

On the Likelihood of Improving the Accuracy of the Census Through Statistical Adjustment

David A. Freedman and Kenneth W. Wachter

Abstract

In this article, we sketch procedures for taking the census, making adjustments, and evaluating the results. Despite what you read in the newspapers, the census is remarkably accurate. Statistical adjustment is unlikely to improve on the census, because adjustment can easily put in more error than it takes out. Indeed, error rates in the adjustment turn out to be comparable to—if not larger than—errors in the census. The data suggest a strong geographical pattern to these errors even after controlling for demography, which contradicts a basic premise of adjustment. Complex demographic controls built into the adjustment mechanism turn out to be counter-productive.

Proponents of adjustment have cited “loss function analysis” to compare the accuracy of the census and adjustment, generally to the advantage of the latter. However, these analyses make assumptions that are highly stylized and quite favorable to adjustment. With more realistic assumptions, loss function analysis is neutral or favors the census. At the heart of the adjustment mechanism, there is a large sample survey—the post enumeration survey. The size of the survey cannot be justified. The adjustment process now consumes too large a share of the Census Bureau’s scarce resources, which should be reallocated to other Bureau programs.

Keywords: Census; adjustment; heterogeneity; correlation bias; demographic analysis; dual-system estimation; non-sampling error; loss function analysis

1 Introduction

The census has been taken every ten years since 1790. Counts are used to apportion Congress and redistrict states. Furthermore, census data are the basis for allocating federal tax money to cities and other local governments. For such purposes, the geographical distribution of the population matters rather than counts for the nation as a whole. Data from 1990 and previous censuses suggested there would be a net undercount in 2000; the undercount would depend on age, race, ethnicity, gender, and—most importantly—geography. This differential undercount, with its implications for sharing power and money, attracted considerable attention in the media and the court-house.

There were proposals to adjust the census by statistical methods, but this is advisable only if the adjustment gives a truer picture of the population and its geographical

distribution. The census turns out to be remarkably good, despite the generally bad press reviews. Statistical adjustment is unlikely to improve the accuracy, because adjustment can easily put in more error than it takes out.

In this article, which is an expanded version of Freedman and Wachter [15], we sketch procedures for taking the census, making adjustments, and evaluating results. (A sketch is what you want: detailed descriptions cover thousands of pages.) We have new data on errors in the adjustment, and on geographical variation in error rates. We discuss alternative adjustments, and point out critical flaws in oft-cited methods for comparing the accuracy of the census and adjustment. We close with pointers to the literature, including citations to the main arguments for and against adjustment, and a summary of the policy recommendations that follow from our analysis.

2 The Census

The census is a sophisticated enterprise whose scale is remarkable. In round numbers, there are 10,000 permanent staff at the Bureau of the Census. Between October 1999 and September 2000, the staff opened 500 field offices, where they hired and trained 500,000 temporary employees. In spring 2000, a media campaign encouraged people to cooperate with the census, and community outreach efforts were targeted at hard-to-count groups.

The population of the United States is about 280 million persons in 120 million housing units, distributed across 7 million “blocks,” the smallest pieces of census geography. (In Boston or San Francisco, a block is usually a block; in rural Wyoming, a “block” may cover a lot of pastureland.) Statistics for larger areas like cities, counties, or states are obtained by adding up data for component blocks.

From the perspective of a census-taker, there are three types of areas to consider. In “city delivery areas” (high-density urban housing with good addresses), the Bureau develops a Master Address File. Questionnaires are mailed to each address in the file. About 70 percent of these questionnaires are filled out and returned by the respondents. Then “Non-Response Followup” procedures go into effect: for instance, census enumerators go out several times and attempt to contact non-responding households, by knocking on doors and working the telephone. City delivery areas include roughly 100 million housing units.

“Update/leave” areas, comprising less than 20 million households, are mainly suburban and have lower population densities; address lists are more difficult to construct. In such areas, the Bureau leaves the census questionnaire with the household while updating the Master Address File. Beyond that, procedures are similar to those in the city delivery areas.

In “update/enumerate” areas, the Bureau tries to enumerate respondents—by interviewing them—as it updates the Master Address File. These areas are mainly rural, and post-office addresses are poorly defined, so address lists are problematic. (A typical address might be something like Smith, Rural Route #1, south of Willacoochee, GA.)

Perhaps a million housing units fall into such areas. There are also special populations that need to be enumerated—institutional (prisons and the military), as well as non-institutional “group quarters.” (For instance, 12 nuns sharing a house in New Orleans are living in group quarters.) About 8 million persons fall into these two categories.

3 Demographic Analysis

Demographic analysis estimates the population using birth certificates, death certificates, and other administrative record systems. The estimates are made for national demographic groups defined by age, gender, and race (Black and non-Black). Estimates for sub-national geographic areas like states are currently not available. According to demographic analysis, the undercount in 1970 was about 3 percent nationally. In 1980, it was 1 to 2 percent, and the result for 1990 was similar. The undercount for Blacks was estimated at about 5 percentage points above non-Blacks, in all three censuses.

Demographic analysis starts from an accounting identity:

$$\text{Population} = \text{Births} - \text{Deaths} + \text{Immigration} - \text{Emigration}.$$

However, data on emigration are incomplete. And there is substantial illegal immigration, which cannot be measured directly. Thus, estimates need to be made for illegals, but these are (necessarily) somewhat speculative.

Evidence on differential undercounts depends on racial classifications, which may be problematic. Procedures vary widely from one data collection system to another. For the census, race of all household members is reported by the person who fills out the form. In Census 2000, respondents were allowed for the first time to classify themselves into multiple racial categories: this is a good idea from many perspectives, but creates a discontinuity with past data. On death certificates, race of decedent is often determined by the undertaker. Birth certificates show the race of the mother and (usually) the race of father; procedures for ascertaining race differ from hospital to hospital. A computer algorithm is used to determine race of infant from race of parents.

Prior to 1935, many states did not have birth certificate data at all; and the further back in time, the less complete is the system. This makes it harder to estimate the population aged 65 and over. In 2000, demographic analysis estimates the number of such persons starting from Medicare records. Despite its flaws, demographic analysis has generally been considered to be the best yardstick for measuring census undercounts. Recently, however, proponents of adjustment have favored another procedure, the DSE (“Dual System Estimator”).

4 DSE—Dual System Estimator

The DSE is based on a special sample survey done after the census—a PES (“Post Enumeration Survey”). The PES of 2000 came to be called ACE (“Accuracy and Coverage Evaluation Survey”): acronyms seem to be unstable linguistic compounds. The

ACE sample covers 25,000 blocks, containing 300,000 housing units and 700,000 people. An independent listing is made of the housing units in the sample blocks, and persons in these units are interviewed after the census is complete. This process yields the “P-sample.”

The “E-sample” comprises the census records in the same blocks, and the two samples are then matched up against each other. In most cases, a match validates both the census record and the PES record. A P-sample record that does not match to the census may be a “gross omission,” that is, a person who should have been counted in the census but was missed. Conversely, a census record that does not match to the P-sample may be an “erroneous enumeration,” in other words, a person who got into the census by mistake. For instance, a person can be counted twice in the census—because he sent in two forms. Another person can be counted correctly but assigned to the wrong unit of geography: she is a gross omission in one place and an erroneous enumeration in the other.

Of course, an unmatched P-sample record may just reflect an error in ACE; likewise, an unmatched census record could just mean that the corresponding person was found by the census and missed by ACE. Fieldwork is done to “resolve” the status of some unmatched cases—deciding whether the error should be charged against the census or ACE. Other cases are resolved using computer algorithms. However, even after fieldwork is complete and the computer shuts down, some cases remain unresolved. Such cases are handled by statistical models that fill in the missing data. The number of unresolved cases is relatively small, but it is large enough to have an appreciable influence on the final results (Section 9).

Movers—people who change address between census day and ACE interview—represent another complication. Unless persons can be correctly identified as movers or non-movers, they cannot be correctly matched. Identification depends on getting accurate information from respondents as to where they were living at the time of the census. Again, the number of movers is relatively small, but they are a large factor in the adjustment equation (Section 9). More generally, matching records between the ACE and the census becomes problematic if respondents give inaccurate information to the ACE, or the census, or both. Thus, even cases that are resolved through ACE fieldwork and computer operations may be resolved incorrectly. We refer to such errors as “processing error.”

The statistical power of the DSE comes from matching, not from counting better. In fact, the E-sample counts came out a bit higher than the P-sample counts, in 1990 and in 2000: the census found more people than the post enumeration survey.¹ As the discussion of processing error shows, however, matching (like so many other things) is easier said than done.

Some persons are missed both by the census and by ACE. Their number is estimated using a statistical model, assuming that ACE is as likely to find people missed by the census as people counted in the census—“the independence assumption.” Following this assumption, a gross omission rate estimated from the people found by ACE is

extrapolated to the sort of people who are unlikely to be found, although the gross omission rate for the latter group may well be different. Failures in the independence assumption lead to “correlation bias.” Data on processing error and correlation bias will be presented later.

5 Small-Area Estimation

The Bureau divides the population into “post strata” defined by demographic and geographic characteristics. For Census 2000, there were 448 post strata. One post stratum, for example, consisted of Asian male renters age 30–49, living anywhere in the United States. Another post stratum consisted of Blacks age 0–17 (male or female) living in owner-occupied housing in big or medium-size cities with high mail return rates, across the whole country. Persons in the P-sample are assigned to post strata on the basis of information collected during the ACE interview. (For the E-sample, assignment is based on the census return.)

Each sample person gets a “weight.” If the Bureau sampled 1 person in 500, each sample person would stand for 500 in the population and be given a weight of 500. The actual sampling plan for ACE is more complex, so different people get different weights, ranging from 10 to 6000.² To estimate the total number of gross omissions in a post stratum, the Bureau simply adds the weights of all ACE respondents who were identified as (i) gross omissions and (ii) being in the relevant post stratum.

