Methods for Census 2000 and Statistical Adjustments

David A. Freedman and Kenneth W. Wachter April 2004

Abstract

This article outlines procedures for taking the US census, making adjustments, and evaluating the results. The census turns out to be 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 are comparable to if not larger than errors in the census. The data suggest a strong geographical pattern to such errors even after controlling for demographic variables, which contradicts basic premises of adjustment. In fact, the complex demographic controls built into the adjustment process seem on whole to have been counter-productive.

1. Introduction

The census has been taken every ten years since 1790, and provides a wealth of demographic information for researchers and policy makers. Beyond that, counts are used to apportion Congress and redistrict states. Moreover, 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 more than counts for the nation as a whole. Data from 1990 and previous censuses suggested there would be a net undercount in 2000. Furthermore, 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 turned out to be remarkably good, despite much critical commentary. Statistical adjustment was unlikely to improve the accuracy, because adjustment can easily put in more error than it takes out.

We will sketch procedures for taking the census, making adjustments, and evaluating results. (Detailed descriptions cover thousands of pages; summaries are a necessity.) Data will be presented on errors in the census, in the adjustment, and on geographical variation in error rates. Alternative adjustments are discussed, as are methods for comparing the accuracy of the census and the adjustments. There are pointers to the literature, including citations to the main arguments for and against adjustment. The present article is based on Freedman and Wachter (2003), which may be consulted for additional detail and bibliographic information.

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 rangeland.) 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 special populations.

3. Demographic Analysis

DA (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 DA, the undercount in 1970 was about 3 percent nationally. In 1980, it was 1 to 2 percent, and the result for 1990 was similar. DA reported the undercount for Blacks at about 5 percentage points above non-Blacks, in all three censuses.

DA 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 collect 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, DA estimates the number of such persons starting from Medicare records. Despite its flaws, DA has generally been considered to be the best yardstick for measuring census undercounts. Recently, however, another procedure has come to the fore, 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 was renamed ACE ("Accuracy and Coverage Evaluation Survey"). 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. 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 though 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 in the sample blocks. As the discussion of processing error shows, however, matching is easier said than done.

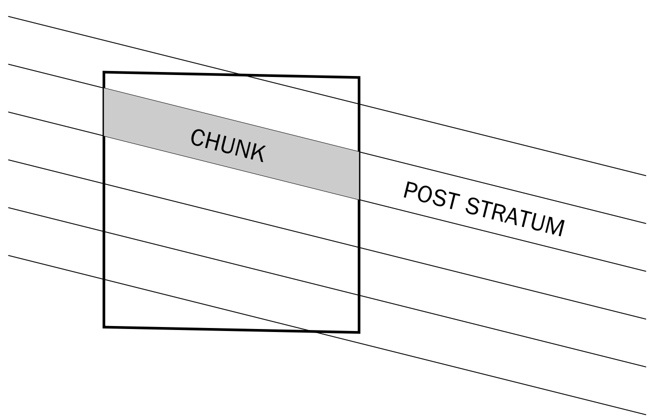
Some persons are missed both by the census and by ACE. Their number is estimated using a statistical model, assuming that ACE is as likely to find people missed by the census as people counted in the census—"the independence assumption." Following this assumption, a gross omission rate estimated from the people found by ACE can be extrapolated to people in the census who were missed by ACE, although the true gross omission rate for that 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 1 person in 500 were sampled, each person in the sample 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 are given different weights. To estimate the total number of gross omissions in a post stratum, one 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. In more detail, ACE data are used to compute an *adjustment factor* for each post stratum. 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.



How to adjust small areas like blocks, cities, or states? Take any particular area. As the sketch indicates, this area will be carved up into "chunks" by post strata. Each chunk 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 here. 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.

Heterogeneity is 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 in Section 11 below, show that heterogeneity is a serious concern.

The adjustment issue was often framed in terms of sampling: "sampling is scientific." However, from a technical perspective, sampling is not the point. 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, the gains and losses must balance. This point was often overlooked in the political debate. In 2000, even more so than in 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.

