# Sampling-Based Adjustment of the 2000 Census–A Balanced Perspective


Margo Anderson
Department of History
University of Wisconsin-Milwaukee
Milwaukee, WI 53201

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

Stephen E. Fienberg
Department of Statistics and Center for Automated Learning and Discovery
Carnegie Mellon University
Pittsburgh, PA 15213-3890

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

Bruce D. Spencer
Department of Statistics
Northwestern University
Evanston, IL 60208-4070

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


April 17, 2000


**Abstract**

Long before the issue of adjustment for differential undercount of blacks and other minorities arose in connection with the 1980 and 1990 censuses, public controversy had attended the publication of census results. Recently, Brown, et al. [8] critiqued the results of sampling-based adjustments to the 1980 and 1990 censuses. They concluded that the proposed design for Census 2000, which includes an updated version of such adjustment, is too risky. Their account of past practices contains many errors and misleading statements, however, and is deeply flawed. Here we attempt to provide a more balanced perspective of past practices and the fascinating statistical and legal issues that Census 2000 raises.

## Contents

In the decade after the 1990 census, the US Census Bureau's mantra was a "one-number census," meaning that for each geographical level (nation, state, city, census block) there would be a single census result. This was the impetus for what the Bureau called the Integrated Coverage Measurement program (ICM), which would have incorporated sampling for non-response follow-up and a postenumeration survey into the census counts. The result of the *Glavin v. Clinton*[1] case is that, for the purpose of apportionment among the states, sampling is not to be used, by interpretation of the Census Act.[2] That same act as it was revised in 1976, however, requires sampling to be used for all other census purposes, if feasible. Because in this recent decision the Supreme Court did not express any judgment regarding this provision, and because statisticians at the Census Bureau believe that use of sampling following the regular enumeration process will improve the accuracy of the census numbers for other purposes than apportionment, the Bureau plans to proceed with that component of its original sampling plan. Thus the Census Bureau is confronted in 2000 with the result it sought to avoid–a two-number census.

The first consequence of the two-number census of 2000 is that the Census Bureau was allocated an extra two billion dollars to do the extra work. More consequences will ensue once the census is complete, at the end of December, 2000. The Census Bureau expects to release both the original census enumeration counts and the sampling-based adjusted counts on April 1, 2001, and each state legislature will have to choose one or the other set of figures to use for establishing legislative and congressional districts within the state. The legislatures are unfortunately very likely to decide the matter on partisan grounds, with Republicans favoring the unadjusted numbers, and Democrats the adjusted ones. To many if not most Democrats, the issue is civil rights, focused specifically on the chronic differential net undercount of blacks in each census, at least since 1940.[3] When the constitution was signed, each black, as a slave, counted for 60% of a white person. An unadjusted census historically has raised the figure to about 95%.[4] Democrats hope that the sampling adjustment of the Census of 2000 will close at least some of this racial gap. Republicans, fearing that unseen bureaucrats may be able to skew the sampled numbers, favor unadjusted numbers. The perception that the areas typically undercounted in the census–inner-city minority areas–tend to elect Democrats, while the areas typically overcounted–white, rural areas–tend to vote Republican, might have something to do with the partisan stance on the sampling and adjustment issue.

Whatever a state legislature chooses to do, litigation is almost sure to result. Those attacking sampling will cite the Constitution's language mandating an "actual enumeration," while those defending it will point out that sampling had not been invented at the time, [5] and that the word "actual" was added by the Style Committee without debate by the Constitutional Convention.[6] Additionally, there may be controversy over the extent to which the Post-Enumeration Survey results improve, or disimprove, the raw census counts.

Brown et al. [8] present their case against the PES; our purpose here is to explain why we disagree. The major missing element in the treatment of Brown et al. is the long list of known defects in traditional census enumeration methods. Indeed, these defects are deemed by many to be so substantial and great in their impact on census accuracy that sample-based adjustment (with its known shortcomings) is expected to

[1] 19 F. Supp. 2d 543 (E.D. Va. 1998), *aff'd*, Department of Commerce v. United States House of Representatives, 119 S. Ct. 765 (1999).

[2] Title 13, U.S.C. §195 (1976).

[3] See the detailed historical discussion of this issue in Anderson and Fienberg [3].