To a first approximation, the estimated undercount in a post stratum is the difference between the estimated numbers of gross omissions and erroneous enumerations.³ The Bureau computes an “adjustment factor”; when multiplied by this factor, the census count for a post stratum equals the estimated true count from the DSE. About two-thirds of the adjustment factors exceed 1: these post strata are estimated to have undercounts. The remaining post strata are estimated to have been overcounted by the census; their adjustment factors are less than 1.⁴

How does the Bureau adjust small areas like blocks, cities, or states? Take any particular area. Each post stratum has some number of persons counted by the census in that area. (The number may be zero.) This census number is multiplied by the adjustment factor for the post stratum. The process is repeated for all post strata, and the adjusted count is obtained by adding the products; complications due to rounding are ignored for now. The adjustment process makes the “homogeneity assumption,” that undercount rates are constant within each post stratum across all geographical units. This is not plausible, and was strongly contradicted by census data on variables related to the undercount. Failures in the homogeneity assumption are termed “heterogeneity.” Ordinarily, samples are used to extrapolate upwards, from the part to the whole. In census adjustment, samples are used to extrapolate sideways, from 25,000 sample blocks to each and every one of the 7 million blocks in the United States. That is where the homogeneity assumption comes into play.

The political debate over adjustment is often framed in terms of sampling: “sam-

pling is scientific.” However, from a technical perspective, sampling is not the issue. The crucial questions are about the size of processing errors, and the validity of statistical models for missing data, correlation bias, and homogeneity—in a context where the margin of allowable error is relatively small.

6 State Shares

All states would gain population from adjustment. Some, however, gain more than others. In terms of population share, gains and losses must balance—a subtle point often overlooked in the political debate. In 2000, even more than 1990, share changes were tiny. According to Census 2000, for example, Texas had 7.4094 percent of the population. Adjustment would have given it 7.4524 percent, an increase of $7.4524 - 7.4094 = .0430$ percent, or 430 parts per million. The next biggest winner was California, at 409 parts per million; third was Georgia, at 88 parts per million.

Ohio would have been the biggest loser, at 241 parts per million; then Michigan, at 162 parts per million. Minnesota came third in this sorry competition, at 152 parts per million. The median change (up or down) is about 28 parts per million. These changes are tiny, and most are easily explained as the result of sampling error in ACE. “Sampling error” means random error introduced by the luck of the draw in choosing blocks for the ACE sample; you get a few too many blocks of one kind or not quite enough of another: the contrast is with “systematic” or “non-sampling” error like processing error.

The map (Figure 1) shows share changes that exceed 50 parts per million. Share increases are marked “+”; share decreases, “-”. The size of the mark corresponds to the size of the change. As the map indicates, adjustment would have moved population share from the Northeast and Midwest to the South and West. This is paradoxical, given the heavy concentrations of minorities in the big cities of the Northeast and Midwest—and political rhetoric contending that the census shortchanges such areas (“statistical grand larceny,” according to New York’s ex-Mayor Dinkins). One explanation for the paradox is correlation bias. The older urban centers of the Northeast and Midwest may be harder to reach, both for census and for ACE.

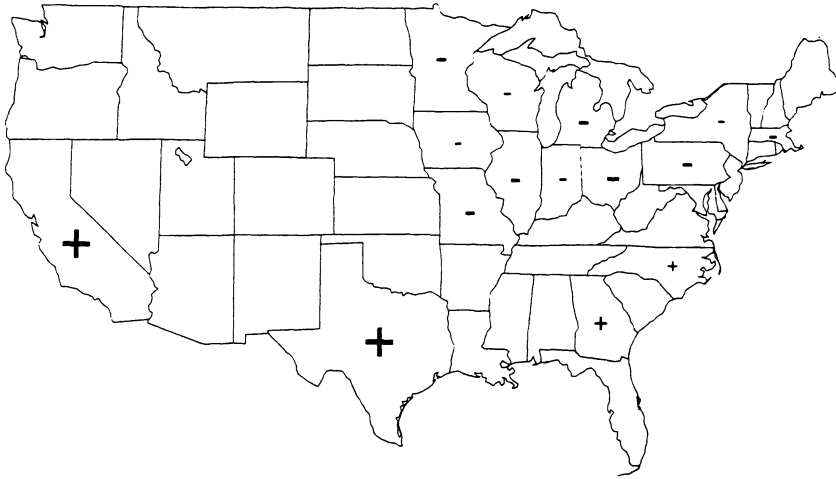


Figure 1: ACE adjustment: State share changes exceeding 50 parts per million⁵

7 The 1990 Adjustment Decision

Table 1: Errors in the adjustment of 1990

The adjustment	+5.3
Processing error	-3.6
	<hr/>
Corrected adjustment	+1.7
Correlation bias	+3.0
	<hr/>
Demographic analysis	+4.7

A brief look at the 1990 adjustment decision provides some context for discussions of Census 2000. In July 1991, the Secretary of Commerce declined to adjust Census 1990. At the time, the undercount was estimated as 5.3 million persons (Table 1). Of this, 1.7 million persons were thought by the Bureau to reflect processing errors in the post enumeration survey, rather than census errors. Later research has shown the 1.7 million to be a serious underestimate. Current estimates range from 3.0 million to 4.2 million, with a central value of 3.6 million. (These figures are all nation-wide, and net.) Thus, the bulk of the 1990 adjustment resulted from errors not in the census but in the PES. Processing errors generally inflate estimated undercounts, and subtracting them leaves a corrected adjustment of 1.7 million. (There is an irritating numerical coincidence here, as 1.7 million enters the discussion with two different roles.) Correlation

bias, estimated at 3.0 million, works in the opposite direction, and brings the undercount estimate up to the demographic analysis figure of 4.7 million.⁶ The message is simple: on the scale of interest, most of the estimated undercount is noise.

8 Evaluating Census 2000

We see widespread—although by no means universal—agreement on two chief points. First, Census 2000 succeeded in reducing differential undercounts from their 1990 levels. Second, there are serious questions about the accuracy of proposed statistical adjustments. Mistakes in statistical adjustments are nothing new. Studies of the 1980 and 1990 data have quantified, at least to some degree, the three main kinds of error: processing error, correlation bias, and heterogeneity. In the face of these errors, it is hard for adjustment to improve on the accuracy of census numbers for states, counties, legislative districts, and smaller areas. Statistical adjustment can easily put in more error than it takes out, because the census is already very accurate.

In 1990, there were many studies on the quality of the adjustment. For 2000, evaluation data are only beginning to be available. However, the Bureau's preliminary estimates, based largely on the experience of 1990, suggested that processing error in ACE contributes about 2 million to the estimated undercount of 3.3 million.⁷ (Errors in ACE will be discussed in more detail, below.) Errors in the ACE statistical operations may from some perspectives have been under better control than they were in 1990. But error rates may have been worse in other respects. There is continuing research, both inside the Bureau and outside, on the nature of the difficulties. The Bureau investigated a form of error called "balancing error"—essentially, a mismatch between the levels of effort in detecting gross omissions or erroneous enumerations. We think that troubles also occurred with a new treatment of movers (discussed in the next section) and duplicates. Some 25 million duplicate persons were detected in various stages of the census process, and removed.⁸ But how many slipped through?

Besides processing error, correlation bias is an endemic problem that make it extremely difficult for adjustment to improve on the census. Correlation bias is the tendency for people missed in the census to be missed by ACE as well. Correlation bias in 2000 probably amounted, as it did in 1990, to millions of persons. These people cannot be evenly distributed across the country. If their distribution is uneven, the DSE creates a distorted picture of census undercounts. Heterogeneity is also endemic: undercount rates differ from place to place within population groups treated as homogeneous by adjustment. Heterogeneity puts limits on the accuracy of adjustments for areas like states, counties, or legislative districts. Studies of the 1990 data, along with more recent work discussed below, show that heterogeneity remains a serious concern.⁹

9 The Adjustment Decision: March 2001

In March 2001, the Secretary of Commerce—on the advice of the Census Bureau—decided to certify the census counts rather than the adjusted counts for use in redistricting (drawing congressional districts within state).¹⁰ The principal reason was that, according to demographic analysis, the census had overcounted the population by perhaps 2 million people. Proposed adjustments would have added another 3 million people, making the overcounts even worse. Thus, demographic analysis and ACE pointed in opposite directions. The three population totals are shown in Table 2.¹¹

Table 2: The population of the United States

Demographic analysis	279.6 million
Census 2000	281.4 million
ACE	284.7 million

If demographic analysis is right, there is a census overcount of .7 percent. If ACE is right, there is a census undercount of 1.2 percent. Demographic analysis is a particularly valuable benchmark, because it is independent (at least in principle) of both the census and the post enumeration survey that underlies proposed adjustments. While demographic analysis is hardly perfect, it was a stretch to blame demographic analysis for the whole of the discrepancy with ACE. Instead, the discrepancy pointed to undiscovered error in ACE. Evaluations of the ACE data are ongoing, so conclusions must be tentative. However, there was some information on missing data and on the influence of movers available in March 2001, summarized in Table 3.¹²

Table 3: Missing data in ACE, and impact of movers

Non-interviews	
P-sample	3 million
E-sample	6 million
Imputed match status	
P-sample	3 million
E-sample	7 million
Inmovers and outmovers	
Imputed residence status	6 million
Outmovers	9 million
Inmovers	13 million
Mover gross omissions	3 million

These figures are weighted to national totals, and should be compared to (i) a total census population around 280 million, and (ii) errors in the census that may amount to a

few million persons. For some 3 million P-sample persons, a usable interview could not be completed; for 6 million, a household roster as of census day could not be obtained (lines 1 and 2 in the table). Another 3 million persons in the P-sample and 7 million in the E-sample had unresolved match status after fieldwork: were they gross omissions, erroneous enumerations, or what? For 6 million, residence status was indeterminate—where *were* they living on census day? (National totals are obtained by adding up the weights for the corresponding sample people; non-interviews are weighted out of the sample and ignored in the DSE, but we use average weights.) If the idea is to correct an undercount of a few million in the census, these are serious gaps. Much of the statistical adjustment therefore depends on models used to fill in missing data. Efforts to validate such models remain unconvincing, despite some over-enthusiastic claims in the administrative and technical literature.¹³

The 2000 adjustment tried to identify both in-movers and out-movers, a departure from past practice. Gross omission rates were computed for the out-movers and applied to the in-movers, although it is not clear why rates are equal within local areas. For out-movers, information must have been obtained largely from neighbors. Such “proxy responses” are usually thought to be of poor quality, inevitably creating false non-matches and inflating the estimated undercount. As the table shows, movers contribute about 3 million gross omissions (a significant number on the scale of interest) and ACE failed to detect a significant number of out-movers. That is why the number of out-movers is much less than the number of in-movers. Again, the amount of missing data is small relative to the total population, but large relative to errors that need fixing. The conflict between these two sorts of comparisons is the central difficulty of census adjustment. ACE may have been a great success by the ordinary standards of survey research, but not good enough for adjusting the census.

