process

Census taking in the United States involves four key steps: (1) development of a national address list (or frame); (2) an initial procedure for collecting responses from households that can be linked to the address frame; (3) a follow-up procedure to obtain responses from households not included in the initial procedure; (4) an assessment process to estimate the size of the population not covered by the initial or follow-up counting procedure.

Each step in the census process is susceptible to various sources of error. Housing units can be omitted from the national address frame. Occupied housing units can be incorrectly listed as vacant during the initial counting procedure. Confusion about the census residence rules—which determine an individual's membership status in a household—may result in persons being listed on multiple census forms or being omitted from any census form. Obtaining accurate information from hostile or evasive households during the follow-up procedure is an often daunting and error-prone task. The assessment process, involving an independent survey after other census operations have concluded, introduces sampling error and additional sources of bias.

Designing a census thus amounts to deciding which methods should be applied at each step of the process and how resources should be apportioned among the census activities—presumably, with the goal of obtaining population estimates of acceptable accuracy in the most cost-effective manner. Making informed choices among alternative designs and developing a consensus in support of the final census design therefore would seem to require articulation and agreement on the standards against which alternative designs should be judged.

[1] Much of the content of the present report was prepared before the Supreme Court issued its ruling prohibiting the use of sampling for reapportionment purposes in January, 1999. We have kept the focus of the report on the original proposal for Census 2000 than included *both* sampling for non-response followup *and* integrated coverage measurement for adjustment, nonetheless because that was the focus requested originally by the Donner Foundation. Moreover, the revisions in the Census 2000 plan are still in some flux and thus it is difficult to discuss the revised plan in detail. The Supreme Court ruling does have a number of clear ramifications and we discuss several of these later in the report.

## 4 Criteria for Evaluating Census Accuracy

There are three basic types of error that enter the census process. Persons can be counted more than once, they can be counted once but assigned to the wrong geographic location, or they can be missed entirely by census operations. Duplications can occur, for example, because an individual owns two homes—perhaps a primary residence and a vacation home—and returns census forms sent to both locations. A child of divorced parents might be listed on the census questionnaire completed by each parent, or a college student might be counted both at campus and at a parent's home. A duplication error results in an erroneous enumeration.

Exclusions can occur, for example, because a housing unit is missed in constructing the address list and never receives a census form. A household may fail both to return the mail questionnaire and to respond to follow-up attempts at enumeration. An individual may be inadvertently omitted from the household roster, because the household respondent misunderstands the rules determining who should be listed on the census form. An exclusion error results in an omission. Research from the 1990 census suggests that missed housing units, missed households, and missed persons within households contribute on the order of 70% the total number of omissions (e.g., see Anderson and Fienberg [4], and Killion [46]).

A rostering or geocoding error that incorrectly assigns an individual to a geographic location (e.g., a census block) results in two errors: an erroneous enumeration in the assigned block and an omission from the block in which the individual actually resides.

In an operation on the enormous scale of the decennial census, mistakes inevitably creep into the process, even though quality control procedures are implemented to detect and correct them. In assessing the accuracy of alternative census designs, it is important to gain an understanding of the various types of census error, the magnitudes of these errors, and the geographical variation in error rates.

One commonly cited measure of accuracy is the net national undercount, which is obtained by subtracting the number of erroneous enumerations from the number of omissions. In the 1990 census, the net national undercount was estimated at approximately 4 million persons—corresponding to a net error rate of less than 2% (e.g., the discussion of 1990 in the books by Bryant and Dunn [13], and by Choldin [15]).

The net national undercount, however, is an inadequate measure of census accuracy, because it masks several important features. First, the 1990 net undercount estimate of 1.6% was obtained by subtracting an estimated 16 million erroneous enumerations from an estimated 20 million omissions. The sum of these two errors yields a gross error rate of about 8% (see, e.g., Ericksen [25], Ericksen, Fienberg, and Kadane, [26], and U.S. General Accounting Office [70]). Later calculations at the Census Bureau have showed that the gross error rate at the census block level is also of this order of magnitude [69] and so there is little cancellation of errors until one begins to aggregate. Moreover, the causes of erroneous enumerations and omissions suggest that these errors are unlikely to cancel one another at subnational levels of geography. Thus, a census design with small net error but large gross error can produce state population estimates, for example, that vary substantially in their degree of accuracy.

The existence of differential undercounting in the census was initially discovered during World War II when an unexpectedly large number of black males registered for the military draft (Anderson and Fienberg, [5], Price [54]). The differential undercount of demographic groups has persisted over the past half-century, even as the net undercount has been reduced. Because members of undercounted demographic groups are more likely to reside in some geographic areas than in others, the differential undercount of demographic groups implies a differential undercount of geographic areas. The presence of such undercounting in population estimates is objectionable partly because it seems to violate a basic notion of fairness. It has been argued (e.g., Mitroff, Mason, and Barabba, [47]) that for the census not to count uncooperative respondents is fair. The damage resulting from not being counted, however, falls not just on the persons not counted, but on their neighbors as well, who are underrepresented politically and denied their fair share of federal funds as a result of the omission.

Geographic areas with relatively high undercount rates may also be shortchanged in political representation and revenue allocation. For these purposes, the most relevant consideration is not the accuracy of population estimates but rather their accuracy relative to estimates of other areas at the same level of geography. For example, in considering census accuracy for reapportionment, the accuracy of state population shares seems the most appropriate measure.

# 5 Traditional Methods, Post-Enumeration Surveys, and the 1990 Experience

The concept of "census undercounts" is a function of the development of modern probability sampling methods and their use in statistical methodology. Census officials from George Washington and Thomas Jefferson to the present have been aware of the extraordinary difficulties of counting a dynamic, diverse, and mobile population, and that much of the concern for census accuracy has fueled improvements in census methods over the years. Until the 1940s, census officials had hunches, but few tools to measure accuracy, and certainly no mechanism to "correct" for undercounts, overcounts, or miscounts, beyond doing the enumeration over again. Beginning in the 1940s, statisticians and demographers worked to develop methods of measuring census accuracy, using the techniques of demographic analysis and dual systems estimation. The Census Bureau began evaluating the population coverage of enumeration procedures in the 1950 census. Sample "post enumeration survey," statistical inference from sample to national rates, and comparison to the vital registration system through demographic analysis replaced the hunches and laments about census error with increasingly reliable point estimates of error and intervals of confidence about those estimates.

Anderson and Fienberg [5] document the rise in the uses of sampling in the census context, beginning with the sample long form in 1940 and the uses of the PES in 1950, as well as the National Vacancy Check Survey and the Postenumeration Post Office Check Survey in 1970, which led to the inclusion of over 1.5 million people in the census on the basis of generalizing from sample data. From 1950 to 1980, significant innovations–including the mail questionnaire, optical scanning, and matching technologies–were developed, tested, and adopted as components of the "traditional" census operation (e.g., see Anderson and Fienberg [5], Goldfield [40]). During this period, each census also achieved a reduction in the net national undercount from that of the preceding census.

While sampling has long been understood to be an efficient mode of data collection for a variety of purposes, its use as a methodological adjunct to supposedly traditional census enumeration methods awaited the demands placed on the census by the events of the 1930s as well as the crucial developments of clustering and multi-stage sampling due to Neyman and others. These technical developments made it possible to begin to fit the structure of sample designs with the structure of census operations, a theme that has continued to the present plans for the 2000 census. The overriding justification for the use of sampling methods, beginning in 1940 and moving forward to the present, has been to improve the accuracy and the quality of census counts and other data. Part of the justification of the new sample long form in 1940 was the reduction of respondent burden.

The purpose of the 1950 PES was coverage evaluation, with the goal of identifying areas to focus on for future administrative efforts to increase accuracy. At about the same time the Census Bureau expended considerable resources to develop a statistical model for census error and began an organized effort to assess key components. The measure of correlated response error due to interviews led, for example, to the trial in 1960 of the mail-out, mail back approach to enumeration which worked to eliminate the role of the interviewers. The bureau's strategy over the intervening years was to approach the problem of accuracy in a piecemeal fashion. From historical demographic analysis, the net national undercount is now estimated to have declined from 5.4% in the 1940 census to 1.2% in the 1980 census (Robinson, et al., [58]).

There were some signs by 1980, however, that census operations might be approaching limits in the ability to enumerate the population accurately using only traditional methods. First, the rate at which households returned a completed census form by mail dropped by 13% between 1970 and 1990 (Edmonston and Schultze, [24]). Second, the extensive array of auxiliary programs intended to improve population coverage was criticized as being costly, inefficient, and prone to introducing erroneous enumerations into the census process, thus inflating the number of gross errors (Citro and Cohen, [16]). Finally, although the net undercount continued to decline, differentials in coverage between demographic groups persisted. Blacks, for example, were undercounted at a rate that averaged 3.8% higher than the rate for non-Blacks and remained remarkably consistent from 1940 to 1980 (Robinson, et al., [58]).

By the time of the 1980 census, increasing recognition of the persistent differential undercount and its political salience led undercounted cities and states to file lawsuits calling for sample-based adjustment of the official census estimates. Although a postenumeration program was carried out in 1980, the Census Bureau opposed adjustment on grounds that the methods used were not adequate for purposes other than evaluation. The Census Bureau's position was supported by Commerce Department officials in both the Carter and Reagan administrations. The federal courts supported the right of the Commerce Department to make such a determination.

With the recognition during the 1980s that most coverage improvement efforts had failed to address the differential undercount, attention focused anew on using the PES to correct the raw enumeration results. The Census Bureau initiated an ambitious research program to design a postenumeration survey that could be used to adjust 1990 census enumeration results, if necessary, to correct for differential undercoverage of demographic groups and geographic areas. In early 1987 the Census Bureau announced plans for a 300,000-household survey in 1990. Later that year, however, Commerce Department officials halted development of the PES program. New York City filed a lawsuit in 1988, and the Commerce Department consented in the summer of 1989 to reinstate the survey but with a sample size of only 150,000 households. (It turned out ultimately to be of size approximately 165,000 households.) Work on the 1990 PES resumed after a hiatus of almost two years and with a substantial reduction in survey precision attributable to the smaller sample size. The Secretary of Commerce agreed to consider *de novo* the question of adjusting the 1990 census and to reach a decision by July 1991. Thus, in the end, the 1990 census design was a compromise, a traditional census complete with coverage improvement programs and a PES for possible use in correcting the count.

After an extensive evaluation of the 1990 PES, the Director of the Census Bureau and a majority of members of the Census Bureau's Undercount Steering Committee recommended adjustment to improve the accuracy of the 1990 census estimates. In July 1991, however, the Secretary of Commerce acted against those recommendations and decided not to adjust the 1990 census. In the ensuing litigation, the Secretary's decision was upheld by the district court, overturned by the appellate court, and ultimately reaffirmed by the Supreme Court in 1996, not as being correct, but as being within his prerogatives.

Although statisticians in the Census Bureau and many in the academic community supported sample-based adjustment of the 1990 census, there were significant technical criticisms of the 1990 PES methodology. Appendix A of this report examines these criticisms in detail, rebutting some of them and agreeing with others. Ultimately, we conclude that, when judged using neutral criteria, the available evidence about the 1990 census favors adjustment.

Despite the generally acknowledged success of the 1990 adjustment as an improvement over the raw enumeration counts, the small sample size had two major impacts on accuracy. (1) The level of sampling error was a serious component of the overall census error; and (2) the data at low levels of geography were sufficiently sparse as to require smoothing of the adjustment results across state lines.

The public discourse on the 1990 census and adjustment has conflated 'sample size' problems with a more general condemnation of sampling in the context of censustaking, and has led to mythmaking that 'sampling' itself is a suspect enterprise. Wachter and Freedman's February 1996 testimony before a Senate subcommittee, for example, depicts a census officialdom who have lost control of their methodology, and raises fears of numbers 'vulnerable' to 'statistical error.' In the current context of partisan dispute, this testimony effectively drowns out their more measured endorsements of the bureau as "the world's finest statistical agency" whose officials "need and deserve Congressional support." Statements regarding the vulnerability of the proposed methodology for 2000 to such errors continue to appear in the work of these and other critics (e.g., see Brown et al. [12], p. 6). For a response to some of these, see Appendix A of this report.

Similarly, the frequently repeated statement that "55% of the DSE national undercount estimate [in the 1990 census] is due to bad data or processing errors" obfuscates the technical discussion needed of how to remedy the differential undercount.[2] And this is not simply rhetoric on the part of the critics, as the error statement plays a very subtle role in one of the formal arguments set forth about the lack of robustness of the Bureau's 1990 loss function analysis (see Brown et al. [12]), arguments which are otherwise difficult to justify. For details on this point, again see Appendix A.

# 6 Dual System Estimation For Improving the Count

Much has been written about dual systems estimation (DSE) in the statistical literature, and the approach is widely accepted and used in a wide array of applications. Hogan [43] summarizes the operational version of the methods used in 1990, and he builds on methodology described earlier in Wolter [78] and Hogan and Wolter [44]. Anderson and Fienberg [4] provide a detailed tutorial on the how the method works in the context of the census. Thus we restrict ourselves here to a brief non-technical description.