[4] The 95% figure corresponds to the approximately 5% differential undercount between blacks and others that has persisted over the 50 years of measurement by the U.S. Bureau of the Census.

[5] see Anderson and Fienberg [1].

[6] See the extensive discussions in Madison's notes on the Constitutional Convention [19].

be a measurable improvement. We respond to what we take to be the major points raised by Brown et al., and explain why they are either wrong or misleading. [7]

# 1 Error in the Census

Census errors occur in two basic types: omissions and erroneous enumerations. What counts as an error, however, depends critically on the level of geographic detail at issue. For example, at the level of states, a person counted in one part of a state when she lives in another would not be regarded as an error, while at the level of census blocks, this could be regarded as two errors – one census block has an omission while another has an erroneous enumeration.

In 1990, there were more than 10 million erroneous enumerations and 15 million omissions in the census.[8] How should these be regarded? For the purpose of ascertaining the national population, it would be appropriate to regard this as a net (national) undercount of 5 million. Brown et al.[9] focus on this quantity for most of their arguments. At the level of census blocks, aggregates of which are relevant for determining legislative and congressional district boundaries, however, it is just as or more appropriate to add the errors, yielding 25 million gross errors. Since the constitutional and political purpose of the Census is apportionment among and within the states, gross error seems the more relevant measure of error. For each specific use of census data, gross error represents an upper bound on the amount of relevant error. But for most of these purposes the bulk of erroneous enumerations tend to occur in different units or domains than do the omissions and so only some of them cancel out. It is only by focusing on net census error and ignoring the real problem of error in the census and how it is distributed that Brown et al. can conclude that the PES of 1990 added more error than it removed.[10]

In looking to the Census of 2000, we agree with the prediction of Brown et al.[11] that erroneous enumerations are likely to increase, because of the emphasis in census operations on reducing omissions by making extra census forms available. A study of the 1980 and 1990 censuses[12] found that many "coverage improvement" programs had added very substantially to erroneous enumerations, even as they reduced omissions. This was why there was so much support in the statistical community for the sampling for non-response followup part of the Bureau's Integrated Coverage Measurement plan; it was widely believed that using sampling for this problem would dramatically reduce the numbers of erroneous enumerations in the 2000 census.

# 2 Criteria for Evaluating Census Accuracy

Brown, et al., discuss two criteria for judging the relative accuracy of unadjusted and adjusted census data: (1) the estimated population counts of places (e.g., states, counties, cities), and (2) the relative population

[7] Daponte, Fienberg, Kadane, and Steffey [12] contains much more detailed treatment of a number of the issues discussed, as well as a point-by-point response to an earlier version of the Brown et al. paper.

[8] See U.S. Census Bureau, "Report to Congress – The Plan for Census 2000" (July 1997) [28]. Brown et al. [8] cite 13 million and 17 million, while Bryant [9], Ericksen and DeFonso [15] give even higher figures. The General Accounting Office [29] provides a range of estimates of gross error.

[9] [8], see especially pp. 348-349, p.361, and p. 374.

[10] Brown, et al. [8], p. 370. The issue not addressed by their misleading net error calculation is the geographic variation of the net undercount at aggregates of blocks used for the drawing of local, state, and congressional political districts.

[11] [8] p. 355, top.

[12] Ericksen et al. [16].

shares of such places.[13] Because political representation and federal revenues are generally allocated to places on the basis of relative population size, they argue that the second criterion is of greater importance.

We agree that a primary goal of decennial census operations should be to obtain the most accurate estimates of population shares that are possible with available methods. But we part company with Brown, et al., in two important respects. First, we do not accept their subsequent claims about the potential inaccuracy of the 1990 adjusted state population shares relative to unadjusted shares. In another section below we consider this issue further and uncover problematic assumptions and misleading interpretations in the alternative loss function analyses presented in their Appendix C, and in a technical appendix we address the matter in detail.

Second, although we concur that accuracy of population shares are of importance, we believe that the estimates of accuracy should take into account 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.[14] This is also an issue of shares in a sense, but it is the shares for demographic groups that are the focus here and not simply the shares for various geographic areas.[15]

The existence of differential undercounting in the census was initially discovered during World War II when a startlingly large number of black males registered for the military draft [16]. The differential undercount of demographic groups has persisted over the past half-century, even as the net undercount has been reduced. Moreover, 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.