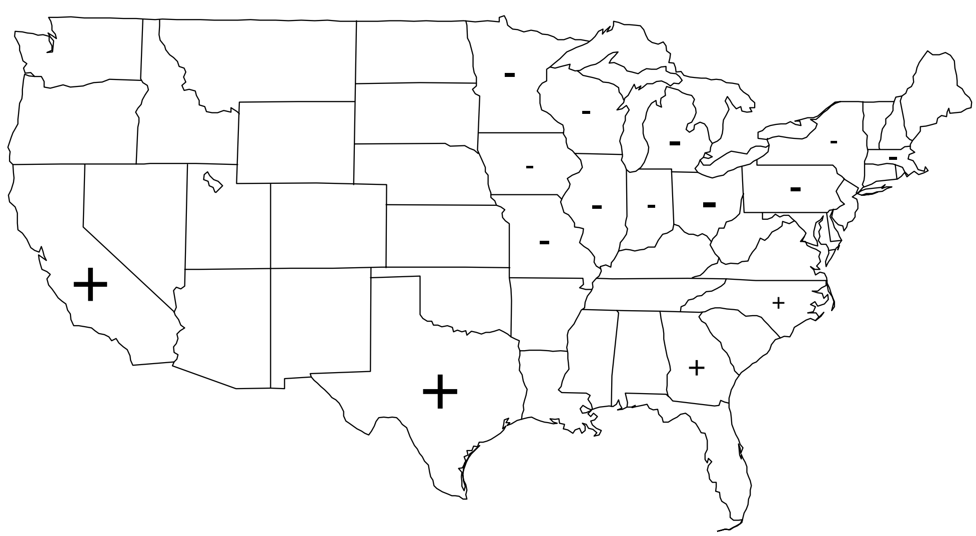


Figure 1. ACE Adjustment: State Share Changes Exceeding 50 Parts Per Million

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.

7. The 1990 Adjustment Decision

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. 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; given the data that are available, parceling the figures down to local areas would require heroic assumptions.)

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 meanings.) 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 (Table 1). On the scale of interest, most of the estimated undercount is noise.

Table 1. Errors in the adjustment of 1990

The adjustment	+5.3		
Processing error	−3.6		
Corrected adjustment		+1.7	
Correlation bias		+3.0	
Demographic Analysis			+4.7

8. Census 2000

Census 2000 succeeded in reducing differential undercounts from their 1990 levels. That sharpened questions about the accuracy of proposed statistical adjustments. Errors in statistical adjustments are not 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.

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. Troubles 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. But how many slipped through? And how many of those were missed by ACE?

Besides processing error, correlation bias is an endemic problem that makes it 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. Surely these people are unevenly distributed across the country ("differential correlation bias"). The more uneven is the distribution, the more distorted a picture of census undercounts is created by the DSE.

9. The Adjustment Decision for Census 2000

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). The principal reason was that, according to DA, 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, DA and ACE pointed in opposite directions. The three population totals are shown in Table 2.

Table 2. The population of the United States

Demographic Analysis	279.6 million
Census 2000	281.4 million
ACE	284.7 million

If DA is right, there is a census overcount of .7 percent. If ACE is right, there is a census undercount of 1.2 percent. DA 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 DA is hardly perfect, it was a stretch to blame DA for the whole of the discrepancy with ACE. Instead, the discrepancy pointed to undiscovered error in ACE. When the Secretary made his decision, there was some information on missing data and on the influence of movers, summarized in Table 3.

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.

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

The 2000 adjustment tried to identify both inmovers and outmovers, a departure from past practice. Gross omission rates were computed for the outmovers and applied to the inmovers, although it is not clear why rates are equal within local areas. For outmovers, 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 outmovers. That is why the number of outmovers is so much less than the number of inmovers. 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 nearly good enough for adjusting the census.

10. Gross or Net?

Errors can reported either gross or net, and there are many possible ways to refine the distinction. (Net error allows overcounts to balance undercounts; gross error does not.) Some commentary suggests that the argument for adjustment may be stronger if gross error is the yardstick. Certain places may have an excess number of census omissions while other places will have an excess number of erroneous enumerations. Such imbalances could be masked by net error rates, when errors of one kind in one place offset error of another kind in another place. In this section, we consider gross error rates.

Some number of persons were left out of Census 2000 and some were counted in error. 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 error of 9.5 million in Census 2000.