The dual systems method is basically a human-population version of the well-known capture-recapture method for estimating the size of a population which is used in a variety of areas of science. For each block in the sample followup, the method takes the census results and the sample followup and combines them, thus directly adding those households and individuals that are found in the sample but not in the census. One way to represent the data for a dual systems approach is in terms of a set of $2 \times 2$ table of counts, one for each census block in a postenumeration survey (PES). The rows of the table correspond to being included or not in the census, and the columns correspond to being included in the second source. Then we can display the data for a given block as follows:

[2] Leo Breiman [10]. Other evaluations of the PES do not accept Breiman's characterization. See for example, Howard Hogan [43], Mary Mulry and Bruce Spencer [51], or the comments by Thomas Belin and John Rolph [7] [8], and Eugene Ericksen, Stephen Fienberg and Joseph B. Kadane[26], in the same issue of *Statistical Science*, as Breiman's article. Further, even were the statement a correct and proper interpretation of the known data, it would be effectively irrelevant to the usefulness of the DSE approach. The net national undercount, which is the difference between omissions and erroneous enumerations, is not why the Census Bureau carried out the PES and did DSE. The closer the estimates for omissions and erroneous enumerations come to balancing out, the closer the net undercount estimate is to zero, and the bigger the impact of "errors." If there was a zero net estimated undercount a single error would account for it all. Thus, the seemingly explosive statement may be of little importance when we come to understand the improved accuracy from using PES-adjusted data. See also Anderson and Fienberg [4] on this issue.

| | | PES | | |
|---|---|---|---|---|
| | | In | Out | Total |
| | In | $a$ | $b$ | $n_1$ |
| Census | Out | $c$ | $d$ | $N - n_1$ |
| | Total | $n_2$ | $N - n_2$ | $N$ |

(1)

We get to observe $n_1$ from the census and $n_2$ from PES. By matching the two sources we find $a$, and thus get $b$ and $c$ by subtraction. Finally, $d$ (and thus $N - n_1$, $N - n_2$, and $N$) is unknown and we estimate it. If being counted in the census is independent of being counted in the PES, then our estimate of the unobserved $d$:

$$\hat{d} = \frac{bc}{a} \tag{2}$$

and we add this estimate for the missing cell to the three observed counts to get our estimate of the population size, $N$:

$$\begin{aligned} \hat{N} &= a + b + c + \hat{d} \\ &= n_1 + c + \tfrac{bc}{a}. \end{aligned} \tag{3}$$

Because we only have a sample of census blocks we need to generalize to those blocks not in the PES and produce adjustment factors for them as well. This requires the use of some type of statistical model. In 1990, the sum of the $c$ cells when weighted up to represent the nation as a whole to the nation was 18.8 million while the sum of the $d$ cells was 1.5 million (See Cohen, White, and Rust [17]).

While conceptually simple statistically, the rationale for the method is often difficult for non-technical people to follow and even Chief Justice Rehnquist mis-described it in his 1996 opinion on the 1990 census adjustment controversy–mixing up the the third and fourth cells (with entries $c$ and $d$ in the $2 \times 2$ table above. (see [77]).

There are two additional complications that were the focus of debate in the statistical community over the methodology. First, before applying the usual capture-recapture formulae the Census Bureau does an adjustment to the block by block counts for erroneous enumerations in the census. Second, to improve the accuracy of the projection of the adjustments from the sample blocks to the rest of the nation, the Bureau used what was referred to as a smoothing or regression model. The method and an exchange of views about whether it actually provided demonstrable improvements over the raw census counts in 1990 can be found in a series of four articles in a 1993 issues of the journal *Jurimetrics*, written by four of the statistical experts who participated in the trial that followed the 1991 decision not to adjust (see [30], [61], [33], [73]). For a more up-to-date review of these issues see the report of the most recent National Academy of Sciences Panel on census methodology (Cohen, White, and Rust, [17]) which concludes that the PES in 1990 removed much of the substantial bias in the unadjusted census counts and that the residual unremoved bias is likely to have a limited effect on the estimated undercount. In other words, in 1990 the adjusted counts were more accurate than the unadjusted counts. We concur in this judgment.

## 7 Sampling, Statistical Estimation, and the 2000 Census Design

The experience of the 1990 census contained important lessons for planning the 2000 census. The dual-track strategy adopted in 1990 involved conducting a 1980-style census, a postenumeration survey, and extensive subsequent evaluation prior to reaching a decision about whether to adjust the 1980-style census. Such an approach proved unworkable, because the length of time required for a thorough comparative evaluation of the unadjusted and adjusted estimates extends beyond the legislatively mandated deadlines for reporting official population estimates. Also, for effective operational planning, it is crucial to reach consensus on key components of census design as early as possible. Development of the 1990 PES suffered considerably from the two-year hiatus in activity between 1987 and 1989.

In planning for the 2000 census, the Census Bureau began early this decade to examine alternative methods concurrently with reviews by scientific advisory groups. As noted earlier, the concept of integrated coverage measurement as a replacement for the 1990 dual-track strategy won widespread support from statisticians as well as from other advisory groups of interested parties. Integrated coverage measurement not only offers the promise of substantial reductions in the magnitude of differential undercounting, but also provides a significant opportunity to re-engineer census operations (see, e.g., Edmonston and Schultze, [24]). Increases in the efficiency of the census design should be realized by removing coverage improvement programs that have failed to correct the undercount and have provided diminishing marginal returns, often adding substantial gross error to the census and thus complicating the task of adjustment for undercoverage.

Components of the 2000 census design include a comprehensive national address list, improved administration of a redesigned mail questionnaire, streamlined field operations that include the use of sampling in following up nonresponding households, and earlier implementation of an expanded coverage measurement survey. These design changes are expected to yield adjusted population estimates in 2000 with greater accuracy, relative to the unadjusted counts, than the corresponding 1990 estimates.

Several factors linked to differences in the 1990 and 2000 census designs are expected to contribute improvements in accuracy. First, conducting coverage measurement survey operations closer to Census Day 2000 should reduce the magnitude of processing error involved in matching census and survey responses. Second, the 2000 coverage measurement survey will benefit from a substantially increased sample size of 750,000 households—five times larger than the 1990 PES. The larger sample size will enable production of direct state estimates—that is, the adjusted population estimate for a particular state, as well as for smaller geographic areas within the state, will be based only on information collected from that state. The reasons for this restriction seem largely a consequence of parochial political acceptability, although Thompson and Fay[67] do point to language in the 1996 Supreme Court decision regarding the 1990 proposed adjustment which suggests the desirability of such an approach. Statistically, this restriction prevents the use of valid and helpful cross-state-border information. Using such information permits us to attain the same level of uncertainty with a much lower sample size. In fact, much of the planned increase in sample size from 150,000 in 1990 to 750,000 in 2000 is required to compensate for this state-only restriction.

Thus, the design changes proposed for 2000 are expected to address two of the four most prominent technical criticisms of the 1990 PES methodology: (1) the magnitude of processing error in the coverage measurement survey, and (2) the use of smoothing models to combine information from multiple states in producing estimates within an individual state (but see the comments above on the lack of validity of this criticism). The proposed 2000 census design has drawn criticism on the two charges remaining from 1990: (3) variability of undercount rates within specific demographic groups (also known as within-poststratum heterogeneity), and (4) lack of independence between the census and survey operations because individuals missed in the census are more likely to be omitted also in the postenumeration survey (termed the problem of correlation bias). For reasons discussed in the Appendix A, we conclude that these technical difficulties are unlikely to alter the expected superiority of adjusted 2000 population estimates relative to estimates produced without the use of statistical sampling and estimation to supplement traditional census operations.

In summary, following Anderson and Fienberg [4] [5], we note that the proposed design for Census 2000 built on these earlier efforts to use sampling to improve accuracy. (1) The increase in the PES sample size allows fitting sampling better with the design of the census. (2) The lessons from 1990 are expected to lead to improvements in the accuracy of the PES and thus in the dual systems corrected counts. (3) Sampling for nonresponse followup (SNRFU), had it been allowed by the Supreme Court, would have replaced the faulty and inaccurate coverage improvement programs with a scientifically based survey which is expected to replace the biases in the late coverage improvement programs by a small component of sampling error. A compelling feature of SNRFU is that it should reduce the level of EEs and thus reduce a major component of the gross error. Moreover, since a large amount of the error in the 1990 PES was associated with EEs to the extent that SNRFU reduces the number of EEs, the PES/ICM will have much improved accuracy. Finally, the integration of enumeration, SNRFU and ICM means greater control over the census process and thus the prospect of the reduction of yet additional sources of error.

# 8 Can Demographic Analysis Offer a Viable Alternative to the Census 2000 Plan?

Demographic analysis uses a basic accounting identity to estimate a current population: the number of people living in an area now is the same as the number of people who lived there at some past time, plus births, minus deaths, plus people who moved in, minus people who moved out. Demographic methods for assessing census coverage combine data from previous censuses, vital statistics on births and deaths, and other administrative records, such as Medicare data, to produce national population estimates by age, race or ethnicity, and sex. Applying this approach to the United States as a whole, demographers believe that births and deaths are recorded relatively accurately. National migration statistics, however, both for immigration and emigration, are subject to large errors and considerable uncertainty.

Because of the lack of detailed information on internal migration and other state-level components, however, demographic analysis is generally regarded as unable to provide reliable estimates for such subnational areas as states (Steffey and Bradburn, [66]). Some critics of PES methodology have recently made the contrary suggestion that demographic methods could be used in 2000 to correct the national undercount and apportion the correction satisfactorily among smaller areas of geography (e.g., Brown, et al., [12], and Darga [19]). Appendix B considers the advisability of such suggestions. We find the current and likely future state of demographic analysis methods to be problematic except for the broad purposes of coverage evaluation for which they were originally designed. We have seen no empirical evidence that could support the call to use demographic analysis, by itself, for any type of adjustment.

# 9 Experiences in Other Countries

Although the proposed new uses of sampling in producing population estimates have been the subject of intense, often politically tinged, debate in the United States, such procedures have been introduced successfully and with relatively little controversy in other industrialized countries with similar traditional census operations.

As in the United States, censuses in Great Britain are taken every ten years and encounter similar undercoverage problems. Several adjustments are applied to census numbers in preparing local area population estimates that are widely used by the central government in allocating resources (e.g., see Diamond and Skinner, [22], as well as Diamond, [21], and Skinner, [63]). Two approaches were taken in estimating the overall undercount in England and Wales in their 1991 censuses: demographic methods and a postenumeration survey, the latter using a substantially different design and estimation procedure than that used in the 1990 U.S. census. After considerable analysis, adjustments using the demographic estimation approach were incorporated into the final estimates and have been generally accepted by local authorities. Diamond and Skinner note, however, that problems of estimating internal migration complicate the application of demographic methods to distribute the estimated national undercount among smaller geographic areas. Some have called for a redesign of the postenumeration survey used in England and Wales, possibly resulting in procedures resembling more closely the independent survey and dual system estimation carried out in the 1990 U.S. census.

In Australia, population estimates based on census counts adjusted for undercount have been released as the official population estimates since 1976 (e.g., see Choi, Steel, and Skinner, [14], and Steel, [65]). Censuses are taken every five years, and estimates are produced for states, territories, and local government areas. Population estimates are used to determine the number of seats each state has in the federal House of Representatives and the allocation of funds to states and local government areas.

The methods used in Australia to measure and correct for census coverage error share many features with those used in the 1990 U.S. census. An independent household survey is conducted after the enumeration, census and survey results are matched to determine missed people and some types of erroneous enumerations, and dual system estimates are produced. Adjustment of the Australian census was prompted by the high undercount rate found by the 1976 Post Enumeration Survey and by the considerable discrepancy between the 1976 census count and 1976 population estimates that were based on updated 1971 census results. There has been general acceptance and remarkably little controversy surrounding the adjustment decision (Steel, [65]).

Statistics Canada has also addressed the matter of adjustment for census coverage error using its special reverse record check methodology and other coverage programs, and has adjusted intercensal estimates following the 1991 and 1996 Canadian censuses. Net undercoverage in Canada exceeds that in the U.S., but it does not appear to differentially impact substantial minority populations in the same fashion as is the case in the United States. Dick, [23], describes the methodology used following 1991, and the techniques were further updated following 1996. These particular uses of adjustment have also generated remarkably little controversy.

# 10 Special Issues

The current controversy over census methods in the United States raises several important non-technical issues that are somewhat distinct from the report's focus on census operations and integrated coverage measurement. We examine two charges made against the proposed 2000 census design: that the proposed uses of statistical sampling (a) are prohibited by existing statutes and the U.S. Constitution and (b) would render the census process vulnerable to political manipulation.

As noted earlier, the Supreme Court recently ruled that existing federal legislation prohibits the use of sampling in producing population estimates to be used in reapportioning the House of Representatives. But the ruling does not address other uses of population estimates, such as allocation of federal program funds, nor does it preclude Congress from revising the existing statute to remove the prohibition for reapportionment. The Supreme Court pointedly did not address pre-emptively the constitutionality of possible legislative changes, as Justice O'Connor did point out in the majority opinion, but Justice Scalia took the occasion to opine on the matter in his concurring opinion. It is uncertain whether–and if so, when–the constitutional question will come before the Supreme Court.