Yet correcting the differential undercount was not among the official criteria considered in 1990. In the statement accompanying his July 15, 19991 decision not to use adjusted counts,[17] Commerce Secretary Mosbacher ignored completely the demonstrable improvements adjustment made in the shares for demographic groups and relied heavily on the provision in the Final Guidelines on adjustment issued by his office in 1990 which required that the evidence must show "overwhelmingly" that the count can be improved by statistical adjustment "at all geographic levels."[18] We can profitably separate out several issues here: (1) the overstrictness of the criterion requiring overwhelming evidence at all levels, (2) the focus on states, (3) the lack of attention to demographic groups as relevant units. Asking for overwhelming evidence at all levels simply does not seem reasonable to us. What is the right decision if there is overwhelming evidence at high levels of geography and at worst ambiguity at lower levels? We believe that a single-minded focus on geographic shares and a demand for overwhelming evidence at all levels are not the appropriate way to approach the evaluation of census accuracy in 2000.

The longstanding evidence of the differential undercount, and the experience and evaluations from the 1990 census strongly suggest the relevance of a more balanced approach to assessing the accuracy and improvement associated with sampling-based adjusted counts in the 2000 census.

---

[13] [8], pp. 360-362 and p. 365.

[14] See Steel [26].

[15] For a discussion of some of these issues, see Spencer [24].

[16] Anderson and Fienberg [3], Chapter 2.

[17] Mosbacher [20].

[18] For a relevant excerpt of the language of the final guidelines, see Anderson and Fienberg, Appendix D [3], especially, pp. 253–4.

## 3 Background for Census 2000

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.

After decades of improvement in coverage there were some signs by 1980 that census operations might be approaching a limit in the ability to accurately enumerate the population using only traditional methods. First, the rate at which households returned a completed census form by mail dropped by 13% between 1970 and 1980 as the mail census was extended from 60% to 90+% of households.[19] Second, the extensive array of auxiliary programs intended to improve population coverage in the 1980 census was criticized as being costly, inefficient, and prone to introducing erroneous enumerations into the census process, thus inflating the number of gross errors.[20] 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 the difference remained remarkably consistent from 1940 to 1980.[21]

The 1990 census presented clear evidence that extensive efforts to improve the traditional enumeration methods had failed to resolve census coverage deficiencies. For the first time, the trend of declining net undercount reversed, rising by 50 percent from 1.2% in 1980 to an estimated 1.6% of the 1990 population. Differential undercount persisted, with the Black vs. non-Black differential increasing to 4.4%, the highest level in the six decades for which estimates were available. Mail return rates continued to drop, indicative of declining public cooperation, and unit costs continued to escalate, rising another 25% after almost doubling from 1970 to 1980.

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. The grounds for reaffirmation were not really statistical, however. Rather, they were based on the authority granted to the Secretary of Commerce under the Administrative Procedures Act.

In planning for the 2000 census, the Census Bureau began early this decade to examine alternative methods concurrently with reviews by scientific advisory groups. The concept of integrated coverage mea-

[19] See Edmonston and Schultze, [14].

[20] Citro and Cohen [10].

[21] See Robinson, et al., [22].

surement 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 offered the promise of substantial reductions in the magnitude of differential undercounting, but also would have provided a significant opportunity to re-engineer census operations.[22] Increases in the efficiency of the census design could have been realized by removing coverage improvement programs that have failed to correct the undercount and have provided diminishing marginal returns, often adding gross error to the census and thus complicating the task of adjustment for undercoverage.

The January 1999 Supreme Court decision and congressional resistance will prevent the streamlining of field operations that might have been achieved with the use of sampling in following up nonresponding households and will delay implementation of the Accuracy and Coverage Evaluation (ACE) survey. Nevertheless, expansion of the ACE survey and related improvements are expected to yield adjusted population estimates in 2000 with greater accuracy, relative to the unadjusted counts, than the corresponding 1990 estimates. The ACE survey in Census 2000 will benefit from a substantially increased sample size of 300,000 households–twice as large than the 1990 PES.[23]

Thus, the design changes proposed for accuracy and coverage evaluation (ACE) in 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 "questionable" smoothing models to combine information from multiple states in producing estimates within an individual state. In the first instance improvement has been made, and in the second instance the Bureau has backed away from the use of statistical smoothing models in several parts of the methodology. The proposed 2000 census design will likely draw 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). The anti-adjustment arguments on within-poststratum heterogeneity are all indirect, Further, despite the dependence, either of an intrinsic sort or induced by heterogeneity, there appear to be ample empirical and theoretical arguments to suggest that adjustment moves the census counts in the right direction. Thus, for reasons more fully discussed in our detailed evaluation of the topic,[24] we predict 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.