Before presenting comparable numbers for ACE, we mention some institutional history. The census is used as a base for post-censal population estimates. This may sound even drier than redistricting, but $200 billion a year of tax money are allocated using post-censal estimates. In October 2001, the Bureau revisited the adjustment issue: should the census be adjusted as a base for the post-censals? The decision against adjustment was made after further analysis of the data. Some 2.2 million persons were added to the Demographic Analysis. Estimates for processing error in ACE were sharply increased. Among other things, ACE had failed to detect large numbers of duplicate enumerations in the census, because interviewers did not get accurate census-day addresses from respondents. That is why ACE had over-estimated the population. The Bureau's work confirmed that gross errors in ACE were well above 10 million, with another 15 million cases whose status remains to be resolved. Error rates in ACE are hard to determine with precision, but are quite large relative to error rates in the census.

11. Heterogeneity in 2000

This section demonstrates that substantial heterogeneity remains in the data, despite elaborate post stratification. In fact, post stratification seems on the whole to be counter-productive. Heterogeneity is measured as in Freedman and Wachter (1994, 2003), with SUB ("whole-person substitutions") and LA ("late census adds") as proxies—surrogates—for the undercount: see the notes to Table 5. For example, .0210 of the census count (just over 2%) came from whole-person substitutions. This figure is in the first line of the table, under the column headed "Level". Substitution rates are computed not only for the whole country, but for each of the 435 congressional districts: the standard deviation of the 435 rates is .0114, in the "Across CD" column. The rate is also computed for each post stratum: across the 448 post strata, the standard deviation of the substitution rates is .0136, in the "Across P-S column," The post strata exhibit more variation than the geographical districts, which is one hallmark of a successful post stratification.

To compute the last column of the table, we think of each post stratum as being divided into "chunks" by the congressional districts. We compute the substitution rate for each chunk with a non-zero census count, then take the standard deviation across chunks within post stratum, and finally the root-mean-square over post strata. The result is .0727, in the last column of Table 4. If rates were constant across geography within post strata, as the homogeneity assumption requires, this standard deviation should be 0. Instead, it is much larger than the variability across congressional districts. This points to a serious failure in the post stratification. If the proxies are good, there is a lot of heterogeneity within post strata across geography.

Similar calculations can be made for two coarser post stratifications. (i) The Bureau considers its 448 post strata as coming from 64 PSGs. (Each PSG, or "post-stratum group," divides into 7 age-sex groups, giving back $64 \times 7 = 448$ post strata.) The 64 PSGs are used as post strata in the second line of Table 4. (ii) The Bureau groups PSGs into 16 EPS, or "evaluation post strata." These are the post strata in the third line of Table 4. Variability across post strata or within post-strata across geography is not much affected by the coarseness of the post stratification, which is surprising. Results for late census adds (LA) are similar, in lines 4–6 of the table. Refining the post stratification is not productive. There are similar results for states in Freedman and Wachter (2003).

The Bureau computed "direct DSEs" for the 16 evaluation post strata, by pooling the data in each. From these, an adjustment factor can be constructed, as the direct DSE divided by the census count. We adjusted the United States using these 16 factors rather than the 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.

Table 4. Measuring heterogeneity across Congressional Districts (CD). In the first column, post stratification is either (i) by 448 post strata; or (ii) by the 64 post-stratum groups, collapsing age and sex; or (iii) by the 16 evaluation post strata. "SUB" means whole-person substitutions, and "LA" is late census adds. In the last two columns, "P-S" stands for post strata; there are three different kinds, labelled according to row.

		Standard Deviation		
Proxy & Post Stratification	Level	Across CD	Across P-S	Within P-S across CD
SUB 448	.0210	.0114	.0136	.0727
SUB 64	.0210	.0114	.0133	.0731
SUB 16	.0210	.0114	.0135	.0750
LA 448	.0085	.0054	.0070	.0360
LA 64	.0085	.0054	.0069	.0363
LA 16	.0085	.0054	.0056	.0341

Notes. The level of a proxy does not depend on the post stratification, and neither does the SD across CDs. These two statistics do depend on the proxy. A "substitution" is a person counted in the census with no personal information, which is later imputed. A "late add" is a person originally thought to be a duplicate, but later put back into the census production process. Substitutions includes late adds that are not "data defined," i.e., do not have enough information for matching. Substitutions and late adds have poor data quality, which is why they may be good proxies for undercount. Table 5 in Freedman and Wachter (2003) uses slightly different conventions and includes the District of Columbia.

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 (7630, 7486). The correlation between the 435 pairs of changes is .87, as shown in the third line of Table 6. 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. On this basis, and on the basis of Table 4, we suggest that 448 post strata are no better than 16. (For some geographical areas with populations below 100,000, however, the two adjustments are likely to be different.)