In the absence of a ruling, we cite the findings of prominent legal scholars who studied the question as part of a National Academy of Sciences panel study concluded that sampling was permissible in producing estimates for reapportionment provided that the sampling was preceded by a good faith effort to enumerate the entire population (Edmonston and Schultze, [24]).

Legal arguments concerning statistical adjustment have been advanced by census litigants on both sides of the undercount issue. Opponents of adjustment have claimed that the language in Article 1 of the U.S. Constitution requiring an "actual enumeration" expressly prohibits the use of statistical sampling and estimation in producing decennial population totals. Proponents of adjustment have countered by arguing that coverage measurement surveys and associated estimation methods are not only permissible–and consistent with long-standing imputation methods used by the Census Bureau–but are required to correct the known differential undercount under the Constitutional guarantee of equal representation (e.g., see Choldin [15]). Others argue that it would be difficult to imagine that the words "actual enumeration" were meant to exclude the possibility of scientifically-based methods of sampling and estimation, since it was unlikely that the Framers of the constitution had any knowledge of such ideas. The method of probability sampling on which we now rely so heavily, for example, was not really set forth as a serious method until the late nineteenth century. The report prepared by Anderson and Fienberg [3] provides a more careful examination of this issue.

Critics of the 2000 census design have also charged that the use of integrated coverage measurement would render the census process newly susceptible to corruption and manipulation of population estimates for political advantage. For several reasons we find this charge completely without merit. First, the Census Bureau is staffed almost exclusively by career civil servants. Because of the openess of the Bureau and its longstanding and continuing efforts to assess accuracy and quality, the professional staff enjoys in the scientific community a time-honored reputation for integrity (see for example, Moore [48]). Second, the statistical sampling and estimation procedures, as well as other components of the 2000 census design, will be prespecified, thus establishing the ground rules well in advance of Census Day 2000. Third, the Census Bureau's activities will continue to be subject to independent review and oversight from various bodies, including the Census Monitoring Board, Congress, the General Accounting Office, and the National Academy of Sciences. To come to a balanced judgment about the question of political manipulability, it is important for us to consider the issue on the context of a 1990-style census without integrated coverage measurement. The number of people found in a geographic area depends critically on the effort spent finding them. Larger resources devoted to finding people, leaving census offices open longer in particular areas, and more intensive use of "coverage improvement" programs, all increase the nominal count in an area, whether the enumerations are correct or erroneous. Such decisions are made administratively and are not reviewed in any systematic technical fashion. Consequently, on balance it appears that a 1990-style census without ICM is more vulnerable to political manipulation than would be a 2000-style census incorporating ICM.

There is one aspect of political manipulability that deserves further mention. The Census Bureau is part of the Department of Commerce, arguably the most political of the departments of the federal government. Under both Democratic and Republican administrations, the Secretary of Commerce has been a former party national chair or other close political confidant of the President, in much the same way that the Postmaster General used to be. Perhaps, the nation would be well-served by moving the Census Bureau to a less politically-charged environment, either as the nucleus of an independent, non-political statistical agency (e.g., see the related recommendation in Norwood [52]) or to some other reporting structure.

The current state of the law leads to the likelihood of further political conflict. The recent Supreme Court decision [72] is based on an interpretation of the 1976 Census Bureau statute, Title 13 of the U.S. Code, which states:

> Except for the determination of the population for purposes of apportionment among the several States, the Secretary [of Commerce] shall, if he considers it feasible, authorize the use of the statistical method known as 'sampling' in carrying out the provisions of this title. 13 USC §195.

If nothing else, the use of sampling as part of the 1990 PES showed that sampling was feasible for adjustment purposes. Further, the Secretary, having proposed the use of the sampling for the entire 2000 census, can hardly now find sampling not to be feasible. The provision of Section 195 of Title 13 would then appear to require the Secretary to use sampling for every other purpose of the census (including for redistricting of congressional and state legislatures). According to recent press reports, the majority in the U.S. House of Representatives is likely to deny funds that would permit the use of sampling for such purposes (thus making it "infeasible?). We may confidently predict a summer of intense political maneuvering on this subject.

## 11 Summary and Conclusions

Statistical sampling and estimation methods have been used for more than 20 years to measure and correct for census coverage in other Western industrialized nations. These methods have been under development at the U.S. Census Bureau for more than 50 years. Although such technical issues as heterogeneity and correlation bias remain subjects of ongoing research, the postenumeration survey methodology is now a mature and proven census innovation. Evidence from extensive evaluations of the 1990 census and PES operations suggest, on balance, that incorporating sample-based coverage measurement and adjustment into production of the 2000 census population estimates will improve their accuracy at most levels of geography and, at worst, will have a neutral effect on overall accuracy for the smallest geographic areas.

Evaluation studies, such as those carried out with the 1990 census, should play an important role in informing decisions about the design of the next census. In evaluating the quality of population estimates it is equally important, however, to consider also such factors as the processing and quality assurance procedures in both census and survey operations, the consistency of current estimates with past results, and comparisons with other independent sources of information, including estimates obtained from demographic analysis (Steel, [65]). Population estimates should also be judged on their face validity—that is, their consistency with what is thought to be known about the patterns of undercount in the population. On all these counts, we find the 1990 evidence to support the use of statistical sampling and estimation in the 2000 census.

Finally, it must be recognized that the development of a viable design for a project as large as a national census requires the ability to make key decisions at various points in the decade prior to Census Day. Such decisions should not be suspended or impeded by the presence of technical disagreements between well intentioned statisticians. A disinterested statistician familiar with the Australian census operations observed that complete consensus among members of the American statistical community was unlikely to occur because some critics of the 1990 PES methodology were requiring that sampling-based adjustment methods meet unattainable standards of proof (Steel, [65]). We concur with that assessment and believe it is time to carry on the tradition of innovation in American censuses by incorporating integrated coverage measurement in the 2000 census, thus using the best available scientific methods to provide the country with population estimates of the highest possible quality.

# A A Critique of Arguments Against Sampling-Based Coverage Adjustment in Census 2000

## A.1 Introduction

There are few strong critiques of the Post Enumeration (PES) methodology associated with the 1980 and 1990 censuses and the proposals from the U.S. Census Bureau of the methodology to be used in Census 2000. Many if not all of these critiques emanate from a single location: the Department of Statistics at the University of California. For commentary on 1980 Freedman and Navidi [34], [35], and for extensive critiques of the 1990 PES methodology, see for example Breiman [10], [11], Freedman [32] [33], Freedman and Wachter (with various co-authors) [38], [36], [37], [39], Wachter[73], and Wachter and Speed [74].

Kenneth Darga [19], a demographer from Michigan, repeats many of Breiman's claims regarding bad data in the post enumeration survey, but adds little to the litany of criticism contained in the foregoing references, although he reifies the method of demographic analysis that we believe is totally inappropriate, and fails to describe in any detail the methodology proposed for census 2000, except to indicate that is similar to that of 1990.

Thus, to respond in any detail to the criticisms of the plans for Census 2000, we have turned to two as yet unpublished Technical Reports from the Department of Statistics at the University of California, Berkeley:

1. Brown, L.D., Eaton, M.L., Freedman, D.A., Klein, S.P., Olshen, R.A., Wachter, K.W., Wells, M.T., and Ylvisaker, D., "Statistical Controversies in Census 2000," Department of Statistics, University of California, Berkeley, Technical Report 537 (15 October 1998). [12].

2. Stark, P.B., "Differences between the 1990 and 2000 Census Adjustment Plans, and Their Impact on Error," Department of Statistics, University of California, Berkeley, Technical Report 550 (16 March 1999). [64]

The first report is jointly authored by a number of the earlier critics and summarizes many of the arguments presented in their earlier papers on the 1990 census. The second report, authored by Philip Stark, includes a basic exposition of the capture/recapture methods used for census coverage measurement, the same criticisms of the 1990 PES made by earlier critics and more recently by Darga, and more current information about operational aspects of the 2000 census design. Both reports are available from the WWW at: http://stat-www.berkeley.edu/tech-reports/index.html

This appendix excerpts statements from these technical reports–which we refer to as TR537 and TR550, respectively–and responds to a series of what we take to be incorrect or misleading statements, or ones whose import can only be understood in an appropriate context. While we agree with these authors about many issues and concerns that they raise regarding the proposed Census 2000 methodology, we disagree with most of their analyses and conclusions.

## A.2 Three Principal Allegations Regarding the 1990 PES and Census 2000

We begin by examining three of the most serious allegations made by critics against the 1990 PES and, by implication, the use of similar methodology in 2000:

1. "[A]djustment adds more error than it removes." (TR550, section 1)

2. “The 1990 PES probably would have reduced the accuracy of state population counts and shares.” (TR550, abstract)

3. “The 1990 adjustment fails some simple plausibility checks.” (TR550, abstract)

**Claim 1: Adjustment adds more error than it removes.**

This argument appears in both technical reports. TR537 contains the statement, “The greater part of the adjustment in 1990 turns out, then, to be attributable to errors in the post enumeration survey: the proposed adjustments were more artifactual than real.” In TR550, Stark repeats the claim that “more that half of the adjustment reflects statistical bias resulting from poor data quality and faulty methodological assumptions.” This claim originates primarily from analyses reported in Breiman [10], suggesting that at least 55% and perhaps 80% or more of the net national undercount estimate in 1990 is attributable to bias in the PES.

Although we disagree with some of the details of Brieman’s analyses and his characterization of the sources of PES error as producing “bad data,” there is a much more fundamental flaw in the logic of this often repeated argument. A hypothetical outcome, adapted from Anderson and Fienberg [4], will serve in its extreme to illustrate the fallacy. Suppose that the traditional 2000 census operations produce 30 million erroneous enumerations, all of which happen to occur east of the Mississippi River. Suppose also that 30 million omissions occur, all west of the Mississippi River. Thus, the true net national undercount is zero. Now suppose that the 2000 accuracy and coverage evaluation (ACE) estimates that 30 million erroneous enumerations occurred east of the Mississippi and 30 million plus one omissions occurred west of the Mississippi. Thus, the 2000 ACE yields a net national undercount estimate of one person and is astonishingly successful in identifying both components of gross census error and in correcting for the geographic inequities they would have introduced in the population estimates. However, applying the rationale of the Berkeley critics, one is led to the conclusion that 100 percent of the net national undercount estimate is due to error, or “bad data,” in the 2000 adjustment.

Relying on this flawed metric, Breiman proceeds to conclude that “the fact [sic] that 80% of the undercount estimate is due to poor data quality indicates that the DSE estimate state and local adjustments are largely reflections of bad data in an unpredictable pattern.” This conclusion is accepted by the authors of the recent technical reports and is used to argue against sampling-based adjustment in the 2000 census. Interestingly, however, there is strong empirical evidence to refute this claim—evidence kindly provided to us by Lawrence Brown in a personal communication.

Figure 1 is a scatterplot of data for the 50 states and the District of Columbia. On the horizontal axis is the change from 1980 to 1990 in a state’s share of the national population computed from the unadjusted counts in the two censuses. On the vertical axis is the difference between a state’s unadjusted 1990 share and its share computed from the production PES adjustment in 1991. Rather than exhibiting an “unpredictable pattern,” the data clearly lie close to a straight line. Least squares regression through the origin–the intercept term is not significantly different from zero–yields a slope estimate of 0.102 and accounts for 78.7% of the variation in adjustment-based changes in state shares. These findings suggest that, instead of adding noise to the estimated population shares, the 1990 PES-based adjustment does something quite systematic and credible-namely, adding or subtracting about another 10 percent to the estimated change in state share from 1980 to 1990.

