Counting and Estimation: Methodology for Improving the Quality of Censuses

The U.S. 2000 Census Adjustment Decision*

Margo Anderson
Department of History
University of Wisconsin-Milwaukee
Milwaukee, WI 53201, USA
margo@csd.uwm.edu

Stephen E. Fienberg
Department of Statistics &
Center for Automated Learning and Discovery
Carnegie Mellon University
Pittsburgh, PA 15213-3890, USA
fienberg@stat.cmu.edu

July 31, 2001

*Paper presented at the International Conference on Quality in Official Statistics, Stockholm, Sweden, May 14-15, 2001.

Abstract

In most countries census data are recognized as being fallible, i.e., subject to various errors of undercounting and erroneous enumeration. Part of the difficulty faced by statistical agencies is that the public tends to think of the census as an exercise in counting. When the census process is more properly viewed as one of estimation then the statistical issues come to the fore, including bias-variance tradeoffs. This paper reconsiders the methodology used in the 1990 and 2000 U.S. decennial censuses to correct for census error through adjustment, namely the post-enumeration survey. The 2000 post-enumeration Accuracy and Coverage Evaluation (ACE) survey re-measured the census results for over 300,000 households in aprroximately 11,000 blocks nationwide. In March of this year, the U.S. Bureau of the Census made a decision not to adjust the 2000 census using the results of ACE. We review some of the quality considerations surrounding that decision, implications for censuses in other countries, and some related estimation problems growing out of the use of multiple lists.

1 Introduction

In the 1990 U.S. census of population and housing, roughly 1 in 10 residents of the U.S. were not properly counted, with omissions exceeding erroneous enumerations by about 5 million, for a net undercount of just under 2%. The gross error of approximately 25 million persons (the sum of the omissions and erroneous enumerations) made it clear to most statistical observers that it was time to think about the census process in terms of estimation instead of simply counting (Ericksen and DeFonso, 1993). The documentation of these errors came from a post-enumeration survey of 165,000 households in over 5,000 blocks nationwide. Had the two types of errors been spread uniformly across the parts of the country and among groups there would have been little fuss. But when it became clear that the two types of errors occurred in a differential fashion, especially across racial groups, the political decision not to adjust the census to correct for this differential undercount led to a protracted political and legal dispute. The dispute ended in a 1996 Supreme Court decision which allowed the political decision to use unadjusted data to stand because it could not be conclusively demonstrated to be "arbitrary and capricious." In Anderson and Fienberg (2001a), we provide a detailed account of these events and their technical statistical basis.

The political controversy over adjustment continued throughout the 1990s as the Census Bureau proposed an ambitious plan to incorporate sampling for non-response followup in addition to a 750,000 household post-enumeration survey. Democrats generally favored sampling and adjustment while Republicans opposed them. This controversy led to a legal challenge that also resulted in a Supreme Court decision, which this time declared that sampling could not be used for congressional apportionment.[1] This decision in effect ruled out the plan to use sampling for non-response followup. The Bureau then redesigned its post-enumeration survey and relabeled it as the Accuracy and Coverage Evaluation Survey (ACE). ACE involved over 300,000 households in about 11,000 blocks or block-clusters across the country, and the Bureau planned to produce sample-adjusted counts based on ACE for all purposes other than congressional apportionment. In June 2000, then Census Bureau Director Kenneth Prewitt (2000) issued a report declaring the use of adjustment as operationally and statistically feasible and setting the Bureau on the course of adjustment unless there were problems with ACE and the raw census enumeration counts could be shown to be superior. A senior level committee, the Executive Steering Committee for ACE Policy (ESCAP) was formed within the bureau to evaluate whether the data produced in census 2000 supported the initial determination to proceed with adjustment.

The Census Bureau now has the results of the 2000 census and the ACE post-enumeration survey in hand, although its analyses of them are ongoing. The census appears to have had a high level of gross error, although not as great as in 1990, for a net undercount of about 1.2%. In March of this year, the ESCAP recommended to the Secretary of Commerce in the new Republican administration "that unadjusted data be released as the Census Bureau's official redistricting data." The basis for this recommendation was the result of the committee's inability "to conclude that the adjusted data are more accurate for use in redistricting." (Bureau of the Census, 2001). The full report is available online as are technical reports providing backup details.