10 Gross or Net?

Errors can be reported either “gross” or “net,” and there are many possible ways to refine the distinction. Given the uncertainties, we find that error rates in the adjustment are comparable to—if not larger than—error rates in the census, whether gross or net. For context, proponents of adjustment have lately favored measuring errors in the census on a gross basis rather than net, citing concerns about geographical imbalances. Some places may have an excess number of census omissions while other places will have an excess number of erroneous inclusions. Such imbalances could indeed be masked by net error rates. However, adjustment is hardly a panacea for geographical imbalance. The adjustment mechanism allows cancellation of errors within post strata—the homogeneity assumption at work. Much of the gross error is netted out, post stratum by post stratum; the rest is spread uniformly across geography within post strata. Adjustment fixes geographical imbalances in the census only if you buy the homogeneity assumption: location is accident, demography is destiny.

Proponents of adjustment have also objected to a comparison between undercount

estimates and estimated processing error (as in Section 7), on the grounds that net errors can after all be negative. We are unsympathetic to this complaint. For most areas with substantial populations—all states and 433/435 congressional districts—the adjustment is positive. Furthermore, estimated processing error is positive for all states and all congressional districts.¹⁴ For such areas, adjustment adds to the population totals, and these increments mainly result from errors in the adjustment rather than errors in the census. All that said, a comparison of gross error rates will be instructive: see Table 4, where rows are numbered to match the paragraphs below.

Table 4: Errors in the census and ACE. Millions of persons. March figures.

	Positive Error	Negative Error	Gross Error	Para- graph
Census	1.00	4.26	5.26	(i)
ACE	1.75	.90	2.65	(ii)
Census	.10	2.51	2.61	(iii)
Census	3.1	6.4	9.5	(iv)
ACE			12.8	(v)

Sign convention. A “positive” error makes a population estimate too high, while a negative error makes the estimate too low. Thus, correlation bias counts as a negative error for ACE. Generally, processing error in ACE is positive.

(i) ACE would have added 4.26 million persons nationwide in certain post strata, and subtracted 1.00 million in other post strata. The net change is $4.26 - 1.00 = 3.26$ million but the gross is $4.26 + 1.00 = 5.26$ million. In effect, these are net and gross errors in the census, as estimated by ACE.¹⁵

(ii) For comparison, the Bureau’s estimated biases in ACE (as of March 2001) add 1.75 million persons to the adjustment in certain post strata and subtract .90 million in other post strata, for a net error of $1.75 - .90 = .85$ million and a gross error of $1.75 + .90 = 2.65$ million.¹⁶ The latter is about half the proposed gross adjustment. Thus, proponents of adjustment must think there are 5.26 million gross errors in the census that are detected by ACE and can be fixed by adjustment. But half of these represent errors in the adjustment mechanism itself, rather than errors in the census—even on the March figures for biases in ACE, which were largely based on extrapolation from 1990.

(iii) This comparison, however, is far too generous to ACE, because biases in ACE are counted against the census. Instead, we can estimate errors in the census from a bias-corrected ACE, sticking with the March figures for bias. On this basis, gross error rates in the census are virtually the same as those in ACE. In line (iii), for instance, the negative error for the census can be computed from the data in lines (i) and (ii), as

$4.26 - 1.75 = 2.51$ million, and similarly for the positive error.¹⁷ The gross error in the census is 2.61 million; in ACE, 2.65 million.

(iv) Some number of persons were left out of Census 2000 and some were counted in error. Even if ACE had been done with surgical precision, there is no easy way to estimate the size of these two errors separately. Many people were counted a few blocks away from where they should have been counted: they are both gross omissions and erroneous enumerations. Many other people were classified as erroneous enumerations because they were counted with insufficient information for matching; they should also come back as gross omissions in the ACE fieldwork. With some rough-and-ready allowances for this sort of double-counting, the Bureau estimated that 6–8 million people were left out of the census while 3–4 million were wrongly included, for a gross error in the census of 9–12 million; the Bureau's preferred values are 6.4 and 3.1, for a gross of 9.5 million.¹⁸ Much of this nets out within post strata: see line (i).

(v) For comparison, gross errors in ACE amount to 11.7 million after weighting to national totals, with an additional 1.1 million for correlation bias: here, cancellation is not allowed within post strata.¹⁹ Doubtless, the 11.7 million double-counts some errors; and in any event, much of the error will net out within post strata. Still, on this basis, gross error rates in ACE are substantially larger than those in the census.

It is puzzling to see proponents of adjustment reciting gross error rates for the census, like those in line (iv) of the table, as if such data justified their position.²⁰ Errors that cancel within post strata cannot matter to the adjusters, or at least to those who care about logical consistency, because such errors—according to their theories—affect the accuracy neither of the census nor of the adjustment. Moreover, gross error rates in ACE are comparable to, if not larger than, gross error rates in the census.

11 Error Rates in ACE: October 2001

In October 2001, the Bureau decided not to adjust the census as a base for post-censal population estimates. This sounds even drier than redistricting, but \$200 billion a year of tax money are allocated using such estimates. The decision was made after further analysis of the data, carried out between March and October. The Bureau added 2.2 million to the demographic analysis; and processing error in ACE went from 2 million to 5–6 million. Moreover, the Bureau confirmed that gross errors in ACE were well above 10 million, with another 15 million cases whose status remains to be resolved.²¹ Any way you slice it, a large part of the adjustment comes about because of errors in the adjustment process rather than the census.

Before the October decision, we tried to reconcile the figures in Table 2 for the population of the United States—279.6 million from demographic analysis, 281.4 million from the census, and 284.7 million from ACE. There are good (albeit post hoc) arguments for increasing the demographic analysis figure, perhaps by 2 million; the census seemed about right to us, or even a little high; and in our view, the net processing error in ACE was probably 5–6 million, partially offset by correlation bias amounting to

2–3 million. In short, the Bureau’s preliminary estimates for processing error in ACE needed to be doubled or tripled, and so did Bureau estimates for correlation bias. In the main, these forecasts are confirmed by the October decision document: U. S. Census Bureau [25]. Bureau estimates for correlation bias, however, are still based on a fiction—that is there is no correlation bias for women.²²

12 Heterogeneity in 2000

Table 5: Measuring heterogeneity. In the first column, post stratification is either (i) by the Bureau’s 448; or (ii) by the 64 post-stratum groups, that is, collapsing age and sex; or (iii) by the 16 evaluation post strata. “II” means whole-person substitutions, and “LA” is late census adds. In the last two columns, “P-S” stands for post strata; these are of three different kinds, according to rows of the table.

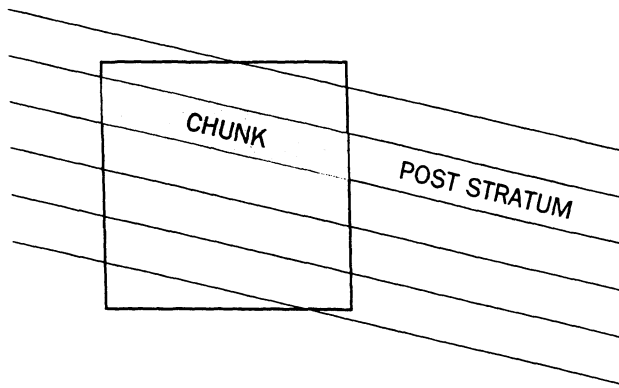
Proxy & Post Stratification	Level	Standard Deviation		
		Across states	Across P-S	Within P-S across states
II 448	.0208	.0069	.0134	.0201
II 64	.0208	.0069	.0131	.0128
II 16	.0208	.0069	.0133	.0089
LA 448	.0085	.0036	.0070	.0118
LA 64	.0085	.0036	.0069	.0074
LA 16	.0085	.0036	.0056	.0046

Note. The level does not depend on the post stratification, and neither does the SD across states. These two statistics do depend on the proxy.

In this section, we show that substantial heterogeneity remains in the data, despite the Bureau’s elaborate post stratification; in fact, the post stratification seems on the whole to be counter-productive. We measure heterogeneity as in Freedman and Wachter [13], with “whole-person substitutions” and “late census adds” as “proxies” (surrogates) for undercount.²³ For example, 2.08% of the census count came from whole-person substitutions (“II” in the first line of the table, for obscure historical reasons). We compute these substitution rates, not only for the whole country, but for each state and DC: the standard deviation (SD) of the 51 rates is .69 of 1%. We also compute the rate for each post stratum: across the 448 post strata, the SD of the substitution rates is 1.34%: the post strata do show considerably more variation than the states.