An alternative analysis (also provided to us by Lawrence Brown) eliminates the District of Columbia (not a state) and the four Mexican-border states: California, Arizona, New Mexico and Texas. Even with these five omitted, a similar pattern is evident, although less strongly so. Given the large undercounts and minority populations in the four border states one might argue for keeping them in the analysis, they correspond to

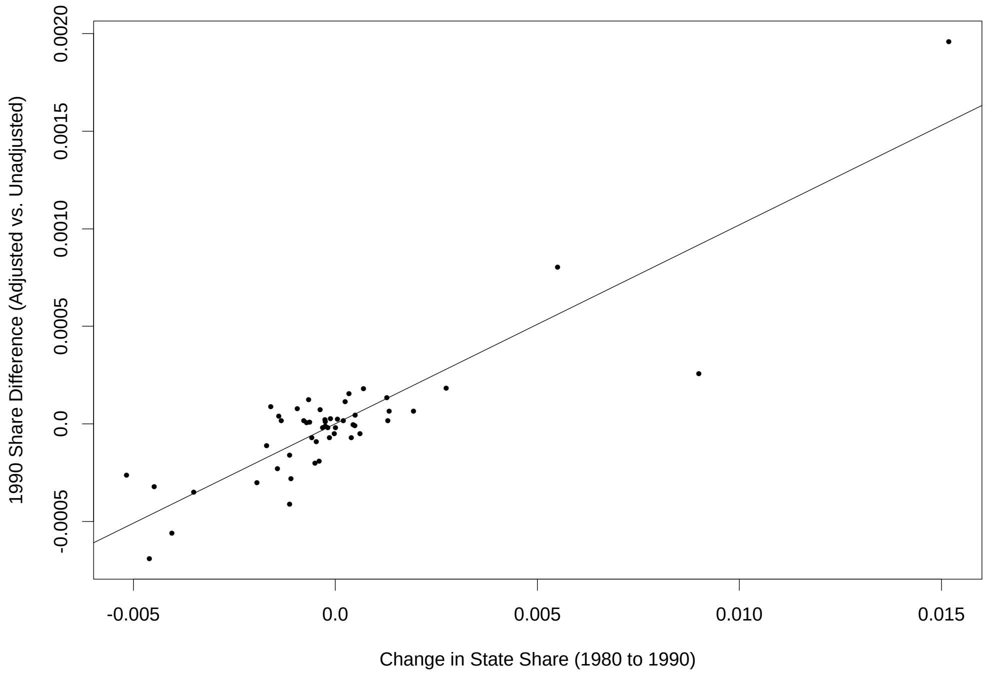


Figure 1: Predictability of Adjusted Shares

high leverage observations, and reasonably so to us. We agree, however, with Brown on the appropriateness of setting aside the District of Columbia. But even were we to accept this alternative analysis as the most apropriate, it still provides evidence counter to the claim that adjustment adds more error than it removes.

**Claim 2: Adjustment in 1990 probably would have reduced the accuracy of state population counts and shares.**

This claim, made in TR550, cites Breiman [10] and TR537 for support and ignores contrary evidence in other sources, including the Census Bureau's P-16a and P-16b reports (see also the recent discussion in Cohen, White, and Rust [17]). For the reasons presented above, we do not find Breiman's analyses or conclusions pertinent to this question. TR537 does not address the relative accuracy of unadjusted and adjusted population counts and makes only the more modest claim that "adjustment could easily have made population shares less accurate rather than more accurate, even for states."

The attempt to cast doubt on the relative accuracy of adjusted state population shares relies largely on a loss function analysis presented in TR537 as an alternative to the extensive total error analysis carried out by the Census Bureau and reported by Mulry and Spencer [51]. The authors of TR537 argue that, "With other assumptions that are perhaps more realistic [than those used by the Census Bureau], the [risk] estimate becomes $-250 \pm 821$; units are parts per 100 million, positive values favoring adjustment and negative values favoring the census. Even at the state level, the case for adjusting the 1990 census was shaky at best. There was a strong likelihood that adjustment would have put in more error than it took out." (p. 12)

The first sentence of this argument is misleading, because the substantial standard error renders the risk difference insignificant. Moreover, several aspects of the TR537 loss function analysis are problematic. Some of the underlying assumptions are clearly in error and appear to result from misunderstandings about the original total error analysis. We consider these in detail below. By comparison, the Census Bureau's analysis estimates that use of the unadjusted census numbers involves significantly greater risk. Thus, the findings presented in TR537 are based on indefensible assumptions and do not provide credible support for the alleged superiority of unadjusted state shares.

In the remainder of this subsection we provide a line-by-line rebuttal of statements found in the TR537 loss function analysis (section 7 of that report). The page numbers in parentheses refer to our printed copy of the document that we downloaded from the UC Berkeley web site.

**Statement**: "We find these assumptions [of the Census Bureau's total error model] to be extremely questionable: for example, $\hat{\beta}$ is severely biased." (p. 13)

**Rebuttal**: The authors have mischaracterized the derivation of $\hat{\beta}$. State-by-state variability may be underestimated, but the data are not adequate to answer conclusively any questions about the bias of $\hat{\beta}$.

**Statement**: "$\hat{H}$ is off by a factor of . . . 84 . . . [because] the [bias] allocation scheme converted variance in $\hat{\beta}$ to bias, and the model assumed that bias in $\hat{\beta}$ vanishes." (p. 13)

**Rebuttal**: The argument seems to be as follows:

1. The trace of $\hat{G}$ would have been twice as big if no smoothing had been done.
2. The trace of $\hat{H}$ would be the same as the trace of $\hat{G}$ if the sample sizes had been the same.
3. The sample size underlying $\hat{H}$ was only 7% of the sample size underlying $\hat{G}$.
4. Assertions 2 and 3 imply that the trace of $\hat{H}$ is equal to twice the trace of $\hat{G}$ divided by 0.07.
5. The Bureau's jackknife estimate of the trace of $\hat{H}$ was only 0.33 times the calculated trace of $\hat{G}$.
6. The Bureau's estimate is too small by a factor equal to (2)(1/0.07)(1/0.33) or about 84 or so.

The argument is false. It is derived considering only the ratio of PES and EFU sample sizes and fails to account for the fact that $\hat{G}$ and $\hat{H}$ are estimating different variance components. In fact, it is well known that the sampling variance for a mean may be expressed as the product of four factors:

**a.** the finite population correction factor. In the current situation, the finite population correction is approximately 1 for both samples, the PES (giving rise to $\hat{G}$) and the evaluation sample (giving rise to $\hat{H}$).

**b.** the reciprocal of the sample size. The sample size for the evaluation sample was approximately 7% of the PES sample size.

**c.** the design effect or "deff", which accounts for the complexity of the sample design. The deff is defined as the ratio of the variance of the statistic of interest under the actual sample design to the variance that would occur if the statistic had been calculated from a simple random sample of the same size. The design effect accounts for clustering and stratification and unequal selection probabilities. The design effect may be smaller for the evaluation sample than for the PES if it was selected to reduce the variability of sampling weights.

**d.** the population variance. The assumption that the population variances are the same is critical. No evidence is presented to show that they are the same. For example, if one is estimating two different proportions from two different simple random samples of the same size, and the expected values are 0.5 and 0.005, the variance for the estimate of the former will be 50 times the variance of the estimate of the latter.

The Census Bureau's estimate of $\hat{H}$ was developed with tested software based on standard methodology. The estimate in this critique seems based on unstated assumptions which lack empirical basis.

**Statement**: "The matrix $\hat{G}$ is too small as well, by a factor of 2 to 3 ..." (p. 13)

**Rebuttal**: This statement is based on Bob Fay's 1992 reanalysis [27]. Even if a factor of 2 or 3 is conceded, the change does not significantly alter the original loss function analysis.

**Statement**: "The [Census Bureau's] allocation assumed (i) bias was proportional to some covariate for post strata within evaluation post strata, and (ii) bias was constant across states within post strata ... [B]y comparison, we allocate bias to states in proportion to estimated undercounts ..." (pp. 13-14)

**Rebuttal**: Mulry and Spencer did not allocate total bias to states. Rather, they allocated individual bias components to states and then performed total error analysis within each state. The authors apparently obtain the numbers in Table 2 of TR537 using a very simplistic argument based on an expression for the risk difference (Census $-$ PES) for the $k$-th area (state):

$$\hat{R}_k = (X_k - \hat{\beta}_k)^2 - \hat{\beta}_k^2 - 2\hat{G}_{kk}.$$

Putting $\hat{\beta}_k = .25X_k$ and $\hat{\beta}_k = .50X_k$ produces the entries in the latter half of Table 2. These entries are derived using clearly false assumptions about $\hat{H}$ and an unsupported simplistic method for assigning PES bias that seems to follow from arguments that focus improperly on net national undercount as the denominator base. The Breiman claim that over half the net undercount is due to error, which is highly misleading and at best irrelevant, is thus pivotal to the reasonableness of the bias allocation.

**Statement**: "Table 2 may in some respects understate the problem. Heterogeneity is probably larger than all the other errors put together, and heterogeneity is not taken into account in Table 2; nor is there any allowance for bias in the imputations." (p. 14)

**Rebuttal**: The original total error analysis does include heterogeneity (correlation bias), and imputation bias was converted to variance and included (see Appendix A.7 of Mulry and Spencer, 1993). Where is the authors' analysis that treats all sources of error? There is no basis for concluding that the results would ultimately favor the census over the PES.

**Statement**: "We think the last line [in Table 2] is the most plausible ... [L]oss function analysis depends very strongly on the assumptions going into the evaluation: the available data do not decide the issue." (p. 14)

**Rebuttal**: The alternative analyses in TR537 range from neutral to favorable toward adjustment. Notwithstanding the problems discussed above, in no case is there evidence in TR537 that the 1990 adjustment would have made matters significantly worse. Although we agree that the decision to proceed with ICM in 2000 cannot be supported solely by a narrow focus on loss function analyses, the suite of analyses discussed here clearly point in that direction.

**Claim 3: The 1990 adjustment fails some simple plausibility checks.**

Three observations are cited in TR550 as evidence of the implausibility of the 1990 adjusted population estimates:

1. Northeastern and midwestern states with large central-city minority populations, i.e., New York, Illinois, Michigan, Massachusetts, and Pennsylvania would have reduced population shares under adjustment.

2. Adjustment shifts gender ratios in certain demographic groups from typical values (e.g., 51% boys) to highly implausible extremes (48% to 56% boys).

3. The PES-based estimates of net national undercount for certain race and gender groups do not agree well with estimates obtained via demographic analysis.

The authors of TR537 do not consider gender ratios, a line of argument due to Darga [19], but they attribute the other discrepancies to failures in model assumptions underlying the capture-recapture approach and dual-system estimation: "Reachability varies from place to place around the country, and differential levels of correlation bias are a major threat to the accuracy of adjusted population shares ... The post enumeration survey may well have been more successful at finding undercounted black males in California and Texas than in New York, Illinois, and Pennsylvania." (pp. 6-7)

Several counterpoints blunt the impact of these arguments. First, results such as the reduction in northeastern and midwestern population shares may be counterintuitive but are not necessarily in error, as conceded in TR537. Moreover, arguments based on comparisons of gender ratios and net national undercount estimates of subpopulations rely on the validity of demographic analyses and do not acknowledge the possible flaws in these methods. The only direct evidence on the size and effects of correlation bias is at the national level and comes from comparisons with demographic estimates. But the latter are subject to large amounts of error and bias of their own (see the discussion in Appendix B).

Most fundamentally, the observed and potential shortcomings of sampling-based coverage adjustment need to be weighed against the deficiencies of the unadjusted population estimates. No one can defensibly argue that PES-based adjustment in 2000 will yield perfect results. Supporters of the 1990 adjustment have long acknowledged the existence of correlation bias and within-poststratum heterogeneity in undercount rates, among other problems, and have sought means of measuring and understanding their effects. See, for example, the discussion and references in Belin and Rolph [7]. The potential effects of violations of DSE assumptions, however, should be considered relative to the severe violations of implicit assumptions in the traditional census–problems that have persistently over the decades produced population estimates that are completely implausible because of clear and substantial bias in the coverage of minority populations at aggregated levels.

Such a perspective seems absent when the authors of TR537 write, "We conclude that correlation bias is a serious and intractable problem. The bias is serious because it can result in adjustments that make state shares worse rather than better. The bias is intractable because it cannot be measured at subnational levels. Efforts to model the bias at the state level in 1990 were quite unsuccessful ... " (p. 9)

Although we regard this as perhaps the most serious complaint in TR537, it concerns the 1990 census and may arise from conditions that are unlikely to be present in 2000. Specifically, problems with erroneous enumerations from coverage improvement efforts (e.g., the parolee and probationer program) could have been a major contributing factor to the estimated biases in the adjustment methodology in 1990. Also, erroneous enumerations were likely a bigger problem for the census than for the postenumeration survey and the related adjustment methodology.

Furthermore, at the level of poststrata, if the probabilities of being enumerated in the survey and in the census were positively correlated, the adjustment at least moves the raw census counts in the correct direction (see e.g., Kadane, Meyer and Tukey, [45]) and for "small" adjustments this property transfers to shares as well (see Cohen, White, and Rust [17]). Finally we note that by far the biggest part of the adjustment comes not from those missed in both the census and the sample but rather from those missed in the census but found in the sample. This fact, combined with the other evidence, suggests that a focus on the potential effects of correlation bias and differential biases is likely misplaced.

## A.3 Further Comments on "Statistical Controversies in Census 2000"

In both TR537 and TR550 we find many other instances in which a disproportionate burden of proof is placed on methods for sampling-based coverage adjustment relative to similar methods used in traditional census operations. Real or potential problems with aspects of the PES and dual system estimation are often judged with little regard for the overall context of census operations and objectives. The rhetoric of these reports is not characteristic of a neutral scientific assessment but rather seems intended to raise doubts and fears about implementation of an extensively tested methodology. For these reasons we feel compelled to provide a close textual reading and further commentary on these reports.

TR550 is notable partly because it contains some new information (i.e., about ACE, the renamed ICM effort) concerning the 2000 census design. However, the predicted comparisons in TR550 of the 2000 "headcount" and ACE results with their 1990 counterparts necessarily involves a great deal of unverifiable speculation. We previously addressed what we regard as TR550's most serious weaknesses: its uncritical reliance on Breiman's metrics and on demographic analysis, including the references to recent work by Darga. Also, we strongly disputed above the claim made in TR550 that the 1990 PES probably would have reduced the accuracy of counts and state shares. Consequently, we do not consider this report any further.