Tables 4–5 and Figure 2 show that an elaborate post stratification does not remove much heterogeneity. We doubt that heterogeneity can be removed by the sort of post stratification—no matter how elaborate—that can be constructed in real census conditions. The impact of heterogeneity on errors in adjustment is discussed by Freedman and Wachter (1994, pp. 479–81). Heterogeneity is more of a problem than sampling error.

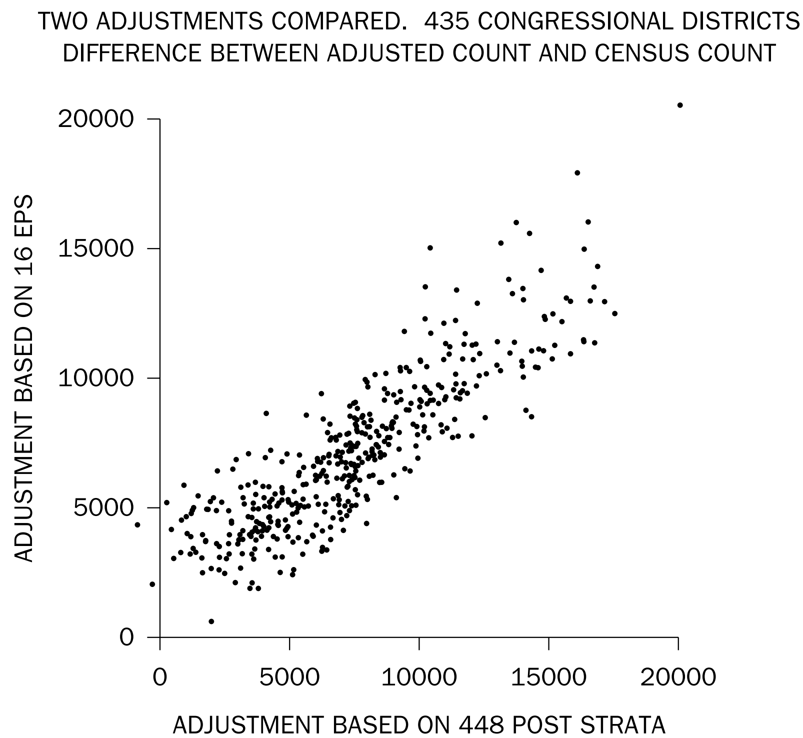


Figure 2. Changes to congressional district populations. The production adjustment, with 448 post strata, is plotted on the horizontal. An alternative, based only on the 16 evaluation post strata (EPS), is plotted on the vertical.

Table 5. Comparing the production adjustment based on 448 post strata to one based on 16 evaluation post strata. Correlation coefficients for changes due to adjustments.

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

Within a state, districts are by case law almost exactly equal in size—when redistricting is done shortly after census counts are released. Over the decade, people move from one district to another. Variation in population sizes at the end of the decade is therefore of 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. Exact equality at the beginning of the decade does not seem like a compelling goal.

12. Loss Function Analysis

A statistical technique called "loss function analysis" has been used to justify adjustment. In effect, this technique attempts to make summary estimates of the error levels in the census and the adjustment. 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 assumptions than on data.

For example, loss function analysis depends on models for correlation bias, and the model used in 2000 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 (2000). A second example: loss function analysis depends on having precise estimates of error rates in ACE. But there is considerable uncertainty about these error rates, even at the national level (Sections 9 and 10). A last 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. That is how processing errors and correlation bias in ACE can be parceled out to local areas without creating unmanageably large variances. But these homogeneity assumptions are not tenable (Section 11).

13. Pointers to the Literature

Reviews and discussions of the 1980 and 1990 adjustments can be found in *Survey Methodology* 18 (1992) 1–74, *Journal of the American Statistical Association* 88 (1993) 1044–1166, and *Statistical Science* 9 (1994) 458–537. 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 (1999), but see Stark (2001) and Ylvisaker (2001). Prewitt (2000) may be a better source, and Zaslavsky (1993) is often cited. Cohen, White, Rust (1999) try to answer arguments on the 1990 adjustment; but see Freedman and Wachter (2003). Skerry (2000) has an accessible summary of the issues. Darga (2000) is a critic. Freedman, Stark and Wachter (2001) have a probability model for census adjustment, which may help to clarify some of the issues.