# 4 Loss Function Analysis of the 1990 Census Data

In the opening section of Brown, et al., the authors consider the 1990 census and claim that "adjustment could easily have made population shares less accurate rather than more accurate, even for states."[25] What

[22] for example, see the recommendations in Edmonston and Schultze[14].

[23] The larger sample size that had been originally proposed for the one number census approach–five times the size of the 1990 PES– would have enabled 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, would have been based only on information collected from that state. This was done largely to cope with apportionment concerns. Following 1999 Supreme Court decision, the Census Bureau simply could not sustain this level of operational effort on sampling and thus reverted to the smaller sample size of ACE, which the Bureau deems sufficiently accurate for adjustment purposes. But the Bureau was also able to relax the politically-imposed constraint that state estimates depend only on the data from that state, and this will permit estimates with greater expected precision despite the reduced sample size.

[24] Daponte et al.[12].

[25] [8], p. 348.

does "could easily" really mean in a technical sense? What appears clear is that their attempt to cast doubt on the relative accuracy of adjusted state population shares relies largely on a loss function analysis presented in their Appendix C as an alternative to the extensive total error analysis carried out by the Census Bureau and reported by Mulry and Spencer. In a short summary of that analysis Brown, et al., argue that, "With other assumptions that are perhaps more realistic [than those used by the Census Bureau], the [risk] estimate becomes $-250$ with a standard error of 821. [The measurement units are parts per 100 million; positive values favor adjustment and negative values favor 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 added more error than it removed, and conclusions depend heavily on assumptions."[26]

The first sentence of this argument is misleading, because the substantial standard error renders the risk difference insignificant. But much more importantly, several technical aspects of the Appendix C loss function analysis are problematic. Some of the authors' underlying assumptions are clearly in error and appear to result from misunderstandings about the original total error analysis. We consider these in detail in Appendix A below. By comparison, the Census Bureau's analysis estimates that use of the unadjusted census numbers involves substantially greater risk. Thus, we conclude that the findings presented in Brown, et al., are based on indefensible assumptions and do not provide credible support for challenging the superiority of adjusted state shares. The Census Bureau and Mulry/Spencer analyses still stand as the standard against which to measure alternatives, and they demonstrate the superiority of the sample-based adjusted counts.

Statistical sampling and estimation methods have been used for more than 20 years to measure and correct for census coverage error in other Western industrialized nations, including England and Australia. 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.

## 5 The Computer Error in the 1990 Adjustment Process

Brown et al. make a big fuss about the computer error that occurred in the processing of the 1990 PES data, and they imply that it would have led to an error in the apportionment of the U.S. House of Representatives:

> The original adjustment would have shifted two congressional seats. The revised adjustment, with the estimated undercount reduced to 4.0 million, would have only shifted one seat. Errors in the PES–the coding error being the dominant factor–would have moved a congressional seat from Pennsylvania to Arizona.[27]

Brown et el. are correct about one thing, that the reapportionment formula is sensitive to small changes in the numbers. But their statement is at best misleading and at worst an intentional misrepresentation of the facts. A careful examination of the state counts that result from the correction of the computer error show that the seat still would have moved.[28] What is true is that a revised PES data set constructed for the

[26] [8], Section VII.

[27] [8], p. 359.

[28] Anderson and Fienberg [4] provide the full calculations for the reapportionment for several different versions of the data from 1990. They also document earlier erroneous statements on this matter by some of the authors of Brown, et al.

purposes of postcensal estimation and to develop sampling frames over the decade, with a differing structure of the data, would not have moved the seat from Pennsylvania to Arizona. An accurate statement would be: "Different poststrata in 1990 could have shifted a seat in Congress, but the computer error would not have." All that this discussion demonstrates is that the apportionment formula is very sensitive to small changes in the population numbers, something that political scientists have known about for years. In fact, a population shift of one person can change the apportionment of the House. In light of this shouldn't we really worry about the overall level of error in the census, i.e., the gross error, which is on the order of 10%?