In the remainder of this appendix we address, in order of their appearance in TR537, statements that we believe require either rebuttal or clarification in order to be understood accurately by readers. The subsection headings below correspond to those in TR537, and, as above, page numbers are for the version we downloaded from the UC Berkeley web site.

### 1. Introduction

**Statement**: "Despite the name, ICM is not an inseparable part of the census operation." (p. 1)

**Rebuttal**: This statement does not recognize that ICM potentially has significant implications for the earlier stages of census operations. The inclusion of sample-based coverage measurement and adjustment as an integral step in completing the census provides a compelling argument for bringing discipline to the census process and eliminating expensive and ineffective coverage improvement procedures that in past censuses have substantially increased gross census error. Without ICM, the 2000 census would almost certainly rely on the same flawed procedures. With ICM there is at least an opportunity for meaningful operational reform.

**Statement**: Estimating share changes of small magnitude is "a tricky business."

**Rebuttal**: This comment is made only in reference to the estimation using adjusted counts and is made to suggest that the error in the adjusted counts is too great to allow for accurate estimation of shares. But what about the error in estimates of shares based on the uncorrected census counts? Further, are population shares, rather than gross counts, the most appropriate units of analysis for all purposes?

**Statement**: Subtraction of erroneous enumerations from gross omissions "entails a substantial loss in accuracy." (pp. 1-2)

**Rebuttal**: Do the authors really mean that we should ignore such a large component of error in the census? Their focus on net national undercount is not appropriate. As we argued in the main report above, even if we had succeeded with a net national undercount of 0, we would still have major accuracy problems if the levels of gross error are on the order of magnitude of those experienced in 1990.

**Statement**: "Efforts to demonstrate the validity of adjustment are more problematic than adjustment itself, because they depend even more heavily on assumptions that turn out to be rather fragile." (p. 2)

**Rebuttal**: This claim is simply not supported by the authors' analyses. Moreover, it rests heavily on the notion that one needs to prove that the adjustment is better for all components and all purposes. Were such a requirement applied to all other changes in methods in the conduct of the census, we would never see any progress in census methodology.

### 2. Integrated Coverage Measurement

**Statement**: "In combination, the two errors of 10% create an error that exceeds 60% of the estimated undercount." (p. 3)

**Rebuttal**: The focus here, on the net national undercount, is simply not appropriate. (See detailed comments above.)

**Statement**: "Erroneous enumerations are likely to be even more numerous in 2000 than in 1990 because in 2000 the Census Bureau will be broadcasting the forms in many different ways." (p. 3)

Rebuttal: This claim ignores the potential for a streamlined, simpler census operation in 2000 with the removal of layers of "coverage improvement." It also ignores improvements in methods for identifying duplicate records.

### 3. Processing Error

**Statement**: "Processing errors are likely to be even more serious in 2000 than they were in 1990: the larger sample size and the tighter timetable make quality control more difficult." (p. 3)

**Rebuttal**: Streamlined field enumeration operations would ease time pressures and reduce processing burden from erroneous enumerations. Further, the timetable is set by legislative mandate and can be changed. In any case, there is little or no empirical evidence to support this claim even with the current timetable.

**Statement**: "Matching people between two large sets of records (the census and the ICM) is a complex and error-prone process. ... [E]rrors introduced by adjustment are far from evenly distributed."
(p. 4)

**Rebuttal**: Matching errors are clearly a focus of concern for statisticians, and they require careful study and documentation. The most recent assessment in Cohen, White, and Rust [17], concludes that "matching is unlikely to have a substantial effect on the resulting adjusted population counts." Further, many other parts of the census operation involve extensive matching—e.g., use of administrative records for improving address list coverage. ICM operations should not be held to a higher standard than other, less well documented, parts of census operations that add unknown biases to the census population estimates.

**Statement**: "There are indications that in 1990 balancing error was appreciable. ... The problem for 2000 will be worse." (p. 4)

**Rebuttal**: We do not agree that there was serious evidence of balancing error in 1990, and the authors offer no empirical support for the claim that 2000 will be worse.

**Statement**: "In 1990, there were about 4 million unresolved cases in the P-sample universe and a similar number in the E-sample ... Four million is a big number relative to a total estimated undercount of 5.3 million." (p. 5)

**Rebuttal**: This statement is misleading and irrelevant. The number cited is obtained from case weights and is compared to net national undercount instead of the estimated national population. Four million is not as big when the denominator is 250 million, so the overall disagreement rate is about 1.8%. Also, the number of the P-sample matches in the 1990 rematching study, when weighted up to the U.S. population, differs from the number in the PES by only 0.18%; this is because the figures quote do not allow for offsetting errors.

**Statement**: The imputation models used for unresolved PES cases were biased and sensitive to plausible changes in model assumptions. "If we assign to the Q-class a match rate equal to the average rate for cases with partial information, the estimated undercount would on our reckoning go up by roughly 400,000 people." (p. 5)

**Rebuttal**: This statement again ignores such issues as offsetting errors, but more importantly, the authors offer no basis for anyone to believe that the alternative assignment suggested is plausible.

**Statement**: "At this time, it is not clear whether or how erroneous enumeration rates will be computed for outmovers. [T]he ICM design [for the 2000 census] is more problematic than the post enumeration survey of 1990." (p. 6)

**Rebuttal**: See Waite and Hogan [75] for details. We do not understand why the ICM design [for the 2000 census] is more problematic than that of the post enumeration survey of 1990.

**Statement**: "Errors in the PES—the coding error being the dominant factor—would have moved a congressional seat from Pennsylvania from Arizona." (p. 6)

**Rebuttal**: This statement is misleading, because it combines the effect of the coding error with effects due to smoothing extreme block clusters and to the restratification of the PES that was done in adapting the PES for intercensal uses (e.g., see Hogan [43].) (The coding error alone would not have shifted a seat.) It also ignores the fact that, under the current reapportionment algorithm, assignment of the final congressional seats is very sensitive to small changes in population. Undocumented biases in other parts of census operations could have an equal or greater effect than the observed PES error effect in 1990.

**5. Heterogeneity**

**Statement**: "The choice of post strata may therefore have a substantial influence on the results. . . . [T]he choice of post strata is somewhat subjective." (p. 9)

**Rebuttal**: Of course the choice of post strata involves judgment, but so does almost every phase of the traditional enumeration and most major statistical surveys. Moreover, we could expect that each choice would yield somewhat different outcomes; however, the differences are very hard to predict in advance. What is important is that most of the sensible choices would likely yield improvements over a traditional census without a post enumeration survey, and this should be really all that matters—improved accuracy.

**Statement**: "[C]onsiderable residual heterogeneity... added appreciably to the uncertainties about adjustments for state and sub-state areas, and significantly complicated the task of comparing the accuracy of the census to that of adjustment." (p. 9)

**Rebuttal**: The claims here regarding heterogeneity are inextricably intertwined with those regarding correlation bias, since the impact of heterogeneity is to induce an effect which is indistinguishable from probabilistic dependence of capture probabilities. The evidence to support the existence of substantial residual heterogeneity problems is mixed at best, and Fay and Thompson [29] to whom the authors refer for support on this point offer none except "concern."

**Statement**: "The Bureau has not finalized its post stratification for 2000. . . . [H]eterogeneity within states may even be more of a problem in 2000 than it was in 1990." (p. 10)

**Rebuttal**: More details are now available in Waite and Hogan [75] and Schindler [62],, but there is certainly no reason to expect that heterogeneity within states is any greater a problem in 2000 than it would have been in 1990. Where is there evidence to support such a claim regarding the increased problems for 2000?

**6. Dual System Estimation**

**Statement**: "[M]istakes in fieldwork are magnified by the [ICM sample] weights. . . . [T]he weights may . . . be adjusted at various stages of the process, for reasons that are not entirely compelling-and the impact of such changes can be quite striking . . . [A] decision to revise the weights of two block clusters out of the 5,300 in the 1990 post enumeration survey subtracted 654,000 people from the adjusted count." (p. 11)

**Rebuttal**: The comments regarding the effects of the weights of the extreme block clusters in 1990 are correct, and it is for precisely this reason that adjusting the weights to ensure robustness seems to be sensible.

**Statement**: "The detailed structure of the [smoothing] model seems quite arbitrary, and details have major impacts . . . [P]re-smoothing and benchmarking account for 1/3 of the shift [in the proposed 1990 adjustment]." (p. 12)

**Rebuttal**: The denominator for this ratio is the 1991 PES net undercount estimate of 2.11%. Freedman, et al. ([38], p. 382) cites Table 15 of Census Bureau's P-16a report ([50]). Also see Rolph [61] and Belin and Rolph [7] for a direct response to the claim that the smoothing model was arbitrary in structure. More importantly, aspects such as pre-smoothing and benchmarking are not part of the proposed Census 2000 adjustment process, and thus this comment seems irrelevant for the 2000 census.

**Statement**: "[S]imulation studies and sensitivity analysis suggest that estimated variances were too small by a factor of 2 or 3." (p. 12)

**Rebuttal**: The alleged culprit is heterogeneity within poststrata but these numbers are mere speculation. Consult Fay [27], Freedman, et al. [38], and Rolph [61]. Even with variances larger by a factor of 2, however, the loss function analyses of the Bureau favored the use of the PES-adjusted results.

**Statement**: "Although the post enumeration survey in 2000 will be 5 times larger than the one in 1990, there will be 2 or 3 times as many post strata; thus, sampling error may still be a problem. . . . [A] reduction in variance [by smoothing] is likely to be accompanied by some increase in bias, and the tradeoff is extraordinarily hard to assess." (p. 12)

**Rebuttal**: Bruce Spencer (personal communication) and others do expect that the magnitude of sampling error in 2000 will be comparable to that in 1990. Increase in sample size is spent on more detailed and less efficient, poststratification to compute direct state estimates. For details, see Waite and Hogan [75], Thompson and Fay [67], and Schindler [62].

### 8. Sampling for Non-Response Follow-up

Although this aspect of the 2000 census design was eliminated by the unfavorable ruling of the U.S. Supreme Court in January 1999, we include commentary on the subject here in case the proposal is revived in the future.

**Statement**: "There is some opinion that sampling improves accuracy since interviewers can be better trained and supervised. Given the proposed sampling rates, this advantage cannot be substantial." (p. 15)

**Rebuttal**: The gains come from the full SNRFU process which the authors ignore.

**Statement**: "Sampling seems inherently more complex than a census." (p. 15)

**Rebuttal**: The census operation is complex in many respects. Although SNRFU adds some complexity, it also permits streamlining of field enumeration efforts. See the Census Bureau's 1997 report to Congress for further details ( [69], p. 27).

**Statement**: "[P]ast experience shows that sample surveys have worse coverage than the census. Of course, sampling for non-response in Census 2000 may not be directly comparable to past surveys." (p. 15)

**Rebuttal**: This argument seems diversionary and not especially relevant. The authors seem to suggest, implicitly, that the alternative of the traditional census would achieve coverage on the order of the 97% achieved by the 1990 PES. This is demonstrably false. Moreover, NRFU sampling attempts to cover mail-nonresponding households, which will likely range in 2000 from a few percent to about 60% of the entire population, depending on the geographic area.

**Statement**: "It is by no means obvious that SNRFU will reduce differential undercounts." (p. 15)

**Rebuttal**: The primary goal of SNRFU is cost-effective field enumeration of mail non-respondents, not reduction of differential undercounts. This is the mission of ICM. There is every reason to believe that SNRFU would reduce erroneous enumerations because it comes with the elimination of many of the programs that in 1990 introduced erroneous enumerations into the census results.

**Statement**: "Late returns to the census create further problems. The Bureau has proposed a solution, assuming in essence that selecting people into the non-response followup sample does not change their propensity to file late census returns, conditional on tract and occupancy status of housing unit. This assumption does not seem plausible. Furthermore, the solution seems to ignore late returns from housing units not in the Master Address File." (p. 15)

**Rebuttal**: Why is the assumption not plausible? And exactly how big a problem would such late returns actually be in the context of the Census 2000 design? With a changed schedule and focus for completion of the mail-out-mail-back census component, and the replacement of the followup procedures by SNRFU, we would expect substantially reduced late returns.

**Statement**: "Other assumptions must be made here: (i) census coverage will be the same whether non-response followup is done on a sample basis or a 100% basis; and, (ii) 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. A test census may provide evidence showing that contamination error is modest. However, power is limited and the real census may be somewhat different." (p. 16)

**Rebuttal**: The viability of many aspects of census operations cannot be rigorously confirmed in a test census environment. Throughout the decade there are important points in the census design process when decisions must be made with the best information that is available. In the absence of information suggesting the operations introduce substantial gross error, deferring decisions about major components of the census design increases the risk of an unfavorable outcome in 2000 that damages the credibility of the census population estimates.