The decision against adjustment for 1990 is explained in U. S. Department of Commerce (1991). On the 2000 adjustment decision, see US Bureau of the Census (2001ab), U. S. Census Bureau (2003). For another perspective on Census 2000, see Citro, Cork, and Norwood (2004). Problems with the PES, especially with respect to detecting duplicates, are discussed at pp. 214ff and 240ff. However, there is residual enthusiasm for a PES in 2010 and a corresponding lack of enthusiasm for Demographic Analysis (p. 8). Cork, Cohen, and King (2004) reach different conclusions (p. 11).

14. Litigation

The Commerce Department's decision not to adjust the 1980 census was upheld after trial. Cuomo v Baldrige, 674 F.Supp. 1089 (S.D.N.Y. 1987). The Department's decision not to adjust the 1990 census was also upheld after trial and appeal to the Supreme Court. 517 U.S. 1 (1996). Later in the decade, the Court found that use of adjustment for reapportionment, that is, allocating congressional seats among the states, violated the Census Act. 525 U.S. 316 (1999). The administration had at the time planned to adjust, so the Court's decision necessitated a substantial revision to the design of ACE. Brown et al. (1999).

Efforts by Los Angeles and the Bronx among others to compel adjustment of Census 2000 were rejected by the courts (City of Los Angeles et al. v. Evans et al., Central District, California); the decision was upheld on appeal to the Ninth Circuit. 307 F.3d 859 (9th Cir. 2002). There was a similar outcome in an unpublished case, Cameron County et al. v. Evans et al., Southern District, Texas. Utah sued to preclude the use of imputations but the suit was denied by the Supreme Court. Utah et al. v. Evans et al. 536 U.S. 452 (2002).

The Commerce Department did not wish to release block-level adjusted counts, but was compelled to do so as a result of several law suits. The lead case was Carter v. U.S. Dept. of Commerce in Oregon. The decision was upheld on appeal to the Ninth Circuit. 307 F.3d 1084 (9th Cir. 2002).

15. Other Countries

For context, this section gives a bird's-eye view of the census process in a few other countries. In Canada, the census is taken every five years (1996, 2001, 2006, . . .). Unadjusted census counts are published. Coverage errors are estimated, using variations on the PES (including a "reverse record check") and other resources. A couple of years later, when the work is complete, post censal population estimates are made for provinces and many sub-provincial areas. These estimates are based on adjusted census counts. The process in Australia is similar; the PES there is like a scaled-down version of the one in the US.

In the UK, the census is taken every ten years (1991, 2001, 2011, . . .). Coverage errors are estimated using a PES. Only the adjusted census counts are published. The official acronym is ONC, for One-Number Census. Failure to release the original counts cannot enhance the possibility of informed discussion. Moreover, results dating back to 1982 are adjusted to agree with current estimates. "Superseded" data sets seem to be withdrawn from the official UK web page (www.statistics.gov.uk). Anomalies are found in the demographic structure of the estimated population (not enough males age 20–24). See Redfern (2004); also see pp. 17 and 48 in

http://www.statistics.gov.uk/downloads/theme_population/PT113.pdf

In Scandinavian countries, the census is based on administrative records and population registries. In Sweden, for example, virtually every resident has a PIN; the authorities try to track down movers—even persons who leave the country. Norway conducted a census by mail in 2001, to complete its registry of housing; but is switching to an administrative census in the future. The accuracy of a registry census is not so easy to determine.

16. Summary and Conclusion

The idea behind the census is simple: you try to count everybody in the population, once and only once, at their place of residence rather than somewhere else. The US Bureau of the Census does this sort of thing about as well as it can be done. Of course, the details are complicated, the expense is huge, compromises must be made, and mistakes are inevitable. The idea behind adjustment is to supplement imperfect data collection in the census with imperfect data collection in a post enumeration survey, and with modeling. It turns out, however, that the imperfections in the adjustment process are substantial, relative to the imperfections in the census. Moreover, the arguments for adjustment turn out to based on hopeful assumptions rather than on data.

The lesson extends beyond the census context. Models look objective and scientific. If they are complicated, they appear to take into account many factors of interest. Furthermore, complexity is by itself a good first line of defense against criticism. Finally, modelers can try to buttress their results with another layer of models, designed to show that outcomes are insensitive to assumptions, or that different approaches lead to similar findings. Modeling has considerable appeal. Technique is seductive, and seems to offer badly-needed answers. However, conclusions may be driven by assumptions rather than data. Indeed, that is likely to be so. Otherwise, a model with unsupported assumptions would hardly be needed in the first place.