## 6 Demographic Analysis as an Alternative to the PES

When considering the issue of using "sampling" to adjust the 2000 census, some statisticians, especially opponents of sampling-based methods, have argued that Demographic Analysis (DA) is a promising methodology with which to estimate the magnitude of net undercount. In fact, Brown, et al.[29] directly compared net undercount estimates obtained from the Post Enumeration Survey with those obtained using a demographic analysis approach to argue that adjusting census estimates based on the Post Enumeration Survey "can result in adjustments that make state shares worse rather than better" [30] In this analysis, Brown, et al. assumed that the results from the DA provided the benchmark against which the PES approach should be compared. They do not show the census sex ratios in their table. It turns out that in 1990 adjustment moved sex ratios for blacks towards the sex ratios from DA, but they remained closer to the unadjusted census than to DA. This suggests that adjustment for black males under-adjusted at the national level. [31]

DA uses data from previous censuses and a detailed analysis of administrative records and vital statistics to arrive at what a demographer believes a later census ought to show. To arrive at estimates of net undercount, one then compares estimates obtained from DA with data from the current census.

There is no set way to carry out a demographic analysis. In fact, demographers differ in their opinions on how a demographic analysis of census data ought to be carried out. A demographer's opinions, abilities, and beliefs affect the result of a demographic analysis. For example, some demographers may focus on race first, on sex second, and on age structure last. Other demographers may believe that age structure is the most salient issue in population estimation and that race plays a relatively minor role. Demographers may also differ with respect to how they view the previous census data upon which the DA is based. The point we make is that while DA is a valid and important approach, the question of "Whose demographic analysis?" should be asked. Ultimately, the demographic analytic approach is highly subjective.[32]

The team of demographers that performed the DA for the 1990 Census that Brown, et al. rely upon as a benchmark admit to assumptions being subjective and uncertain. When considering the migration assumptions to use in their DA, Robinson, Woodrow, and Ahmed wrote: "It is not possible to validate many of the assumptions underlying the analyses in this study. Professional judgment plays a major role

[29] [8], p. 361.

[30] [8], p. 362.

[31] This conclusion rests on an assumption that the national DA estimates for blacks are sufficiently accurate to be useful and that classification of race is consistent between administrative records and census reports. The question that Brown, et al. do not grapple with is why an adjustment for part of the black male undercount should be worse than no adjustment.

[32] Spencer [25] describes the Census Bureau's use of subjectivity in preparing official postcensal population efforts. Daponte, Kadane, and Wolfson [11] address the importance of subjectivity in creating population estimates. They demonstrate that uncertainty manifests itself in various aspects of population estimation. DA relies on a wide variety of data sources (e.g., vital statistics, administrative records, migration assumptions). Each of these data sources estimate the demographic issue at hand with varying degrees of accuracy. Daponte, Kadane, and Wolfson advocate using Bayesian inference to allow demographers to be explicit about the sources and extent of uncertainty that exists in estimating a population.

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."[33] Clearly, the demographers carrying out the Census Bureau's DA on the 1990 census believe that there is room for other demographers to carry out an alternative DA.

While DA is a useful approach, it would be most revealing to learn of the actual degree of uncertainty within a demographic analysis performed by one set of demographers and between various demographic analyses performed by different sets of demographers. Having multiple demographic analyses would demonstrate to the statistical community that like the statistical approaches, the demographic analytic approach produces uncertain population estimates.[34]

Because of the lack of detailed information on internal migration and other state-level components, however, DA is generally regarded as unable to provide reliable estimates for such subnational areas as states.[35] Brown, et al. make 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 [36] We find the current and likely future state of demographic analytic 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 Brown, et al.'s call to use DA, by itself, for adjustment at this time.

# 7 Did Adjustment Work in 1990?

As we have already noted, there is compelling evidence from the Census Bureau's evaluations that adjustment corrects, at least partially, the pernicious differential undercounts for various minority groups. Moreover, the loss function analyses show the superiority of adjustment at the level of states and at least for large places within states. No analyses show that census enumerations counts are superior to sample-based adjusted counts, at any level of geography.

An additional argument in favor of adjustment comes to us from Professor Larry Brown, lead author of Brown et al. Brown's argument is that, were it the case that the adjustments are adding solely noise, the change in state shares between the unadjusted and adjusted 1990 censuses would not have a pattern when compared to the change in state shares between 1980 and 1990.