**Statement**: "[T]he ICM cannot detect any non-sampling errors in SNRFU, because the ICM does not–by design–examine any blocks where non-response followup is done on a sample basis." (p. 16)

**Rebuttal**: While this is a correct description of the design, we do not know what analyses the Census Bureau may have done to support the decision not to use SNRFU in the ICM blocks. Of course one could do direct evaluations of SNRFU, using other traditional forms of sample-based quality checking and evaluation; unfortunately given the recent Supreme Court decision we are unlikely to learn how the Bureau planned to approach this issue.

**9. Other Arguments**

**Statement**: "[T]he central question remains. . . . Are the ICM data good enough for their intended use? Will proposed adjustments to the census take out more error than they put in?" (p. 16)

**Rebuttal**: We believe that the evidence on both these questions supports an answer of yes. Moreover, this statement implies disproportionate burden of proof on ICM relative to other components of census design. Suppose we were to ask the reverse question: Are census estimates obtained without ICM good enough for their intended use? Our assessment here is absolutely not; the recently completed panel of the National Academy of Sciences (Cohen, White, and Rust, [17]) reaches similar conclusions.

**Statement**: "Sampling 3 households in 4 cannot save very much money. . . . [T]he savings may be on the order of . . . $125 million." (p. 16)

**Rebuttal**: This contradicts Census Bureau's estimate of $400 million in savings (pp. 37-39 of the Bureau of the Census report [69]) and no details are provided to support the alternative figure. Also, the mathematics here doesn't make sense. Moving from a 75% sample to a 100% sample increases the sample size by a factor of 1/3 not 1/4.

**Statement**: "Reports like Steffey and Bradburn [66] provide little analytic detail supporting adjustment. Nor are they quite as supportive of sampling as the Bureau makes out . . . " (p. 17)

**Rebuttal**: This statement confuses sampling done to reduce costs (SNRFU) with that done to improve accuracy (ICM). But the report cited was strongly supportive of both uses of sampling, subject of course to careful field testing. For further support with analytical backup at a level far exceeding that in the present critique see Cohen, White, and Rust, [17].

**Statement**: "SNRFU and ICM will both suffer from non-sampling error, which will be as hard to quantify as the errors in the census." (p. 17)

**Rebuttal**: This is hardly a compelling argument for not doing them. The real issue is whether the incorporation of these components (even with their attendant nonsampling errors) reduces the differential undercount bias. There is every reason to believe that this was the case in 1990 and that it would be so again in 2000.

**Statement**: "Indications are that the MAF for 2000 is incomplete and inaccurate. ... This is quite troubling, and may be a strong argument for adjustment." (p. 17)

**Rebuttal**: We agree that the MAF is cause for concern. The Census Bureau has changed its plans for address list development and to us the plans remains problematic (see the report of the General Accounting Office on this issue [71]).

**Statement**: "Adjustment will correct the differential undercount for national race and ethnic groups, and may improve the accuracy of state and local shares. ... However, it is not an argument for SNRFU and ICM, because there is a much simpler way to correct the national undercount. Census figures could be scaled up to match the demographic totals for subgroups of the national population. ... Currently, demographic analysis does not account for ethnicity, but the method could be adapted for that purpose. ... Despite the flaws in demographic analysis and the heterogeneity of undercount rates, we do not see much likelihood of demonstrating that ICM will be more accurate than scaling. On the contrary, ICM may do worse because of some hidden breakdown in the operation." (p. 17)

**Rebuttal**: This statement reiterates the authors' seemingly relentless focus on net national undercount to the total neglect of actual error in the census counts. There is little analysis to support the claims for demographic analysis; our analyses suggest that their use at the national level should be the subject of much greater concern (see Appendix B). The measurement of race and ethnicity in the demographic estimates is incomplete and a growing source of error when these estimates are linked to those using the census's questions on the topic. How would the author's propose to evaluate the quality of the counts and shares that would result from "scaling" the demographic estimates? And how would they show that scaling is superior to ICM? After all, we still don't have proper estimates of error for demographic estimates at the national level.

The final reference to a possible 'hidden breakdown' in the ICM operation smacks of fear-mongering. The Bureau did carry out dress rehearsal tests in 1998, and there is no evidence from them that the ICM component of the tests was more susceptible to breakdown than the "traditional" census component.

### 10. Summary and Conclusions

**Statement**: "[A]djustment adds layer upon layer of complexity to an already complex census. And complex systems, from ocean liners to space telescopes, fail in unexpected ways. The results of adjustment are highly dependent on somewhat arbitrary technical decisions. Furthermore, mistakes are almost inevitable, very hard to detect, and have profound consequences." (p. 18)

**Rebuttal**: First sentence ignores streamlining of other census operations that sampling enables. The last two sentences apply equally to many other parts of census operations. The so-called traditional census in which the authors seem to believe, despite all of the evidence of large levels of error to the contrary, is replete with "failed" components. The lesson from other complex systems is to learn from one's mistakes and produce better systems. We now receive valuable pictures from an improved space telescope, and attempts at technological progress in building ocean liners did not end with the Titanic.

# B Demographic Analysis Revisited

## B.1 Introduction

The U.S. Census Bureau estimated completeness of coverage of the 1990 census using two approaches–the post enumeration survey (PES) and demographic analysis (DA). Results from both approaches lead one to conclude that the 1990 census underenumerated the population. However, both approaches have problems. The problems and controversies surrounding the PES are well-documented elsewhere. This appendix focuses on the demographic analysis conducted by the U.S. Census Bureau under the direction of Gregory Robinson.

Competent demographers performed the DA conducted on the 1990 census and they built on a broad effort from the preceding decade associated with the demographic analysis for the 1980 census (e.g., see Fay, et al. [28]). The amount of work and thought that went into the analysis is astounding. What follows is a brief synopsis of the demographic analysis carried out by the Census Bureau on the 1990 census and a discussion regarding how demographic analysis might be better used in the future. For a detailed account of the 1990 DA methodology see Robinson et al., [58], and the discussion that follows it, along with the original Bureau D-reports that were produced in the spring of 1991, many of which we reference below.

## B.2 Synopsis

At the heart of demographic analysis is the demographic balancing equation:

$$P(t_1) = P(t_0) + \text{Births} - \text{Deaths} + \text{Net Migration}$$

where $P(t_0)$ = the size of the population at time $t_0$. Demographic analysis uses data from previous censuses and a detailed analysis of administrative records and vital statistics to arrive at the population at a later census date.

While demographers might agree on the overall approach known as demographic analysis–that the population estimated from the balancing equation and the enumeration can and should be compared–demographers differ with respect to exactly how the analysis should be carried out. Rather than quibble with the minutiae of the DA that Robinson and his colleagues carried out, we intend to point out where our opinion substantially differs from Robinson's.

It is important to note that depending on the particular emphases placed on various components of the estimation procedure, different figures could result. That is, some demographers may focus on race first, sex second, and age structure last, while others may believe, for example, that the age structure is the most salient issue in population estimates and that race plays a relatively minor role.

## B.3 Race and Ethnicity

In conducting its demographic analysis, the U.S. Census Bureau placed its primary emphasis on racial differences in the enumeration. Throughout the DA, race is the most salient aspect of the analysis. Apparently, DA was conducted separately for each of the racial/ethnic groups that the Census Bureau found relevant. In order to do this, the racial categorizations across data sets first needed to be massaged so that the data would be racially comparable.

Demographic analysis relies upon data collected in earlier censuses, vital registration systems, and administrative record systems. "There are known differences between the reporting of race in these data sets and the concepts employed in the 1990 censuses. . . . Therefore, adjustments need to be made to achieve comparability between the racial grouping in the population as estimated from demographic analysis and as classified in the census." (Robinson, Word and Spencer, [60] p. 1)

Massaging the racial categories, in our view, is a difficult and theoretically questionable task. Essentially, the "massaging" re-classified individuals from one racial category to another. While this may have been possible with some acceptable degree of "accuracy" in the past, one can imagine that as racial categories become, in a sense, continuous rather than discrete, the task of pigeonholing each individual into one discrete category becomes futile.

Rather than working with race, demographers need to think about what race represents and work with the variable that most closely operationalizes the relevant concept. For example, if race is currently being used as a proxy for social disenfranchisement, then in the future, as racial categories become more blurred, another variable, such as income or home ownership, may better operationalize the concept of disenfranchisement.

## B.4 Base Population

The base population represents "the legally resident population for 1980 . . . revised to take account of new data relating to the estimate for 1980 and results of research that has been carried out over the past several years." (Robinson, [55], p. 1). In each of the steps of the revision to the 1980 population, which include "adjustment for overstatement of black birth cohorts . . . [and] use of revised estimates of foreign students in 1980," uncertainty exists. Demographers may differ on the extent to which the base population should be adjusted for under- or overenumeration.

## B.5 Births

In the United States, the number of births can be known with a reasonable amount of certainty. Comparing the number of births from vital statistics with the enumerated population leads us to conclude that the 1990 census did not count about 18% of all infants (less than 1 year of age). The enumerated population of infants was 3,204,504 (U.S. Bureau of the Census, Database C90STF3C1),[3] but according to vital statistics, 4,158,212 births occurred in 1990 and 4,040,958 occurred in 1989 (National Center for Health Statistics, [31]). Strangely, this severe undercount of infants is not mentioned in the Census Bureau's Demographic Analysis.

According to Larry Sink of the U.S. Census Bureau (personal communication, October 3, 1997), the Census Bureau has been well aware of the infant undercount problem and primarily attributes it to age misreporting of infants. For example, parents of a 3-month old might have reported the infant as 3 years of age. Because the DA aggregated the 0-4 years age group, it is impossible to see whether the infant undercount problem in particular was addressed. There is no mention in particular of the infant undercount problem throughout the DA documentation.

## B.6 Deaths

Throughout the Census Bureau's DA, the number of deaths is assumed to be well known. An analysis conducted by the Census Bureau suggest that deaths are under-registered by 0.6 to 1.4 percent, depending on the race-sex group. There appear to be problems with the registration of infant deaths (Ahmed, Unpublished memorandum, March 27, 1991, p. 1) especially of neonatal deaths, with underregistration possibly as high

[3] See http://venus.census.gov/cdrom/lookup/879262146.

as 20 percent. There are issues with the racial classification of deaths, which, depending on the analysis being conducted, contributes to the uncertainty in the results.

### B.7 Migration

The component with probably the highest degree of uncertainty is migration. "The last step in the development of the estimated total resident population in 1990 is to add an estimate of the number of undocumented residents in the United States." (Robinson, [55], p. 2) The Census Bureau went through great pains in adjusting and massaging INS data to obtain an estimate of both legal and illegal migrants. However, the estimates were ultimately the output of the demographers' subjective view of the phenomena.

### B.8 Uncertainty

Robinson, Woodrow, and Ahmed, [59], examined uncertainty in various components of the DA: underregistration of deaths, legal immigration, net migration from Puerto Rico, net migration of native-born citizens, net movement of armed forces overseas, foreign students, and mortality rates before 1935. They write

> It is not possible to validate many of the assumptions underlying the analyses in this study. Professional judgment plays a major role in specifying the models and setting the upper and lower bound ranges that are needed to carry out the evaluations. Other researchers could incorporate their own set of assumptions that might lead to different conclusions. However, we believe that reasonable assumptions are used here.

Various other D-reports (e.g., [57], [56], [60], [41]) also examined uncertainty. Another way of addressing uncertainty would have been to use a Bayesian framework and having the demographer be explicit about the sources and extent of uncertainty.

### B.9 Discussion

The use of the Census Bureau's demographic analysis conducted on the 1990 census seems to have taken on a life of its own. The Census Bureau originally carried out the analysis to document whether undercount existed. Some statisticians, however, seeing the limitations of the PES, point to the demographic analysis as providing very reliable and more accurate estimates of the population than the PES approach on the national level.

The DA conducted by the Census Bureau may appear to be reliable because of a lack of debate among demographers about the approach and the assumptions used. Actually, though, demographers can and do differ on the exact approach to be taken and emphases in the analysis. Perhaps because only one version of DA was conducted by the Census Bureau, it seems that statisticians and policymakers have placed more faith in it than perhaps it deserves. If DA was held up to the same amount of scrutiny as the PES approach was, there might be as much contention surrounding DA.

To understand how demographers truly view the population and the census, in the future, many different "demographic analyses" should be conducted. Conducting many different demographic analyses would allow policymakers and statisticians to see the extent of agreement between demographers and the limitations of the approach.

We suggest that a Bayesian approach be used by the demographers conducting DA. The census bureau could employ demographers to examine only one of the components of population change apiece and then aggregate the picture. However, in the examination of each of the components and factors to disaggregate by, demographers have to be very explicit about the sources and extent of uncertainty. This method has been used by Daponte, Kadane, and Wolfson [18] in a different context.

Having independent DAs conducted would also allow for comparison between the results of the various demographic analyses and allow one to get a sense of the robustness of the estimates. Indeed, because demographers differ on their analytic approach, and because only one "demographic analysis" was performed by the Census Bureau to evaluate the 1990 census, we contend that the demographic analysis approach was not used to its fullest.