On the other hand, we can think of each state as being divided into “chunks” by the post strata, as in the sketch below. (Alabama, for instance, is divided into 251

chunks by the post strata; post stratum #1 is divided into 7 chunks by states.) We compute the substitution rate for each chunk with a non-zero census count, then take the SD across chunks within post stratum, and finally the root-mean-square over post strata. We get 2.01%. If rates were constant across states within post strata, as the homogeneity assumption requires, this SD should be 0. Instead, it is larger than the SD across post strata, and almost as large as the overall imputation level.



We made similar calculations for two coarser post stratifications. (i) The Bureau considers its 448 post strata as coming from 64 PSGs, a PSG being a “post-stratum group.” (Each PSG divides into 7 age-sex groups, giving back $64 \times 7 = 448$ post strata.) We use the 64 PSGs as post strata in the second line of Table 5. (ii) The Bureau groups PSGs into 16 EPS, or “evaluation post strata.” We use these as post strata in the third line of Table 5. Remarkably, there is less rather than more variability within post-stratum group than within post stratum—and even less within evaluation post stratum; the air of paradox may be dispelled by Freedman and Wachter [13, p. 482].²⁴ Results for late census adds (LA) are similar, in lines 4–6 of the table. If the proxies are good, refining the post stratification is counter-productive: with more post strata, there is more heterogeneity rather than less.

13 Alternative Post Stratifications

The Bureau computed “direct DSEs” for the 16 evaluation post strata, by pooling the data in each: we constructed an adjustment factor, as the direct DSE divided by the census count.²⁵ We adjusted the United States using these 16 factors rather than the Bureau’s 448. For states and congressional districts, there is hardly any difference: the scatter diagram in Figure 2 shows results for congressional districts. There are 435 dots, one for each congressional district. The horizontal axis shows the change in population count that would have resulted from adjustment with 448 post strata; the vertical, from adjustment with 16 post strata. There is little to choose between the two. (For some geographical areas with populations below 100,000, however, the two adjustments are

likely to have different consequences.)

TWO ADJUSTMENTS COMPARED. 435 CONGRESSIONAL DISTRICTS
DIFFERENCE BETWEEN ADJUSTED COUNT AND CENSUS COUNT

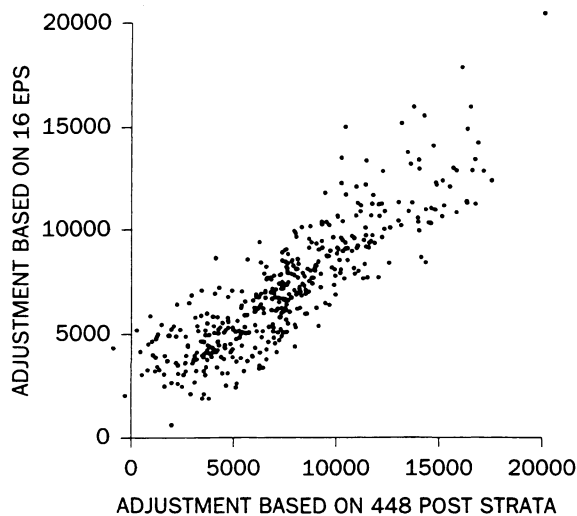


Figure 2: Changes to congressional district populations. The production adjustment, with 448 post strata, is plotted on the horizontal. An alternative, based only on the 16 evaluation post strata (EPS), is plotted on the vertical.

For example, take CD 1 in Alabama, with a 2000 census population of 646,181. Adjustment with 448 post strata would have increased this figure by 7630; with 16 post strata, the increase would have been 7486. The corresponding point is plotted at (7630, 7486). The correlation between the 435 pairs of changes is .87, as shown in the third line of Table 6.²⁶ For two out of the 435 districts, adjustment by 448 post strata would have reduced the population count: their points are plotted just outside the axes, at the lower left.

Within a state, districts are—by court order—almost exactly equal in size when redistricting is done shortly after census counts are released. Over the decade, of course, people move from one district to another. Variation in population sizes at the end of the decade is therefore of considerable policy interest. In California, for one example, 52 districts were drawn to have equal populations according to Census 1990. According to Census 2000, the range in their populations is 583,000 to 773,000.

Table 6: Comparing the production adjustment based on 448 post strata to one based on 16 evaluation post strata. Correlation coefficients for changes due to adjustment.

Changes in state population counts	.99
Changes in state population shares	.90
Changes in congressional district counts	.87
Changes in congressional district shares	.85

Table 6 and Figure 2 show that the Bureau's elaborate post stratification does not remove much heterogeneity. Whatever there was with 448 remains with 16, and that is a lot (Table 5). Experience from 1990 and 2000 teaches a sad lesson. Heterogeneity is not to be removed by the sort of post stratification that can be constructed by the Bureau. The impact of heterogeneity on errors in adjustment is discussed by Freedman and Wachter [13, pp. 479–81]: heterogeneity is likely to be much more of a problem than sampling error.

14 Loss Function Analysis

Proponents of adjustment often rely on a statistical technique called “loss function analysis.” In effect, this technique attempts to make summary estimates of the error levels in the census and the adjustment, generally to the advantage of the latter. However, the apparent gains in accuracy—like the gains from adjustment—tend to be concentrated in a few geographical areas, and heavily influenced by the vagaries of chance. At a deeper level, loss function analysis turns out to depend more on wishful assumptions than on data.

For example, adjustment makes the homogeneity assumption: census errors occur at a uniform rate within post strata across wide stretches of geography. Loss function analysis assumes that and more: error rates in the census are uniform, and so are error rates in ACE. A second example: loss function analysis depends on models for correlation bias, and the Bureau's model assumes there is no correlation bias for women. The idea that only men are hard to reach—for the census and the post enumeration survey—is unlikely on its face. It is also at loggerheads with the data from 1990. See Wachter and Freedman [29]. For such reasons, we cannot believe that loss function analysis will clarify remaining issues about use of adjusted census data.

The discussion now becomes progressively—but unavoidably—more technical. (Readers can skip to Section 15 or 16, without losing the thread of the argument.) Loss function analysis tries to compare accuracy of census and adjusted figures for defined geographical areas. By way of example, take counts for states. The main ingredients for the comparison are the following.

- (a) Census counts.
- (b) Production adjustment counts.

(c) Variances in production adjustment counts.

(d) Biases in production adjustment counts.

However, the quantities in (d) must themselves be estimated, and estimates have variances. Thus, we amend (d) and add (e).

(d) Estimated biases in production adjustment counts.

(e) Variances in estimated biases.

In 1990, estimates for the variances in (c) were questionable. That issue may not arise for 2000, but estimates for the variances in (e) remain problematic. Here is why. As noted above, much of the evaluation data used by the Bureau in March 2001 comes from the 1990 Evaluation Followup, a sample survey done several months after the post enumeration survey was completed. This survey was based on 919 block clusters; the 1990 PES, on 5290; the 2000 ACE, on 11,303. On this basis, variances should be $11,303/919 = 12$ times bigger than the ACE variances. Instead, they are about 4 times smaller. The variances for estimated biases are, by such a reckoning, too small by a factor of $4 \times 12 = 48$. Other calculations give much larger factors, but 48 is surely enough to make the point.²⁷

Where did the missing variance go? Processing error can be estimated from Evaluation Followup in fine-grain geographical detail. However, the sample is small, so variances for direct estimates would be huge. Instead, errors are aggregated to broad population groups (16 evaluation post strata in 2000) and then shared back down to constituent post strata, using proportionality assumptions. Finally, errors are spread across state or substate areas assuming constant error rates within post strata across geography, for correlation bias as well as processing error. Thus, variance in estimated errors is converted to bias by the sharing and spreading—but that particular bias is ignored in loss function analysis.

The statistical theory of loss function analysis. If we use survey data to estimate a parameter, loss can be defined as squared error. Risk is expected squared error, that is, averaged over hypothetical replications of the survey. Loss function analysis tries to make unbiased estimates of risk, as the variance of the estimator plus the square of its bias. Bias has to be estimated, and the variance of the bias estimator has to be accounted for. Unbiased estimators of bias, and unbiased estimators of their variances, are needed to make the calculation work.

For the intended application, consider two competing “estimators” of the population of California at census day in the year 2000: the census itself, and the adjustment based on ACE. The Bureau estimated a risk from the census: variance is nil, and bias is estimated primarily from ACE with some refinement from evaluation studies in 1990. We replicated the Bureau’s calculation, and found the estimated risk to be 167 billion; units are “squared people.” Likewise, there is an estimated risk for adjustment, which is 12.2 billion. Adjustment seems to make the smaller error. This process can be repeated for each state and DC. Summing the results gives a total estimated risk of 362 billion for the census, compared to 46 billion for adjustment. See line 1 of Table 7: a billion is 10^9 , and $362 - 46 = 316$.

Table 7: Loss function analysis for counts. States. Weights inverse to census counts.

Wtd	Cov	Diff	SE	Scale
No	1	316	139	10^9
No	25	316	432	10^9
Yes	1	193	95	10^2
Yes	25	193	303	10^2

The Bureau preferred to divide the estimated risk for each state by its population, at least in its March report. We measure population by the census count: for California, this is 33.1 million.²⁸ The census risk for California is 167 billion/33.1 million \doteq 5050 while the adjustment risk is 12.2 billion/33.1 million \doteq 369. (Here, \doteq means “approximately equal.”) In total over all 51 states (and DC), we get 24,558 for the census risk and 5208 for the risk from adjustment.²⁹ The difference is $24,558 - 5208 = 19,350$, which is rounded to 193×10^2 in line 3 of Table 7; “wtd” indicates division by population counts.

In Table 7, biases are estimated from the Bureau’s “preferred” model.³⁰ If “Cov” is 25, the Bureau’s covariance matrix for estimated biases is multiplied by 25, which brings the variances closer to what might be anticipated on the basis of sample size, as discussed above. Let “Cen” be estimated census loss and “Adj” be estimated adjustment loss. Then “Diff” in the table is $\text{Cen} - \text{Adj}$, which is the estimated gain in accuracy from adjusting. Diff is estimated from sample data, ACE and Evaluation Followup, and is therefore subject to sampling error. The “SE” column gives the standard error for Diff, and gauges the likely magnitude of sampling error in this estimate.