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[37] 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

[33] See Robinson, et al., Report D-8, [23].

[34] Although we cannot easily quantify the error in the actual demographic analyses used by the Census Bureau, we can explore ranges of values based on varying specific assumptions and using differing values from external data sources. In 1990 the Bureau produced a series of reports looking at the error in specific components of the DA, and a listing can be found in Appendix E of Anderson and Fienberg [3].

[35] See Steffey and Bradburn [27].

[36] They describe this simply as "scaling" ([8], p. 369). Darga [13]) similarly suggests that DA data are superior for census purposes to PES-based data.

[37] The line has to go through the origin, (0,0), because shares must sum to 1 and thus changes in shares must sum to 0.

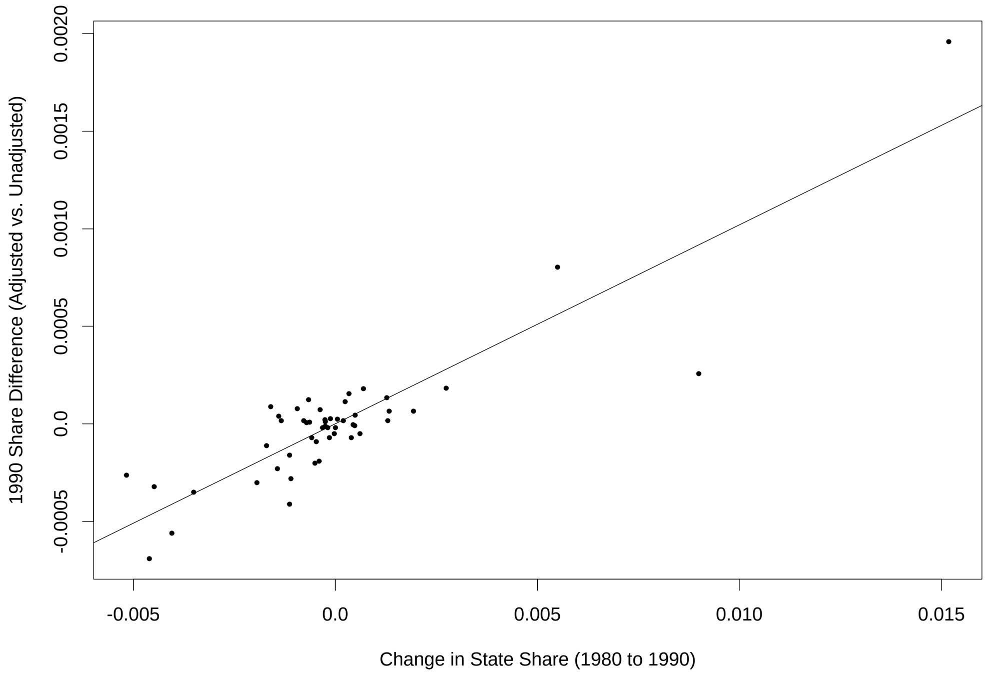


Figure 1: Predictability of Adjusted Shares

adjustment does something quite systematic and credible-namely, adding or subtracting about another 10% to the estimated change in state share from 1980 to 1990.[38]

# 8 The Modified Plan for Census 2000

We agree with Brown et al. that the design of the ACE survey for Census 2000 closely resembles that of the PES for the 1990 census, with the exception that it is double the size and that there will be another decade of experience from working with postenumeration surveys that will influence a number of operational and technical fine points.

We also agree with them that the methodology for the enumeration is complex and prone to error. In fact, there is every reason to believe that the error in the census in 2000 will exceed that from 1990. Brown et al. seem to want to ignore the existence of census error, however, and the strong evidence that we have accumulated over more than 50 years of census evaluations that it differentially impacts upon minority groups. All of the evidence adduced in 1990 and subsequently points to the superiority of sampling-based

[38] 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 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 appropriate, it still provides evidence counter to the claim that adjustment adds more error than it removes.

adjusted census counts. No analysis, by Brown et al. or anyone else, has demonstrated the superiority of the flawed census counts over the adjusted counts.