Affiliations

David A. Freedman is Professor of Statistics and of Mathematics, University of California, Berkeley, CA 94720-3860.

Kenneth W. Wachter is Professor of Demography and of Statistics, University of California, Berkeley, CA 94720-2120.

The authors testified against adjustment in Cuomo v Baldrige (1980 census) and New York v Department of Commerce (1990 census). They have consulted for the Department of Commerce on Census 2000.

Bibliography

Anderson, M. and S. E. Fienberg (1999). *Who Counts? The Politics of Census-Taking in Contemporary America*. New York: Russell Sage Foundation.

Brown, L. D., M. L. Eaton, D. A. Freedman, S. P. Klein, R. A. Olshen, K. W. Wachter, M. T. Wells, and D. Ylvisaker (1999). "Statistical Controversies in Census 2000," *Jurimetrics* 39: 347–75.

Citro, C. F., D. L. Cork, and J. L. Norwood, eds. (2001). *The 2000 Census: Interim Assessment*. Washington, D. C.: National Academy Press.

Citro, C. F,, D. L. Cork, and J. L. Norwood, eds. (2004). *The 2000 Census: Counting under Adversity*. Washington, D. C.: National Academy Press.

Cohen, M. L., A. A. White, and K. F. Rust, eds. (1999). *Measuring a Changing Nation: Modern Methods for the 2000 Census*. Washington, D. C.: National Academy Press.

Cork, D. L., M. L. Cohen, and B. F. King, eds. (2004). *Reengineering the 2010 Census: Risks and Challenges*. Washington, D. C.: National Academy Press.

Darga, K. (2000). *Fixing the Census until It Breaks*. Lansing: Michigan Information Center.

Freedman, D. A., P. B. Stark, and K. W. Wachter (2001). "A Probability Model for Census Adjustment," *Mathematical Population Studies* 9: 165–180.

Freedman, D. A. and K. W. Wachter (1994). "Heterogeneity and Census Adjustment for the Intercensal Base," *Statistical Science* 9: 458–537 (with discussion).
http://www.stat.berkeley.edu/ census/612.pdf

Freedman, D. A. and K. W. Wachter (2003). "On the Likelihood of Improving the Accuracy of the Census Through Statistical Adjustment" *Science and Statistics: A Festscrift for Terry Speed*. Darlene R. Goldstein, Editor. IMS Monograph 40, pp. 197–230.
http://www.stat.berkeley.edu/users/census/612.pdf

Prewitt, K. (2000). "Accuracy and Coverage Evaluation: Statement on the Feasibility of Using Statistical Methods to Improve the Accuracy of Census 2000," *Federal Register* 65: pp. 38373–38398.

Redfern, P. (2004). An alternative view of the 2001 census and future census taking *JRSS Ser. A* 167: 209–248 (with discussion).

Skerry, P. (2000). *Counting on the Census? Race, Group Identity, and the Evasion of Politics*. Washington, D. C.: Brookings.

Stark, P. B. (2001). "Review of *Who Counts?*" *J. Economic Literature* 39: 592–5.

US Census Bureau (2001a). *Report of the Executive Steering Committee for Accuracy and Coverage Evaluation Policy*. With supporting documentation, Reports B1–24. Washington, D. C.
http://www.census.gov/dmd/www/EscapRep.html

US Census Bureau (2001b). *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.
http://www.census.gov/dmd/www/EscapRep2.html

US Census Bureau (2003). *Technical Assessment of A.C.E. Revision II*.
http://www.census.gov/dmd/www/ace2.html

US Department of Commerce (1991). 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, Washington, D.C. Reprinted in part in *Federal Register* 56: 33582–33642 (July 22).

Wachter, K. W. and D. A. Freedman (2000). "The Fifth Cell," *Evaluation Review* 24: 191–211.

Ylvisaker, D. (2001). "Review of *Who Counts?*" *Journal of the American Statistical Association* 96: 340–41.

Zaslavsky, A. M. (1993). "Combining Census, Dual System, and Evaluation Study Data to Estimate Population Shares," *Journal of the American Statistical Association* 88: 1092–1105.

April 2004
Technical Report No. 652
Statistics Department
University of California
Berkeley, CA 94720-3860
www.stat.berkeley.edu/~census/652.pdf