The last column indicates the units. Thus, in the first line of Table 7, the estimated gain in accuracy from adjustment is

$$(316 \pm 139) \times 10^9.$$

Diff is over twice its SE, and such a large value for Diff is hard to explain as the result of sampling error alone. (Diff is “statistically significant.”) However, the calculation rides on Bureau estimates for the sampling variability in the biases—which are too low. Correcting these, as in the second line of Table 7, makes Diff noticeably smaller than its SE, and readily explained as the result of chance. Correcting the covariance matrix for the biases does not change Diff itself, but has a pronounced effect on its estimated SE.

Table 8: Loss function analysis for shares. States. Weights inverse to census shares.

Wtd	Cov	Diff	SE	Scale
No	1	306	297	10^{-9}
No	25	306	778	10^{-9}
Yes	1	57	63	10^{-7}
Yes	25	57	153	10^{-7}

Table 8 turns to state shares; weights are inversely proportional to census population shares. Diff is at the chance level. For congressional districts, the Bureau's loss function uses shares within state, but weights states by the square of the census count. This seems both cumbersome and unnatural—at least to us. We replicated the Bureau's analysis, but also examined numerical accuracy with the squared error loss function and no weights (Table 9).

Table 9 treats congressional districts as 435 areas across the country, with populations ranging from 500,000 to 1,000,000. As before, the estimated gain in accuracy from adjustment is significant if we use Bureau variances for the bias estimates, but insignificant when we correct for under-estimation. The District of Columbia does not come into Table 9, and state boundaries play no special role.

Table 9: Loss function analysis for counts. Congressional districts, unweighted.

Wtd	Cov	Diff	SE	Scale
No	1	137	65	10^8
No	25	137	202	10^8

Tables 7 and 9 show that the statistical significance of loss function analysis for counts is strongly dependent on the modeling—among other things, on the homogeneity assumption for biases. Table 8 shows that, for shares, Diff is at the chance level. Still, Diff is positive in all the summary tables. Perhaps that means adjustment is better than the census? We think not. The Bureau's March estimates for processing error and correlation bias were on the low side. Table 10 doubles the Bureau's allowance for processing error, post stratum by post stratum; it doubles the Bureau's allowance for correlation bias in states likely to have had unusually high levels of correlation bias in the 1990 adjustment (Wachter and Freedman [29] Table 5). This brings processing error to 4 million and correlation bias to 1.5 million; it allows for some geographical variation in rates of correlation bias, a possibility which is excluded by the Bureau's model. The corrected loss function analysis favors the census.

Table 10: Loss function analysis for counts. States. Partial correction for underestimated processing error and correlation bias. Some differentials in correlation bias. Weights inverse to census counts.

Wtd	Cov	Diff	SE	Scale
No	1	-185	99	10^8
No	25	-185	421	10^8
Yes	1	-214	65	10^2
Yes	25	-214	295	10^2

Table 11 gives detail for “Diff” in lines 3 and 4 of Table 8. Five states (CA, IA, MN, MO, TX) account for over half the estimated loss from the census. For the third line of Table 7, four states (CA, FL, GA, TX) account for over half the estimated loss from the census. Unweighted results (line 1 in Tables 7 and 8) are dominated by two states—CA and TX. In short, estimated gains from adjustment are concentrated in a few states, and subject to large uncertainties. Unbiased estimates of risk can be negative; that happens in Tables 7–10, and is explicit in Table 11.

Table 11: Estimated losses in accuracy from the census and from adjustment. State shares. Weights inverse to census shares. Parts per 10 million. Detail for “Diff” in lines 3 and 4 of Table 8. Alabama—Minnesota.

	Cen	Adj
Alabama	-0.6	0.6
Alaska	1.0	0.8
Arizona	-1.3	1.7
Arkansas	-0.4	1.0
California	12.6	2.5
Colorado	-1.1	1.1
Connecticut	-0.2	0.4
Delaware	0.7	0.4
DC	7.7	2.5
Florida	2.8	3.2
Georgia	7.3	2.1
Hawaii	1.1	2.4
Idaho	-0.6	1.0
Illinois	3.2	0.8
Indiana	4.9	0.7
Iowa	10.1	1.5
Kansas	5.2	0.7
Kentucky	-0.2	1.1
Louisiana	1.5	1.0
Maine	-1.8	2.9
Maryland	8.0	7.1
Massachusetts	3.7	0.8
Michigan	3.1	1.1
Minnesota	18.2	0.8

Table 11, Continued: Estimated losses in accuracy from the census and from adjustment. State shares. Weights inverse to census shares. Parts per 10 million. Detail for “Diff” in lines 3 and 4 of Table 8. Mississippi—Wyoming.

	Cen	Adj
Mississippi	-0.9	0.7
Missouri	12.2	0.5
Montana	-1.4	3.2
Nebraska	4.1	0.5
Nevada	0.3	0.8
New Hampshire	-0.3	0.8
New Jersey	-1.5	1.3
New Mexico	-1.0	3.1
New York	-0.9	2.2
North Carolina	0.2	0.8
North Dakota	3.2	0.7
Ohio	9.1	1.1
Oklahoma	-0.6	1.4
Oregon	-0.7	0.7
Pennsylvania	2.9	2.9
Rhode Island	0.3	0.5
South Carolina	1.0	1.9
South Dakota	3.4	0.9
Tennessee	-0.1	0.6
Texas	15.4	2.5
Utah	-1.0	1.1
Vermont	-1.0	1.4
Virginia	2.2	1.7
Washington	-0.9	3.1
West Virginia	1.0	2.9
Wisconsin	5.0	0.6
Wyoming	-0.5	0.9
Total	134.3	76.9

Given the levels of ACE processing error reported in U. S. Census Bureau [25], loss function analysis is an academic exercise. However, this sort of analysis seems to have played a salient role in Bureau deliberations over the 1990 adjustment, and was even a factor in the decision for Census 2000: see U. S. Census Bureau [24].³¹ We think it is time to stop using loss function analysis. The assumptions are too fanciful.

15 Artificial Population Analysis

In essence, loss function analysis justifies the homogeneity assumption by making an even stronger assumption: not only are error rates in the census constant within post strata across geography, so are error rates in ACE. Proponents of adjustment may cite report B14: using proxy variables for overcounts and undercounts, that report creates artificial populations where truth is known. Bias in loss function analysis that

results from failures in the homogeneity assumption can then be measured—for the artificial populations. However, some proxies favor adjustment and some do not: Fay and Thompson [10, p. 82], Freedman and Wachter [13, pp. 484–5].

Detailed results in B14 are rather mixed. Moreover, the B14 artificial populations are, well, artificial. Overcounts and undercounts are measures of difficulty in data collection. Intuition and data analysis suggest the following two criteria for proxies.

- (i) The proxies for overcount and undercount should be positively correlated, but not perfectly correlated.
- (ii) Proxies should be correlated with other indicators of poor data quality.

Bureau report B14 used four artificial populations, and Table 1 in that report lists the proxies. Populations #2 and #4 violate condition (i), using the same proxy for overcounts as for undercounts:

non-substituted persons in #2, non-mailbacks in #4.

With artificial population #3, there must be an inverse rather than a direct relationship between the two proxies, also violating condition (i):

persons with 2 or more allocations,

persons with non-allocated age and birth-date.

With artificial population #1, the proxies violate condition (ii), because they relate to goodness rather than badness of data:

non-substituted persons,

persons with non-allocated age and birth-date.

In consequence, the impact of heterogeneity on loss function analysis remains to be determined.³²

16 Pointers to the Literature

Reviews and discussions of the 1980 and 1990 adjustments can be found in *Survey Methodology* 18 (1992) 1–74 and *Statistical Science* 9 (1994) 458–537. *Journal of the American Statistical Association* 88 (1993) 1044–1166 has a lot of useful descriptive material. Although tilted toward adjustment, the collection does include an insightful paper on heterogeneity—Hengartner and Speed [18]—and a comment on the imputation model by Wachter [27]. Other exchanges worth noting include *Jurimetrics* 34 (1993) 59–115 and *Society* 39 (2001) 3–53: these are easy to read, and informative. Pro-adjustment arguments are made by Anderson and Fienberg [1, 2], but see Stark [23] and Ylvisaker [30]. Prewitt [21] may be a better source, although he too must be taken with several grains of salt. Proponents of adjustment often cite Zaslavsky [31] to demonstrate the comparative advantages of adjustment; however, that paper makes all the mistakes discussed in Section 14 above, and others too. Cohen, White, and Rust [7] try to answer arguments on the 1990 adjustment, but miss many points.³³ Skerry [22] has an accessible summary of the arguments, leaning against adjustment. Darga [8, 9] is the sternest of critics. Freedman, Stark and Wachter [12]

have a probability model for census adjustment, which may help to clarify some of the issues.

17 Policy Implications

Census adjustment has become an expensive program for the Bureau, especially in terms of senior management time. The cost of ACE is driven in part by its complexity, and in part by the sample size. The large sample is needed because there are so many post strata, so the sample is spread very thin. However, given the results in Tables 5–6, the number of post strata would be very hard to justify. Moreover, the large sample size, while helpful on the sampling error side, must contribute to non-sampling error. Bigger samples are harder to manage, and non-sampling error is the critical issue (Sections 4–12).

The sample size for the post enumeration survey should therefore be scaled back dramatically. If 448 post strata are cut back to 16, as suggested by Sections 12–13, the size of the sample can be reduced by a factor of $448/16 = 28$ while maintaining the average number of households per post stratum. The sample can therefore be reduced from 300,000 households to $300,000/28 \doteq 10,000$, although that seems too optimistic. If the program is to be continued, the focus should be research and evaluation not adjustment, and a sample in the range 10,000–25,000 households should be adequate.³⁴

If the adjustment program is scaled back, the savings could well be used elsewhere:

- Demographic Analysis (DA)
- American Community Survey (ACS)
- Maintaining the Master Address File (MAF)
- Community outreach between census years
- Research into counting methods
- Non-response followup.