Thus we agree that Census 2000 is at considerable risk. But we, like most of the professional statistical community that has studied the issue, believe that a sample-based adjustment has the clear possibility of improving the accuracy of the counts in 2000. What actually happens, however, will affect the data used by the nation for years to come.

## Appendix : Revisiting the Technical Details of the Loss Function Analysis

Brown, et al.'s Appendix C contains criticism of three components of the total error analysis: (1) the estimated bias $\hat{\beta}$ of the dual system estimator $X$ of undercount in state population shares; (2) the estimated covariance matrix $\hat{G}$ of the dual system estimator; and (3) the estimated covariance matrix $\hat{H}$ of the bias estimator. Brown, et al., question the assumptions of the Census Bureau's total error model in estimating these three components. We consider, in turn, the criticisms of each of the three components and explain why the Appendix C criticisms are either invalid or without material effect.

Concerning the estimated bias component, Brown, et al., claim that $\hat{\beta}$ is severely biased. This statement lacks support. State-by-state variability may be underestimated, but the data are not adequate to answer conclusively any questions about the bias of $\hat{\beta}$.

More importantly, Brown, et al., mischaracterize the derivation of $\hat{\beta}$: "The Bureau's allocation assumed that (i) bias was proportional to some covariate for post strata within evaluation post strata, and (ii) bias was constant across states within post strata. Bias was allocated down, then reaggregated to the 51 areas." They are are wrong on (i) but correct on (ii).[39] Suppose that the DSE for state A was twice the DSE for state B for every poststratum. Then the component biases for state A would be twice those for state B for every poststratum. Still, the estimated state-level relative bias would not necessarily be the same (even apart from correlation bias) because the components of the DSE would not necessarily be proportional between postrata of the two states.

Brown, et al., apparently obtain the numbers in their Table 2 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. That is, Brown, et al., allocate bias to states in proportion to estimated undercounts. These entries are derived using clearly false assumptions about $\hat{H}$ (see below), which is used in the calculation of the estimated covariance between the risk for different states, and an unsupported and simplistic assumption for assigning PES bias that seems to follow from arguments that focus improperly on net national undercount as the denominator base. What is this latter problem? Throughout their article, Brown, et al., repeat the claim from Leo Breiman [40] that over half the net undercount is due to error. As the discussants of the Breiman article pointed out and others have subsequently noted in length,[41] this claim is highly misleading and at best irrelevant. [42] Thus, there is in effect no empirical support for Brown et al.'s pivotal assumption regarding the reasonableness of the bias allocation.

Concerning the variance components, Brown, et al., state that the matrix $\hat{G}$ is too small by a factor of 2

[39] Mulry and Spencer used direct estimates of the bias for *each* component error for subgroups formed by age, sex, and minority/non-minority. These were ratio-adjusted to aggregate direct estimates of component biases for minority evaluation post-strata and non-minority evaluation poststrata. The component biases for age × sex × minority groups were allocated to component poststrata proportional to the DSE for the poststratum. This allocation of expected component errors (gross errors) is very different than allocating total bias proportionally to poststrata. The only component bias for which this was not done was model bias (correlation bias), which was estimated directly at the poststratum level using models described in Bell [6]. Thus, Brown et al. are wrong with respect to (i).

[40] Breiman [7].

[41] For example, see Ericksen, et. al [17] and Belin and Rolph [5], and the discussion in Anderson and Fienberg [3], p. 295.

[42] A more relevant denominator might be the gross error. Then a corrected version of Breiman's catalogue of biases amount to something on the order of 10% of the gross error in the census. This is a level of error suggesting room for improvement but not an argument for rejecting adjustment.

to 3.[43] Even if one were to concede a factor of 2 or 3, which we are not necessarily ready to do, the change does not significantly alter the original loss function analysis!

Brown, et al., further claim that $\hat{H}$ is off by a rather startling factor of 84, because the Census Bureau's bias allocation scheme converted variance in $\hat{\beta}$ to bias, and the model assumed that bias in $\hat{\beta}$ vanishes.[44] This argument seems to be based on the following premises:

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 concerning $\hat{H}$ in Brown, et al's Appendix C is clearly false. They consider only the ratio of PES and EFU sample sizes and fail to account for the fact that $\hat{G}$ and $\hat{H}$ are estimating different variance components. It is well known that the sampling variance of a sample mean may be expressed as the product of four factors:

- the finite population correction factor;
- the reciprocal of the sample size;
- the design effect or "deff", which accounts for the complexity of the sample design;
- the population variance.