We recommend that in the future, demographic analysis be carried out independently by a number of demographers who are knowledgeable about population processes. We envision a large project that would use a number of demographers. These demographers could ask the Census Bureau to provide various statistics based on administrative and non-administrative data. The important aspect of this is that there be independence among the demographers. We expect that this would result in a number of different estimates of undercount and of the actual population.

Finally, we briefly outline an alternative DA that might have been conducted:

1. We would have first concentrated on age as the most important aspect of the population, rather than race. It is only after one is confident that the estimated population for all races combined, by age and sex is relatively accurate that one should concern oneself with racial classifications and disaggregations.

2. Expected Sex Ratio analysis: We would have liked to have seen an analysis conducted which compared the sex ratios obtained by DA with the sex ratios implied by survivorship ratios over time. Robinson et al incorrectly use the term "expected sex ratio" to refer to the sex ratio obtained from the DA population. By comparing the DA sex ratios with those expected from survivorship ratios, one can ask whether the sex ratios from the DA population fall within a range of uncertainty of mortality differentials between men and women.

3. DA was conducted only on the national level, with the explanation that interstate migration makes conducting DA for a more local level too unreliable. While this may be the case for all age groups, there are certain age groups on which data exists that would have allowed DA to have been conducted locally (at the state, county group, or county level). For example, one could compare the local infant population from the census with the number of births obtained from vital statistics. The local population 80+ from the census can be compared with the local population 80+ from Medicare records. The local population 6-14 from the census can be compared with local school enrollment figures. These various data sources are easily obtainable on an annual basis. It would have been interested to compare the inclusion rates obtained at the local levels for these age groups from DA with those obtained from the PES.

In conclusion, while the DA effort was considerable, we believe that the method was not used to its fullest because of its limitation to only one analysis, and because it did not utilize local data. By not conducting a limited DA at the local level, an opportunity was missed to compare the net undercount estimated from DA with that using the PES.

# C Donner Project Activities

## C.1 Project Participants

The principal Carnegie Mellon participants in this project were Stephen E. Fienberg and Joseph B. Kadane, and their efforts were assisted by Duane Steffey of San Diego State University and Beth Osborne Daponte of the University of Pittsburgh. Several graduate students sat in for some of the conversations. The following scholars made special visits to Carnegie Mellon as part of this project:

- Nov. 16, 1998: Professor Donald Ylvisaker, Department of Statistics, University of California at Los Angeles
- Dec. 4, 1998: Professor M.L. Eaton, Department of Theoretical Statistics, School of Statistics, University of Minnesota
- Dec. 9, 1998: Professor Bruce Spencer, Department of Statistics, Northwestern University
- Dec. 9, 1998: Dr. Mary Mulry, The MARC Group, Irving TX (formerly of the U.S. Census Bureau)
- Dec. 17, 1998: Dr. Jeffrey Passel, Urban Institute (formerly of the U.S. Census Bureau)
- Jan. 21, 1999: Professor Larry Brown, Department of Statistics, University of Pennsylvania

The following scholars were invited to visit us, but were not able to find the time to do so:

- Professor Professor David Freedman, Department of Statistics, University of California Berkeley
- Dr. Howard Hogan, U.S. Bureau of the Census
- Professor Samuel Preston, Department of Sociology, University of Pennsylvania
- Professor Philip Stark, Department of Statistics, University of California Berkeley
- Professor Kenneth Wachter, Department of Statistics, University of California Berkeley
- Professor Martin T. Wells, Cornell University

In addition to those who came to Carnegie Mellon as part of this project, we had several extended discussions with Professor Margo Anderson of the University of Wisconsin-Milwaukee, and we had an opportunity to examine unpublished documents and forthcoming publications by the staff at the U.S. Bureau of the Census and Kenneth Darga.

## C.2 Questions and Issues Relevant To a Re-evaluation of Data and Methods for the 1990 and 2000 Censuses

As part of this project we agreed to carry out several tasks and/or address several questions that were raised as part of the original request for proposals by the Donner Foundation. The numbering in the following list represents that of the Donner Foundation. For each item we have addressed the issues in detail in the main body of this report and in the other two appendices. Here we provide brief responses that summarize our position after reviewing the relevant documentation and materials.

1. After reviewing the papers and reports compiled by the proponents and opponents of sampling compiled both prior to and after the 1990 Census, including the records of the post 1990 Census court cases, can it be rigorously confirmed whether or not the Secretary of Commerce was correct in his decision not to adjust the 1990 Census?

    We reviewed the Secretary's decision at length and the technical basis on which it rested. We, and most of those who visited with us as part of this project, agreed that the criteria for deciding whether or not to use adjusted census counts, outlined in the Federal Regulations by the Secretary of Commerce, are biased against adjustment and fundamentally flawed in other ways. The Secretary's report translated these into formal decision criteria and as a consequence some of them are also technically flawed. Nonetheless, even if we accept the way that the Secretary's report laid out the criteria, our judgment is that the data from the PES evaluations and other sources *fail* to support his decision not to adjust.

2. Evaluations of the arguments proposed by the opponents of sampling in 1990, accompanied by discussion of whether or not these objections were effectively addressed by the proponents of sampling in the 2000 Census.

    The criticisms regarding the use of sampling for the postenumeration survey in 1990 are summarized in the main part of this report and in the technical report by Brown et al., [12], who essentially suggest that every issue which was raised regarding 1990 is relevant to the integrated coverage measurement (ICM) survey in 2000 as well. Our view is that this is simply wrong. Many concerns expressed in connection with 1990 were misleading or irrelevant, and others have been carefully and thoughtfully addressed in the Census 2000 design.

    Moreover, sampling for non-response followup (SNRFU) is a new component in the Census 2000 plan, and criticisms from 1990 are not relevant to it. The principle criticism of SNRFU are that "sampling can be manipulated"—a view that is unscientific and not supportable on any grounds we have seen. The other criticisms suggest that there is little to be gained from SNRFU regarding either savings or technical improvement in accuracy. We disagree, and have explained how SNRFU can actually improve accuracy.

3. Comparisons of the arguments proposed by the proponents of sampling in 1990 and 2000. How have these arguments changed during the past decade? Have they improved in their scientific rigor? If they have improved in their scientific rigor, is this improvement sufficient to warrant the use of sampling in the 2000 Census?

    Again we need to distinguish between sampling for adjustment (the PES in 1990 and the ICM in 2000) and sampling for nonresponse followup (SNRFU in 2000); we address only the former here. The basic argument for the use of sampling for adjustment purposes remains the same. The scientific rigor was not especially lacking in 1990, and it has been strengthened by targeted investigations during the intervening decade. Moreover, the criticisms of the 1990 PES were often either technically flawed or misleading. We elaborate on this point in Appendix A. In particular, the criticisms focus on the net national undercount and seemingly ignore the extent of error in the census and its import.

Thus it is reasonable to conclude that the rigor of the arguments in support of the use of sampling for adjustment in census 2000 is greater than it was in 1990. Since we believe that the use of sampling in 1990 was rigorously justifiable, we of necessity conclude, that it is even more so for Census 2000.

4. Any methodology has inherent strengths and weaknesses. What are the theoretical strengths and weaknesses of the proposed sampling methodology as compared to more traditional methodologies?

   Scientific random sampling is so well established that it hardly needs a defense. Given a sampling frame, statisticians are well used to drawing samples, gathering data, making estimates, and characterizing the resulting uncertainty. In this application, the only issue that could remain is the way in which the sampling data and estimates are used.

   Population estimates produced by the so-called traditional census design are essentially free of sampling error, but traditional methods suffer from substantial biases that result in undercoverage and differential coverage of the population. The sampling methodology, particularly integrated coverage measurement (ICM), proposed as part of the 2000 census would introduce a degree of sampling error and a relatively small amount of nonsampling error from the additional survey and matching procedures in exchange for a substantial reduction of the coverage biases inherent in the traditional census operations. ICM would achieve this improvement in accuracy not only by detecting and correcting coverage errors but also by permitting the reengineering of census operations, thereby reducing the level of error introduced at earlier stages of the process. In our judgment, as well as the judgment of many scientific advisory panels (e.g., most recently, Cohen, White, and Rust [17]), the quality of the population estimates produced by the 2000 census design would be superior to those produced by a traditional census design.

7. Does the sampling methodology proposed for use in the 2000 Census, Integrated Coverage Measurement (ICM), significantly (meaningfully) differ from the Post Enumeration Survey (PES) methodology attempted by the Census Bureau in the 1990 Census? If so, how does the ICM methodology overcome the inaccuracies contained in the 1990 PES methodology?

   There are several reasons to believe that the proposed methodology for the ICM in the original Census 2000 plan represents an improvement over the design of the PES in 1990:

   (a) By structuring the ICM in the context of a one-number census plan with SNRFU, the Census Bureau proposed to make the operational aspects of carrying out the ICM much improved over those for the PES in 1990, and this would have reduced various errors, especially those associated with the timing of the survey.
   (b) The increase in size of the PES from 150,000 to 750,000 households would have reduce sampling error in the ICM and thus improve accuracy.[4]

[4] As we completed this report, the Census Bureau announced another revision in its Census 2000 plan which was a consequence of the Supreme Court ruling that sampling could not be used for apportionment. The Bureau plan as of February 24, 1999 is to carry out a PES of only 300,000 households. Even this two-fold as opposed to five-fold increase in size of the PES households would reduce sampling error in the ICM and thus improve accuracy.

(c) The replacement of the smoothing models with direct state estimates through state-level stratification schemes directly responds to other criticisms of inaccuracy in the 1990 PES. As we explained in the main part of this report, we believe that this is a concession to political parochialism that decreases the accuracy of the resulting estimates.

(d) Detailed changes in plans for movers, the handing of matching, and the stratification scheme all address accuracy in ways that should lead to additional improvements.

Even in the absence of SNRFU, now that the Supreme Court has ruled that it is precluded by the language of the Census Act, and with the reduced ICM sample size of 300,000 households just announced by the Census Bureau, our judgment is that the methodology for 2000 represents the prospect of considerably improved accuracy over that achieved in 1990.

9. Prepare a review of the Census Bureau's and other relevant publications describing the Post Enumeration Survey and Integrated Coverage Measurement methodologies.

   The main part of this report describes some of these publications. A related bibliography can be found in Anderson and Fienberg [5].

## C.3 Materials Reviewed

During the five separate visits by outside scholars, and in a series of separate meeting we reviewed a variety of the primary documents relating to the 1990 census and PES, and to the plans for Census 2000. Among these were Secretary Mosbacher's 1991 report and its various appendices [49]; the Census Bureau's 1991 P-studies and D-studies; the Census Bureau CAPE report [68]; the various critiques of 1990 described in Appendix A; the papers in the special 1993 section on the 1990 Census trial in *Jurimetrics* [30], [61], [33], [73]; the papers in the special 1993 section of *The Journal of the American Statistical Association* [1], [6], [9], [20], [42], [43], [58], [79]; a series of reports from the Committee on National Statistics census panels [53], [66], [24], [76], [17].

We reference other sources and materials explicitly in the Report and other appendices where they seem relevant.

## References


[1] Juha M. Alho, Mary H. Mulry, Kent Wurdeman, and Jay Kim. Estimating heterogeneity in the probabilities of enumeration for dual-system estimation. *Journal of the American Statistical Association*, 88:1130–1136, Sep 1993.

[2] Margo Anderson. *The American Census: A Social History*. Yale University Press, New Haven, 1988.

[3] Margo Anderson and Stephen E. Fienberg. The History of the First American Census and the Constitutional Language on Censustaking: Report of a Workshop. Technical report, Department of Statistics, Carnegie Mellon University, February 1999.

[4] Margo Anderson and Stephen E. Fienberg. To sample to not to sample: The 2000 census controversy. *Journal of Interdisciplinary History*, 30(1):1–36, 1999.

[5] Margo Anderson and Stephen E. Fienberg. *Who Counts? The Politics of Census-Taking in Contemporary America*. Russell Sage, 1999.

[6] Thomas R. Belin, Gregg J. Diffendal, Steve Mack, Donald B. Rubin, Joseph L. Schafer, and Alan M. Zaslavsky. Hierarchichal logistic regression models for imputation of unresolved enumeration status in undercount estimation. *Journal of the American Statistical Association*, 88:1149–1166, Sept. 1993. Comment by Little, Roderick J. A. ; Comment by Wachter, Kenneth W. ; Rejoinder by Belin, Thomas R. and Diffendal, Gregg J. and Mack, Steve and Rubin, Donald B. and Schafer, Joseph L. and Zaslavsky, Alan M.

[7] Thomas R. Belin and John E. Rolph. Can we reach consensus on census adjustment? *Statistical Science*, 9(4):486–508, 1994.

[8] Thomas R. Belin and John E. Rolph. Rejoinder to “Can we reach consensus on census adjustment?”. *Statistical Science*, 9(4):520–521, 1994.

[9] William R. Bell. Using information from demographic analysis in post-enumeration survey estimation. *Journal of the American Statistical Association*, 88:1106–1118, Sep 1993.

[10] Leo Breiman. The 1991 census adjustment: Undercount or bad data. *Statistical Science*, 9(4):458–475, 1994.