Putting a few million dollars into demographic analysis now would make a big difference in 2010. Despite its flaws, DA allows reasonably accurate estimates for the sizes of major population groups, and these estimates could readily be improved—if resources were made available. DA is a more promising tool for census evaluation than ACE.

Decennial census long form data (income, education, occupation, country of origin, and so forth) are collected on a sample of about 1/6 of the respondents. The ACS will collect such data with a rolling sample survey, by interviewing 3% of the nation's households each year. The data will be available continuously, rather than every 10 years. From many perspectives, answers to ACS questionnaires are likely to be of better quality. Furthermore, the burden on the census will be markedly reduced. (The coverage of the ACS, however, is not likely to be as good as the census.)

The accuracy of a mail-out-mail-back census depends on having a list of addresses that is nearly complete, with few duplicates. Building such a list every 10 years is a

huge undertaking. Building it once and then maintaining it might be more productive. With ACS in place, maintaining the MAF seems like a promising activity.

Given the decades of effort spent in developing post enumeration surveys for census adjustment, the decision not to adjust must have been a wrenching one for the Bureau. We are confident they made the right decision. Statistical adjustments were considered in 1980, twice in 1990, and twice again in 2000. These adjustments could not improve the accuracy of the census. The adjustment technology does not work well enough to use. It is time to move on.

Endnotes

1. The Census Bureau provided detailed summary data on the census and the adjustment, by memorandum of understanding with the National Academy of Science and congressional oversight bodies. We were given access to these data; many of our results—like coverage comparisons—are computed from these data. The sum of P-sample non-movers and in-movers is about 1.6 million persons less than the number of E-sample persons (upweighted to national totals).
2. B2 p. 30 and App. 5. Also see R32, Attachment 2, Table 2. “B2” and “R32” are Bureau reports. For bibliographic details, see U. S. Census Bureau [24].
3. The procedure for estimating gross omissions is called “capture-recapture” in the statistical literature: capture is in the census, recapture is in the post enumeration survey. Erroneous enumerations are one major complication, and there are others. For a discussion of the 1990 procedures, see Hogan [19]. As yet, there is nothing comparable for 2000; report Q37 may be the best source.
4. Computed from data described in note 1.
5. Computed from data described in note 1.
6. See Wachter and Freedman [29], with cites to the literature; Breiman [4] is a primary source for alternative error estimates.
7. See B13 and B14 for data sources, and note 14 for methods. The positive component of processing error totals 2.14 million across all post strata, and the negative component .15 million: $2.14 - .15 \doteq 2$ million. Line (ii) of Table 4 combines processing error with correlation bias, giving a smaller net error.
8. B3 Table 8 shows 10.7 million housing units with two forms, and .5 million with three or more. Another 2.4 million forms were flagged as duplicates very late in the process (B1 p. 2). We are reckoning 2 persons per form, although this is rather rough.
9. On heterogeneity in 1990, see Freedman and Wachter [13], Wachter and Freedman [28]. Data for 2000 are presented here, in Table 5. On correlation bias in 1990, see Wachter and Freedman [29].

10. The U. S. Supreme Court had already precluded the use of adjustment for reapportionment, that is, allocating congressional seats among the states; previously, it had upheld Secretary Mosbacher's decision not to adjust Census 1990. See 517 U. S. 1 (1996), 525 U. S. 316 (1999), available on-line at <http://supct.law.cornell.edu/supct/>. For discussion, see Brown *et al.* [5]. The Bureau's recommendation is explained in U. S. Census Bureau [24].

Other litigation. Efforts by Los Angeles and the Bronx among others to compel adjustment have been rejected by the courts (City of Los Angeles *et al.* v. Evans *et al.*, Central District, California); appeals are pending in the Ninth Circuit. Utah has sued to preclude the use of imputations but their suit was denied by the Supreme Court (Utah *et al.* v. Evans *et al.*, <http://supct.law.cornell.edu/supct/>). Members of the House have also sued to compel release of block-level adjusted counts; this case is pending (Waxman *et al.* v. Evans *et al.*, Central District, California), along with a similar case in the Southern District of Texas (Cameron County *et al.* v. Evans *et al.*).

11. B4, Appendix Table 1, Cols. 1–4. We can replicate the census and ACE figures from the data described in note 1.

12. On missing data, see B1 p. 4 and B6 p. 29; report B7 is useful too. Results for movers were computed from data described in note 1.

13. B1 p. 39. Also see Belin and Rolph [3], with a response by Freedman and Wachter [13]. In brief, the “validation” is a coincidence of two rates. One rate is computed from all PES data that required imputation. The other is a benchmark computed from all cases re-interviewed and resolved in Evaluation Followup, a smaller survey done many months after the Post Enumeration Survey, designed to check the quality of the PES. Only stronger cases were sent to Evaluation Followup, and of these, only the strongest could be found and resolved. Thus, the benchmark rate is computed from only 25% of the data, and by no means a random 25% either.

14. The Bureau has provided two sets of “targets” for each of the post strata, computed (i) with correlation bias, and (ii) without correlation bias. Each set comprises 1,000 replicates. The average target in set (i) represents their idea of the adjustment factors, corrected for processing error. The difference between the mean target in set (i) and the corresponding adjustment factor represents the estimated net effect of processing error, post stratum by post stratum. Similarly, the difference between the means of the two sets represents the allowance for correlation bias, post stratum by post stratum.

These target adjustment factors were developed for 416 “collapsed” post strata, e.g., post strata 06-2, 06-4, and 06-6 are pooled due to small sample size; we take the target for a pooled post stratum and treat it as the (common) target for the components.

The targets summarize the Bureau view on errors in ACE, as of March 2001. Error estimates were largely derived from 1990 data. See B13 and B19. In turn, the 1990 data mainly derive from the Evaluation Followup (note 13).

The biases for states and congressional districts were computed from the targets; adjustments were computed from other data described in note 1. By way of example, take

CD 1 in Alabama, whose population according to Census 2000 was 646,181. Adjustment would have added 7630 to this figure, of which 3983 is due to processing error. (This CD is the first in our data file.)

15. If the adjustment factor for a post stratum is greater than 1.00, adjustment adds to the population; in total, such post strata add 4.26 million to the census count: apparently, the census is below the true population for such post strata, and has made a negative error. If the adjustment factor is less than 1.00, adjustment subtracts from the population; in total, such post strata would subtract 1.00 million from the census: apparently, the census is above the true population for such post strata, and has made a positive error.

16. See note 14 on the sources of the data, and the targets. We take the mean target in set (i), with correlation bias, for each post stratum; subtract this from the corresponding adjustment factor; and multiply by the census count for the post stratum. The sum of the positive numbers is 1.75 million; the sum of the negatives, .90 million.

17. The “negative error” in the census comes from omissions, estimated by ACE as 4.26 million; but 1.75 million of this figure reflects errors in ACE, rather than census errors. The positive error in the census can be computed from lines (i) and (ii) as $1.00 - .90 = .10$.

18. Briefing by Census Bureau staff to congressional oversight committees, 21 March 2001.

19. Gross errors in ACE were discussed in Bureau report B19, and were also provided to us in computer-readable format (note 1); disentangling the sign conventions in the computer file seemed more trouble than it was worth, so we report gross error only. Gross errors reported by Freedman and Wachter [15, p. 31] were derived from B19 p. 80, and do not net out processing error within post strata.

20. Anderson and Fienberg [1, 2]; Fienberg [11].

21. The Bureau’s decision is explained in U. S. Census Bureau [25]; funding allocation is mentioned on p. 3. As Figure 1 suggests, however, the impact of adjustment on the distribution of funds would have been minor, at least in relative terms. Also see U. S. Commerce Department [26, Appendix 15]. See U. S. Census Bureau [25, p. 6] on DA; p. 10 reports that there were 3–4 million undetected duplications, which must be added to the previous figure of 2 million for processing error, giving the range 5–6 million; p. 11 discusses the 15 million cases that remain in doubt; and p. 12 mentions 10 million gross errors in mover status, with many other large gross errors mentioned elsewhere. The report notes on p. 13 that missing data create “considerable” uncertainty; the quantification at 500,000 is optimistic, as discussed on p. 14. The numbers may change when further analysis is done. Balancing error is no longer considered a serious problem by the Bureau.

22. Arguments for increasing DA estimates were made by Jeff Passel at the Urban Institute and Bureau demographers, although the two groups reach somewhat different conclusions. Our own estimates were presented at a conference on census adjustment

in Berkeley, on 24 September 2001. For details on correlation bias and data from 1990, see Wachter and Freedman [29]. The Bureau has so far not recognized the problem (B1 p. 45). While B12 p. 16 grants our premises, it denies the arithmetic. We return to this point when discussing loss function analysis.

23. Some persons are counted in the census with no personal information, in which case all their personal characteristics are imputed. “Late census adds” are persons who file a census form too late to be run through the ACE process. These may or may not be “data-defined,” i.e., have enough characteristics for matching. Our late adds include the non-data-defined late adds, whereas our IIs exclude those records.

As a further complication (note 8), 2.4 million forms were found to be duplicates late in the process; of these, about 1.4 million were taken completely off the table, but 1.0 million were put into the ACE process as late adds. These forms correspond to 2.3 million people, and in 2000, the late adds were basically just these people. (Neither imputes or late adds are eligible for matching in ACE; such records are subtracted from the census, on the theory that the corresponding persons—if they really exist—will come back as gross omissions.) For details and an attempted explanation of the logic, see Q43.

For consistency, Table 5 only covers the “ACE target population,” i.e., persons living in group quarters or institutions are excluded, as is remote rural Alaska.