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 $\hat{H}$). The sample size for the evaluation sample was indeed approximately 7% of the PES sample size.

The design effect 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 thus accounts for such complexities as clustering, 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.

The argument in Appendix C relies critically on the assumption that the relevant population variances are the same for the dual system estimator and its bias estimator. The authors present no evidence to show that these variances are the same. On the contrary, there are very plausible reasons why the variances might differ substantially in magnitude. 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

[43] [8], p.374. This statement would appear to be based on an internal census Bureau memorandum by Robert Fay based on his 1992 reanalysis of the PES results, but the discussion in that memorandum is simply speculation and not based on any empirical evaluation of the PES results.

[44] [8], p. 373.

estimate of $\hat{H}$ was developed with tested software based on standard methodology. The estimate developed in the Appendix C critique seems based on unstated assumptions which lack empirical basis.

Finally, Brown et al., additionally state that "Table 2 may, in some respects, understate the problem with loss function analysis. For instance, heterogeneity, which is probably larger than all the other errors combined, is not taken into account in Table 2;"[45] nor, the authors claim, is there any allowance for bias in the imputations.

Both of these allegations are false. The original total error analysis does include heterogeneity (correlation bias), and imputation bias was converted to variance and included.[46] Where is the authors' analysis that treats all sources of error? We have not located it. Thus, we can only conclude that there is no basis for concluding that the results would ultimately favor the census over the PES.

Brown, et al.'s Appendix C concludes with the statement that "[L]oss function analysis depends very strongly on the assumptions going into the evaluation: the available data do not decide the issue." But the alternative analyses presented in Appendix C range from neutral to favorable toward adjustment. Even if one accepts their problematic assumptions as discussed above, in no case do the authors provide evidence that the 1990 adjustment would have made matters significantly worse.

[45] [8], p. 374.

[46] See Appendix A.7 of Mulry and Spencer [21]. Further, when we focus on numerical accuracy, heterogeneity implies undercount estimates that are too small but go in the right direction. See Kadane, Meyer, and Tukey [18].

# References


[1] 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.

[2] 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.

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

[4] Margo Anderson and Stephen E. Fienberg. History, myth making and statistics: A short story about the reapportionment of congress and the 1990 census. Technical report, Department of Statistics, Carnegie Mellon University, 2000.

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

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

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

[8] 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. *Jurimetrics*, 39:347–375, Summer 1999.

[9] Barbara E. Bryant. Census-taking for a litigious, data-driven society. *Chance*, 6(3):44–49, 1993.

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

[11] 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.

[12] Beth Osborne Daponte, Stephen E. Fienberg, Joseph B. Kadane, and Duane L. Steffey. Sampling and Census 2000–methodological issues: Report to the Donner Foundation. Technical report, Department of Statistics Carnegie Mellon University, 1999.

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

[14] 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.

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

[16] Eugene P. Ericksen, Leobardo F. Estrada, John W. Tukey, and Kirk M. Wolter. Report on the 1990 decennial census and the post-enumeration survey, appendix a: The census process., 1991. Report submitted by members of the Special Advisory Panel to the Secretary, U.S. Department of Commerce.

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

[18] Joseph B. Kadane, Michael M. Meyer, and John W. Tukey. Yule's association pardox and ignored stratum heterogeneity in capture-recapture studies. *Journal of the American Statistical Association*, 94:855–859, 1999.

[19] James Madison. *The Debates in the Federal Convention of 1787 Reported by James Madison*. W. W. Norton & Co., April 1987.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] Bruce D. Spencer. Implications of equity and accuracy for undercount adjustment a decision-theoretic approach. In *Conference on Census Undercount, Proceedings of the 1980 Conference*, pages 204–216, Washington, D.C., 1980. U.S. Bureau of the Census, U.S. Department of Commerce.

[25] Bruce D. Spencer. The role of judgement in postcensal estimates. In Committee on National tatistics Panel on Small-Area Estimates, editor, *Estimating Population and Income of Small Areas*, pages 194–199. National Academy Press, Washington, D.C., 1980.

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

[27] Duane L. Steffey and Norman M. Bradburn, editors. *Counting People in the Information Age*. National Academy Press, Washington, DC, 1994.

[28] 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.

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