[11] Leo Breiman. Rejoinder to the discussion of “The 1991 census adjustment: Undercount or bad data”. *Statistical Science*, 9(4):521–527, 1994.

[12] Lawrence D. Brown, Morris L. Eaton, David A. Freedman, Stephen P. Klein, Richard A. Olshen, Kenneth W. Wachter, Martin T. Wells, and Donald Ylvisaker. Statistical controversies in Census 2000. Technical Report 357, Department of Statistics, University of California Berkeley, October 1998.

[13] Barbara Bryant and William Dunn. *Moving Money and Power: The Politics of Census Taking*. New Strategist Publications, Inc., Ithaca, NY, 1995.

[14] C. Y. Choi, D. G. Steel, and T. J. Skinner. Adjusting the 1986 Australian census count for underenumeration. *Survey Methodology*, 14:173–189, 1988.

[15] Harvey Choldin. *Looking For the Last Percent: The Controversy Over Census Undercounts*. Rutgers University Press, New Brunswick, NJ, 1994.

[16] Constance F. Citro and Michael L. Cohen, editors. *The Bicentennial Census. New Directions for Methodology in 1990*. National Academy Press, Washington, DC, 1985.

[17] Michael L. Cohen, Andrew A. White, and Keith F. Rust, editors. *Measuring a Changing Nation: Modern Methods for the 2000 Census*. National Academy Press, Washington, DC, 1999.

[18] Beth O. Daponte, Joseph B. Kadane, and Lara J. Wolfson. Bayesian demography: Projecting the Iraqi Kurdish population, 1977-1980. *Journal of the American Statistical Association*, 92(4):1256–1267, 1997.

[19] Kenneth Darga. *Sampling and the Census: A Case Against the Proposed Adjustments for Undercount*. American Enterprise Institute, Washington, DC, 1999.

[20] John N. Darroch, Stephen E. Fienberg, Gary F. V. Glonek, and Brian W. Junker. A three sample multiple-recapture approach to census population estimation with heterogeneous catchability. *Journal of the American Statistical Association*, 88:1137–1148, 1993.

[21] Ian Diamond. Where and who are the missing million? In *Regional and Local Statistics: Statistics Users Council Conference*, pages 132–145, Esher, 1993. IMAC Research.

[22] Ian Diamond and Chris Skinner. Comment on three papers on census adjustment. *Statistical Science*, 9(4):508–515, 1994.

[23] Peter Dick. Modelling net undercoverage in the 1991 Canadian Census. *Survey Methodology*, 21:45–54, 1995.

[24] Barry Edmonston and Charles Schultze, editors. *Modernizing the U.S. Census, Panel on Census Requirements in the Year 2000 and Beyond*. National Academy Press, Washington, DC, 1995.

[25] Eugene P. Ericksen and Teresa K. DeFonso. Beyond the net undercount: How to measure census error. *Chance*, 6(4):38–43, 1993.

[26] Eugene P. Ericksen, Stephen E. Fienberg, and Joseph B. Kadane. Comment on three papers on census adjustment. *Statistical Science*, 9(4):511–515, 1994.

[27] Robert Fay. Inferences for small domain estimates from the 1990 post enumeration survey. Technical report, U.S. Bureau of the Census, Washington, DC, 1992.

[28] Robert E. Fay, Jeffrey S. Passel, J.G. Robinson, and Charles D. Cowan. *The Coverage of the Population in the 1980 Census, Bureau of the Census*. U.S. Department of Commerce, Washington, DC, 1988.

[29] Robert E. Fay and John Thompson. The 1990 Post Enumeration Survey: Statistical lessons, in hindsight (disc: P92-96). In *Proceedings of the Bureau of the Census Annual Research Conference*, pages 71–91, 1993.

[30] Stephen E. Fienberg. The New York City census adjustment trial: Witness for the plaintiffs. *Jurimetrics Journal*, 34:65–83, 1993.

[31] National Center for Health Statistics. *Vital Statistics of the United States, 1990.* Public Health Service, Washington, DC, 1994.

[32] David A. Freedman. Policy forum: Adjusting the 1990 census. *Science*, 252:1233–1236, 1991.

[33] David A. Freedman. Adjusting the census of 1990. *Jurimetrics Journal*, 34:99–106, 1993.

[34] David A. Freedman and William C. Navidi. Regression models for adjusting the 1980 census (with discussion). *Statistical Science*, 1:3–39, 1986.

[35] David A. Freedman and William C. Navidi. Should we have adjusted the U.S. census of 1980? (with discussion). *Survey Methodology*, 18:3–24, 1992.

[36] David A. Freedman and Kenneth W. Wachter. Heterogeneity and census adjustment. *Statistical Science*, 9(4):476–485, 1994.

[37] David A. Freedman and Kenneth W. Wachter. Rejoinder to "Heterogeneity and census adjustment". *Statistical Science*, 9(4):527–537, 1994.

[38] David A. Freedman, Kenneth W. Wachter, Daniel C. Coster, D. Richard Cutler, and Stephen P. Klein. Adjusting the census of 1990: The smoothing model. *Evaluation Review*, 17:371–443, 1993.

[39] David A. Freedman, Kenneth W. Wachter, D. Richard Cutler, and Stephen P. Klein. Adjusting the U.S. census of 1990: Loss functions. *Evaluation Review*, 18:243–280, 1994.

[40] E. D. Goldfield. Innovations in the decennial census of population and housing: 1940-1990. Paper prepared for the Committee on National Statistics, Commission on Behavioral and Social Sciences and Education, National Research Council, Washington, DC, 1992.

[41] Prithwis Das Gupta. Models for Assessing Errors in Undercount Rates Based on Demographic Analysis. Demographic Analysis Evaluation Project, No. 84: Preliminary Research and Evaluation, Bureau of the Census, 1991. Report D-11.

[42] N. Hengartner and T. P. Speed. Assessing between-block heterogeneity within the post-strata of the 1990 post-enumeration survey estimation. *Journal of the American Statistical Association*, 88:1119–1129, Sep 1993. Comment by Schafer, Joseph L. ; Comment by Ylvisaker, Donald ; Rejoinder by Hengartner, N. and Speed, T. P.

[43] Howard Hogan. The 1990 post-enumeration survey: Operations and results. *Journal of the American Statistical Association*, 88:1047–1060, 1993.

[44] Howard Hogan and Kirk M. Wolter. Measuring accuracy in a post-enumeration survey. *Survey Methodology*, 14:99–116, 1988.

[45] Joseph B. Kadane, Michael M. Meyer, and John W. Tukey. Corellation bias in the presence of stratum heterogeneity. An expanded and revised version with appear in *Journal of the American Statistical Association*, 1999, 1992.

[46] Ruth Ann Killion. The impact of housing unit coverage on person coverge. Statistical Research Division, U.S. Bureau of the Census, Washington, DC, June 24 1993. Memorandum for Thomas C. Walsh.

[47] Ian Mitroff, Richard O. Mason, and Vincent Barabba. *The 1980 Census: Policymaking Amid Turbulence*. Lexington Books, Lexington, MA, 1983.

[48] David S. Moore. Commentary. Another opinion: Statistical methods in the census. *Notices of the American Mathematical Society*, 1999.

[49] Robert A. Mosbacher. *Decision on Whether or Not a Statistical Adjustment of the 1990 Decennial Census of Population Should Be Made For Coverage Deficiencies Resulting in an Overcount or Undercount of Population: Explanation*. Government Printing Office, Washington, DC, 1991. U.S. Department of Commerce, Office of the Secretary.

[50] Mary H. Mulry. Total Error in PES Estimates by Evaluation Post Strata. 1990 post enumeration survey evaluation project, r-6: Statistical support division, Bureau of the Census, 1991. Report P-16a.

[51] Mary H. Mulry and Bruce D. Spencer. Accuracy of the 1990 census census and undercount adjustments. *Journal of the American Statistical Association*, 88:1080–1092, Sep 1993.

[52] Janet L. Norwood. *Organizing to Count: Change in the Federal Statistical System*. Urban Institute Press, 1995.

[53] Panel on Decennial Census Plans. *Counting the People in 1980: An Appraisal of Census Plans, Committee on National Statistics, National Research Council, National Academy of Sciences*. National Academy of Sciences, Washington, DC, 1978.

[54] Daniel O. Price. A check on underenumeration in the 1940 census. *American Sociological Review*, 12:44–49, Feb 1947.

[55] J. Gregory Robinson. Memorandum for Paula J. Schneider, Chief, Population Division, June 4 1991.

[56] J. Gregory Robinson. Uncertainty Intervals for Estimated Black Births, 1915–1934. Demographic Analysis Evaluation Project, No. 77: Preliminary Research and Evaluation, Bureau of the Census, 1991. Report D-4.

[57] J. Gregory Robinson. Uncertainty Intervals for Estimated White Births, 1915–1934. Technical report, Bureau of the Census, 1991. Report D-3.

[58] J. Gregory Robinson, Bashir Ahmed, Prithwis DasGupta, and Karen A. Woodrow. Estimation of population coverage in the 1990 United States Census based on demographic analysis (with discussion). *Journal of the American Statistical Association*, 88:1061–1079, 1993.

[59] J. Gregory Robinson, Karen A. Woodrow, and Bashir Ahmed. Uncertainty Measure for Other Components. Demographic Analysis Evaluation Project, No. 80: Preliminary Research and Evaluation, Bureau of the Census, 1991. Report D-7.

[60] J. Gregory Robinson, David L. Word, and Gregory S. Spencer. Uncertainty for Models to Translate 1990 Census Concepts into Historical Racial Classifications. Demographic Analysis Evaluation Project, No. 81: Preliminary Research and Evaluation, Bureau of the Census, 1991. Report D-8.

[61] John E. Rolph. The census adjustment trial: Reflections of a witness for the plaintiffs. *Jurimetrics Journal*, 34:85–98, 1993.

[62] E. Schindler. Allocation of the ICM sample to the states for Census 2000. In *Proceedings of the Survey Research Methods Section*, page in press, Alexandria, VA, 1998. American Statistical Association.

[63] C. J. Skinner. The use of synthetic estimation techniques to produce small area estimates. Technical Report New Methodology Series NM18, Social Survey Division, Office of Population Censuses and Surveys, 1991.

[64] P. B. Stark. Differences between the 1990 and 2000 census adjustment plans, and their impact on error. Technical Report 550, University of California, Berkeley, 16 March 1999.

[65] David Steel. Comment on three papers on census adjustment. *Statistical Science*, 9(4):517–519, 1994.

[66] D.L. Steffey and N. M. Bradburn, editors. *Counting People in the Information Age*. National Academy Press, Washington, DC, 1994.

[67] John H. Thompson and Robert E. Fay. Census 2000: The statistical issues. In *Proceedings of the Survey Research Methods Section*, page in press, Alexandria, VA, 1998. American Statistical Association.

[68] United States Bureau of the Census. Assessment of accuracy of adjusted versus unadjusted 1990 census base for use in intercensal estimates, 1992. Report of the Committee on Adjustment of Postcensal Estimates, August 7, 1992.

[69] United States Bureau of the Census. Report to Congress — The plan for census 2000, 1997. Typescript February 28, 1996; Originally Issued July, 1997; Revised and Reissued August, 1997.

[70] United States Government Accounting Office. *1990 Census: Reported Net Undercount Obscured Magnitude of Error, GAO/GGD-91-113*. Government Accounting Office, Washington, DC, 1991.

[71] United States Government Accounting Office. *2000 Census: Preparations For Dress Rehearsal Leave Many Unanswered Questions*. Government Accounting Office, Washington, DC, 1998.

[72] United States House of Representatives et al., vs. United States Department of Commerce, et al.; Glavin, Barr et al., vs. Clinton et al. and as argued before the Supreme Court: Department of Commerce v. House of Representatives, 11 F. Supp. 2d 76, appeal dismissed; No. 98-564, 19 F. Supp. 2d 543, affirmed.

[73] Kenneth W. Wachter. The census adjustment trial: An exchange. *Jurimetrics Journal*, 34:107–115, 1993.

[74] Kenneth W. Wachter. Comment on "Hierarchical logistic regression models for imputation of unresolved enumeration status in undercount estimation". *Journal of the American Statistical Association*, 88:1161–1163, 1993.

[75] Preston J. Waite and Howard Hogan. Statistical methodologies for Census 2000: Decisions, issues, and preliminary results. In *Proceedings of the Survey Research Methods Section*, page in press, Alexandria, VA, 1998. American Statistical Association.

[76] Andrew A. White and Keith F. Rust, editors. *Preparing for the 2000 Census: Interim Report II*. Committee on National Statistics, National Research Council, Washington, DC, 1997.

[77] Wisconsin v. City of New York. 116 S. Ct. 1091 (1996).

[78] K. M. Wolter. Some coverage error models for census data. *Journal of the American Statistical Association*, 81:338–346, 1986.

[79] Alan M. Zaslavsky. Combining census, dual-system, and evaluation study data to estimate population shares. *Journal of the American Statistical Association*, 88:1092–1105, Sept. 1993.