24. Freedman and Wachter [13, p. 482] make the connection with analysis of variance, identifying the inequalities that must hold—and those that can be violated. As indicated there, results may depend on weights. Table 5 is unweighted: all states are treated as equals, likewise for post strata and “chunks.” Undoubtedly, some of the effect in Table 5 is due to random variation in the smallest chunks—by the time you spread the population over 448 post strata, 50 states (and D. C.), some of the cells have to be tiny. However, the variance correction in equation (4) of Wachter and Freedman [28] does not affect the pattern in the table, although estimated heterogeneity is reduced. Similarly, we can restrict attention to chunks with a census population in excess of 100, for instance. Or, we can restrict attention, say, to PSGs #1–60 or even to #1–48, #53–60: this markedly reduces the variation due to small chunks without changing the pattern in Table 5. (The excluded PSGs have small populations, and cut across 50 states as well as D.C.) The table does not seem to be an artifact of small-sample variation.

Congressional districts. We were able to compute the analog of Table 5 for congressional districts. The SD across districts is about 1.5 times the SD across states. The SD within post strata across districts is about 3.5 times the SD within post strata across states, for the 448 post strata. Going to 64 or 16 post strata does not change the SD within post strata across districts. These results apply to both proxies.

25. See B19. We used the census count for the ACE target population, as defined at the end of note 23.

26. Some additional summary statistics may be of interest, for instance, for adjustments to congressional districts in Table 6 and Figure 2 (counts). The average district had a

census population of 646,000 in 2000; adjustment by 448 post strata would have added 7500 persons on average, compared to 7200 from the 16 evaluation post strata; the SDs are 3800 and 3000. From the perspective of “loss function analysis,” the coarser post stratification is to be preferred (last paragraph of note 30 below).

27. A “block cluster” contains one block in densely populated areas, and many contiguous blocks in the hinterland. On average, there seem to be about two blocks per cluster. See, e.g., pp. 1048, 1088–89 in Hogan [19] for 1990 sample sizes and B11 p. 3 for 2000 data. There are published discussions of problems with Bureau estimates of variance for bias estimators: see, e.g., Freedman *et al.* [17, pp. 268–69], Freedman and Wachter [13, pp. 532–33]. These papers explain loss function analysis in some detail; also see Mulry and Spencer [20], Brown *et al.* [5]. For 2000, see B13 and B19.

The factor of 4 comes from comparing the traces of the covariance matrices for the adjustment factors and the targets (note 14). For a discussion of the process for generating the targets, see B13 and B19; Fay and Thompson [10, p. 77] may be clearer, or Mulry and Spencer [20, p. 1089], although these refer to the 1990 adjustment. On the 1990 variances, see Freedman *et al.* [16].

28. We are using the ACE target population as defined at the end of note 23. (This is due to a quirk in our computer code.) The full census population of California (including group quarters and institutions) is 33.9 million.

29. The ratio is $24,558/5,208 \doteq 4.715$, compared to 4.895 in B13, Table 1A, line 1. We also replicated the other results in line 1, to the same precision. We have not replicated the targets themselves; preliminary calculations suggest some incongruities (Don Ylvisaker, personal communication), but to resolve these, additional data would be needed.

30. The Bureau’s “preferred model” for correlation bias corresponds to line 1 of Table 1A in B13; this version of correlation bias is built into the set of targets and hence the estimated biases (note 14) that underlie Tables 7–11.

Weights. See B13 p. 5, where the loss function for state counts is weighted inversely to population; for shares, weights are inverse to population share. We follow the Bureau, but weight by the census rather than the adjusted census or bias-corrected adjusted census, to avoid random weights. Due to the aforementioned quirk in our code (notes 28), we weight by census counts for the ACE target population in Tables 7 and 10, rather than the full census count. (Our code for shares uses the full census count.) For reasons that are at best obscure, the Bureau’s loss function for congressional districts uses shares within-state, but weights the states proportional to the square of their census count. We followed suit in replication, but not in Table 9.

Terminology. The ACE target population (defined at the end of note 23) has nothing to do with the targets in note 14.

The model behind loss function analysis. The model, and the procedure for estimating the SE of Diff, are discussed in Freedman *et al.* [17]; also see Freedman and Wachter [14, pp. 368–70] or Brown *et al.* [5, pp. 372–75]. For state counts, say, let C be the

51 × 448 matrix whose (i, j)th entry is the census count for state i intersected with post stratum j. Let $\hat{\phi}$ be the 448 × 1 vector of adjustment factors, and $\hat{\psi}$ the 448 × 1 vector of estimated biases in $\hat{\phi}$. Write $E(\hat{\phi}) = \phi$ and $E(\hat{\psi}) = \psi$. The little model behind loss function analysis assumes that $\hat{\psi}$ is unbiased, so that ψ is the “true” bias in the adjustment factors, and true population counts for the states are given by $C(\phi - \psi)$; estimators are jointly normal, with $\hat{\phi}$ independent of $\hat{\psi}$; finally $\text{cov}(\hat{\phi})$ and $\text{cov}(\hat{\psi})$ are taken as given.

The adjusted state counts (ACE target population) are given by $C\hat{\phi}$; likewise, the biases in these counts are estimated by $C\hat{\psi}$. The covariances are

$$(1) \quad \text{cov}(C\hat{\phi}) = C \text{cov}(\hat{\phi}) C' \quad \text{and} \quad \text{cov}(C\hat{\psi}) = C \text{cov}(\hat{\psi}) C'.$$

Of course, when (1) is applied to data, the covariance matrices $\text{cov}(\hat{\phi})$ and $\text{cov}(\hat{\psi})$ would be replaced by sample-based estimates. The estimated census risk is

$$(2) \quad \text{Cen} = \|C(\hat{\phi} - \hat{\psi} - 1)\|^2 - \text{trace cov}(C\hat{\phi}) - \text{trace cov}(C\hat{\psi}).$$

The estimated risk from adjustment is

$$(3) \quad \text{Adj} = \|C\hat{\psi}\|^2 + \text{trace cov}(C\hat{\phi}) - \text{trace cov}(C\hat{\psi}).$$

Notice that $\text{trace cov}(C\hat{\psi})$ cancels when computing $\text{Diff} = \text{Cen} - \text{Adj}$.

The census population, adjustment factors, and their covariances were supplied by the Bureau (note 1). Biases are computed from the targets (note 14), and their covariance is the empirical covariance of the 1000 sets of targets provided by the Bureau. In Tables 7–11, we are using the targets that correspond to the “preferred model” for correlation bias (see the beginning of this note; the “preferred model” is for correlation bias, not loss function analysis more generally).

The SE of Diff. Index the the 50 states and DC by $k = 1, \dots, 51$. Let μ_k be the error in the census population count for area k. Let X_k be the production dual system estimate for μ_k ; thus, $X = C(\hat{\phi} - 1)$. The bias in X_k is denoted β_k ; this is estimated as $\hat{\beta} = C\hat{\psi}$. Let $G = \text{cov}(X)$ and $H = \text{cov}(\hat{\beta})$, which are assumed known. The estimated risk from the census for area k is $(X_k - \hat{\beta}_k)^2 - G_{kk} - H_{kk}$, while the estimated risk from adjustment is $\hat{\beta}_k^2 + G_{kk} - H_{kk}$. The estimated risk difference is

$$(4) \quad \hat{R}_k = (X_k - \hat{\beta}_k)^2 - \hat{\beta}_k^2 - 2G_{kk}.$$

By a tedious but routine calculation,

$$(5) \quad \text{cov}(\hat{R}_i, \hat{R}_j) = 4\mu_i\mu_j G_{ij} + 2G_{ij}^2 + 4E(X_i X_j) H_{ij}.$$

The displayed covariance can be estimated from sample data as

$$(6) \quad \hat{R}_{ij} = 4(X_i - \hat{\beta}_i)(X_j - \hat{\beta}_j)\hat{G}_{ij} - 2G_{ij}^2 - 4X_i X_j H_{ij},$$

and $\text{var}(\text{Diff})$ can be estimated as $\sum_{ij} \hat{K}_{ij}$. When (4) and (6) are applied to data, the covariance matrices G and H would be replaced by sample-based estimates, whose variability is ignored in (5).

Issues. Of course, G and H are derived from $\text{cov}(\hat{\phi})$ and $\text{cov}(\hat{\psi})$ respectively—equation (1). The calculations take these matrices as known, or at least estimated in a reasonable way. In fact, our calculations suggest that $\text{trace } \widehat{\text{cov}}(\hat{\psi})$ is too small by a factor of 50 or more, $\widehat{\text{cov}}(\hat{\psi})$ being the Bureau's estimate for $\text{cov}(\hat{\psi})$. But scaling $\widehat{\text{cov}}(\hat{\psi})$ by a constant factor may not be a reliable correction. A more sensible thing to do is to make direct estimates of bias, and compute $\text{cov}(\hat{\psi})$ by jackknifing. (This would require much more data than we have.) The calculations also assume that estimates for the biases in the DSE are unbiased, which is rather questionable.

At some level, the Bureau must have been aware of the problems with $\text{cov}(\hat{\psi})$ when it considered adjusting the post-censals in 1992. If it believed in its own estimates for $\text{cov}(\hat{\psi})$, it would have adjusted not by $\hat{\phi}$ but by $\hat{\phi} - \hat{\psi}$. See Fay and Thompson [10, p. 74].

Footnote to a footnote. Loss function analysis “shows” that adjustment by the 16 EPS (Section 13) is comparable to or even better than the production adjustment (448 post strata). We adjust starting from the direct DSEs for the EPS (defined in Section 12), and evaluate using the Bureau's preferred targets. An exact comparison waits on the 16×16 covariance matrix for the direct DSEs, which we do not have; but a factor of 1.5 on estimated MSEs for state counts, unweighted, is plausible—within the peculiar conventions of loss function analysis.

31. On 1990, see Mulry and Spencer [20] or Fay and Thompson [10]. The Undercount Steering Committee cited loss function analysis throughout its report on the 1990 adjustment (U. S. Commerce Department [26] Appendix 4).

32. Also see U. S. Census Bureau [25, p. 17].

33. Some examples give the flavor.

Processing error. Cohen, White, and Rust [7, Chap. 4] defend the adjustment of 1990 without mentioning Breiman's [4] estimates of processing error, although they do take up one of his minor points (pp. 73–74).

Heterogeneity. According to Cohen *et al.* (p. 78), Bureau staff were in the early 1990s “well aware” of problems with the synthetic assumption. This may be so, after the adjustment decisions were made: Fay and Thompson [10]. While those decisions were being made, however, the facts were less evident. For instance, the Bureau's Undercount Steering Committee was “convinced from the data, that, in general, block parts are homogeneous within post-strata”; they saw “no evidence to indicate there are any serious flaws” in the homogeneity assumption at the state level. U. S. Commerce Department [26, Appendix 4 p. 7].

The imputation model. Cohen *et al.* (p. 72) defend the putative validation in 1990 (note 13) as “strong evidence.” They ignore data showing the flimsiness of this evidence (Freedman and Wachter [13] pp. 535–56).

Correlation bias. Cohen *et al.* (p. 82) argue that relatively few people are added to the counts by statistical modeling. The point is that *more* people need to be added—but not in the places expected by the modelers. On the same page, Cohen *et al.* respond to a minor point in Darga [8], relating to correlation bias—but fail to address the major finding: adjustment would have had a perverse effect on many sex ratios.

Smoothing and loss function analysis. Variances for adjustment factors in 1990 are discussed in Freedman *et al.* [16]. The variances were obtained from a “smoothing model.” Cohen *et al.* (pp. 58–62) defend the model, although they do not state the crucial assumptions, nor do they respond to our points. Similarly, Cohen *et al.* (pp. 68–85) defend loss function analysis, without responding to the points raised in Section 14—and in our previous publications (Freedman *et al.* [17]; Freedman and Wachter [13]). Also see Brown *et al.* [5, pp. 364–65, pp. 371–75].

Citro, Cork, and Norwood [6, p. 33] find that the Bureau’s decision not to adjust was “justifiable,” but the “fact that the Bureau did not recommend adjusting the census counts to be provided for redistricting does not carry any implications for the usefulness of statistical adjustment methods based on dual-system estimation.”

34. The 16 evaluation post strata could probably be reconfigured with advantage. Among other things, one of the major variables used to define the post strata—the mail back rate—turned out to have paradoxical features: many post strata with high mail back rates had high rather than low measured undercounts.

Sampling error in the PES is relatively easy to quantify, and is not the major problem, at least at the national level. For example, the standard error for the national undercount estimate of 3.3 million is around 400,000. See B19 Table 20. Reducing the sample size by a factor of 30 would increase the SE to roughly $\sqrt{30} \times .4 \doteq 2$ million. (Of course, our calculations only give rough guidelines.) On this basis, sampling error would still be dominated by uncertainties due to non-sampling error (see Tables 1–4 and Section 11). The extent to which non-sampling error could be reduced with a smaller PES remains to be seen.

In 2000, ACE matched outmovers rather than inmovers. This treatment of movers came about by historical accident (Brown *et al.* [5] p. 367). The Bureau had planned to do sample-based non-response followup (SNRFU). With SNRFU, inmovers cannot be matched back to the census blocks they came from: the records may not be in the census because the people were not selected into the SNRFU sample. The Supreme Court over-ruled the Bureau on SNRFU, as part of the decision cited in note 10, but outmovers stayed in the plan, probably due to inertia. If there is a PES in 2010, let it match inmovers not outmovers—like the PES of 1990.

Authors’ footnote. We thank Liza Levina, Chad Schafer and Don Ylvisaker for many useful comments. Both of us testified for the United States against adjustment, in the Census cases of 1980 and 1990. We have also testified in Congressional hearings, and consulted for the Department of Commerce on the adjustment decision.

Finally, a word about Terry Speed. Terry is a wonderful friend and colleague, who

for years has helped us by answering all kinds of questions, not least about census adjustment – although that is only one of his many interests.

David A. Freedman, Department of Statistics, University of California, Berkeley,
freedman@stat.berkeley.edu

Kenneth W. Wachter, Department of Demography, University of California, Berkeley,
wachter@demog.berkeley.edu

References

- [1] M. Anderson and S. E. Fienberg. *Who Counts? The Politics of Census-Taking in Contemporary America*. Russell Sage Foundation, New York, 1999.
- [2] M. Anderson and S. E. Fienberg. Census 2000 and the politics of census taking. *Society*, 39:17–25, 2001.
- [3] T. R. Belin and J. E. Rolph. Can we reach consensus on census adjustment? *Statistical Science*, 9:458–537, 1994 (with discussion).
- [4] L. Breiman. The 1991 census adjustment: Undercount or bad data? *Statistical Science*, 9:458–537, 1994 (with discussion).
- [5] L. D. Brown, M. L. Eaton, D. A. Freedman, S. P. Klein, R. A. Olshen, K. W. Wachter, M. T. Wells, and D. Ylvisaker. Statistical controversies in Census 2000. *Jurimetrics*, 9:347–375, 1999.
- [6] C. F. Citro, D. L. Cork, and J. L. Norwood, editors. *The 2000 Census: Interim Assessment*. National Academy Press, Washington, D. C., 2001.
- [7] M. L. Cohen, A. A. White, and K. F. Rust, editors. *Measuring a Changing Nation: Modern Methods for the 2000 Census*. National Academy Press, Washington, D. C., 1999.
- [8] K. Darga. *Sampling and the Census*. The AEI Press, Washington, D. C., 1999.
- [9] K. Darga. *Fixing the Census until It Breaks*. Michigan Information Center, Lansing, 2000.
- [10] R. E. Fay and J. H. Thompson. The 1990 post enumeration survey: Statistical lessons, in hindsight. In *Proceedings, Bureau of the Census Annual Research Conference*, Washington, D. C., 1993. Bureau of the Census.
- [11] S. E. Fienberg. The New York City census adjustment trial: Witness for the plaintiffs. *Jurimetrics*, 34:65–83, 1993.

- [12] D. A. Freedman, P. B. Stark, and K. W. Wachter. A probability model for census adjustment. *Mathematical Population Studies*, 9:165–180, 2001.
- [13] D. A. Freedman and K. W. Wachter. Heterogeneity and census adjustment for the intercensal base. *Statistical Science*, 9:458–537, 1994 (with discussion).
- [14] D. A. Freedman and K. W. Wachter. Planning for the census in the year 2000. *Evaluation Review*, 20:355–377, 1996.
- [15] D. A. Freedman and K. W. Wachter. Census adjustment: Statistical promise or illusion? *Society*, 39:26–33, 2001. Sections 1–9 are adapted from this source.
- [16] D. A. Freedman, K. W. Wachter, D. Coster, R. Cutler, and S. Klein. Adjusting the census of 1990: The smoothing model. *Evaluation Review*, 17:371–443, 1993.
- [17] D. A. Freedman, K. W. Wachter, R. Cutler, and S. Klein. Adjusting the census of 1990: Loss functions. *Evaluation Review*, 18:243–280, 1994.
- [18] N. Hengartner and T. P. Speed. Assessing between-block heterogeneity within the poststrata of the 1990 post-enumeration survey. *Journal of the American Statistical Association*, 88:1119–1125, 1993.
- [19] H. Hogan. The 1990 post-enumeration survey: Operations and results. *Journal of the American Statistical Association*, 88:1047–1060, 1993.
- [20] M. H. Mulry and B. D. Spencer. Accuracy of 1990 census and undercount adjustments. *Journal of the American Statistical Association*, 88:1080–1091, 1993.
- [21] K. Prewitt. Accuracy and coverage evaluation: Statement on the feasibility of using statistical methods to improve the accuracy of Census 2000. *Federal Register*, 65:38373–38398, 2000.
- [22] P. Skerry. *Counting on the Census? Race, Group Identity, and the Evasion of Politics*. Brookings, Washington, D. C., 2000.
- [23] P. B. Stark. Review of *Who Counts?* *Journal of Economic Literature*, 39:592–595, 2001.
- [24] U. S. Census Bureau. *Report of the Executive Steering Committee for Accuracy and Coverage Evaluation Policy*. With supporting documentation, Reports B1–B24. Washington, D. C., 2001. <http://www.census.gov/dmd/www/EscapeRep.html>.
- [25] U. S. Census Bureau. *Report of the Executive Steering Committee for Accuracy and Coverage Evaluation Policy on Adjustment for Non-Redistricting Uses*. With supporting documentation, Reports 1–24. Washington, D. C., 2001. <http://www.census.gov/dmd/www/EscapeRep2.html>.

- [26] U. S. Commerce Department. Office of the Secretary. *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 the Population, Explanation*. Three volumes. Reprinted in part in *Federal Register* 56: 33582–33642 (July 22). Washington, D. C., 1991.
- [27] K. W. Wachter. Ignoring nonignorable effects. *Journal of the American Statistical Association*, 88:1161–1163, 1993.
- [28] K. W. Wachter and D. A. Freedman. Measuring local heterogeneity with 1990 U. S. census data. *Demographic Research*, an on-line journal of the Max Planck Institute. Volume 3, Article 10, 2000. <http://www.demographic-research.org/Volumes/Vol3/10/> .
- [29] K. W. Wachter and D. A. Freedman. The fifth cell. *Evaluation Review*, 24:191–211, 2000.
- [30] D. Ylvisaker. Review of *Who Counts?* *Journal of the American Statistical Association*, 96:340–341, 2001.
- [31] A. M. Zaslavsky. Combining census, dual system, and evaluation study data to estimate population shares. *Journal of the American Statistical Association*, 88:1092–1105, 1993.