In the next section we briefly describe the dual systems estimation (DSE) methodology used to adjust census counts based on the ACE results, some anomalies that arise in estimation practice, and departures from DSE assumptions of special interest to the Census Bureau. Then, in Section 3, we

[1]For the full text of this decision, see Department of Commerce v House of Representatives. 1999. 515 U.S. 316.

review the Census Bureau "concerns" about ACE and the accuracy of the adjusted counts raised by the ESCAP Report and other critics. Overall we conclude that the evidence available at the time of the ESCAP review and recommendation favored the use of the adjusted data. For a related set of conclusions with a complementary focus to the present paper, see Mulry and Spencer (2001b). In Section 4, we revisit the issue of errors in the census and reexamine the various estimates for omissions and erroneous enumerations and their sum, gross error. We also discuss some other issues of concern arising in the 2000 census and we discuss alternative approaches to DSE under consideration, including those utilizing additional lists and/or samples to assess dependence and heterogeneity.

In Section 5, we briefly turn some broader issues that arise in the debate about adjustment, and the rhetoric over counting versus estimation. In particular we contrast the U.S. experience on these issues with those in Australia and Canada. We conclude the paper with some observations on the implications of the Census Bureau's decision not to adjust the 2000 census results.

2 ACE Methodology for Census 2000

2.1 Background

Following the completion of the original enumeration, in June 2000, the Census Bureau implemented the Accuracy and Coverage Evaluation (ACE) Survey to gather information from the inhabitants of approximately 11,000 blocks or block clusters across the nation, matching the information collected with the results of the original enumeration for these blocks. In all, the final design for the PES involved checking on the occupants of 314,000 households nationwide. The process began before the actual enumeration with the selection of the sample blocks, and utilized improved methods including telephone interviews with those households that returned their census forms early in the mail-out mail back phase in April. This section draws heavily on Chapter 5 of Anderson and Fienberg (2001a), and the similarities between ACE and the 1990 post-enumeration survey.

In effect, the same sample of blocks produces two samples: the P-sample, consisting of all people living in the sample blocks at the time of the ACE interview, and the E-sample, consisting of all census enumerations coded to the sample blocks. These two samples were to exclude people living in group quarters, and the homeless. The key step for the dual systems estimation approach was to be the matching of persons between the two samples and the placement of people into the 3 observable cells of the 2×2 table cross-classifying census enumeration status with ACE status. Exact matching of names, addresses, etc., was not to be a requirement. Someone who is enumerated in a sample block was to be treated as correctly enumerated if he or she was counted anywhere in the block containing the address or in a ring of adjacent blocks, known as a "search area." This search was done on a targeted and sample basis (see the discussion of this feature below) and was used to reduce the numbers of omissions and erroneous enumerations.

The design of ACE allows for the use of information from the E-sample to measure the proportion of erroneous census enumerations. Erroneous enumerations consist of either duplications, fictitious enumerations, people born after census day or who died before census day, people counted in the wrong location, and census enumerations with insufficient information to allow both matching and follow-up reinterviews. Movers pose special problems for this determination. People who moved from outside the search area into the sample block after census day and are nonetheless counted there as part of the enumeration are deemed to be erroneous enumerations. If they moved from one address in the search area to another the matching process deems them to be correctly enumerated as long as they are counted only once.

The details of this process of matching, evaluation of status, and field followup were different in 2000 compared with 1990 and their implementation were essential parts of the evaluation of the success of the ACE adjustment process (see Davis, 2001, for a brief description). This makes comparisons between 1990 and 2000 especially difficult. For example, responding to criticisms of the imputation process used to resolve household information late in the fall of 1980 (e.g., see Citro and Cohen, 1985), the Bureau consciously kept whole household imputations to a minimum in 1990 and instead used other untested "last resort" coverage improvement procedures (e.g., see Ericksen and DeFonso, 1993, and Edmonston and Schultze, 1995). The net result was low numbers of imputations. Changes in the field followup procedures in 2000 precluded the use of many of these failed coverage improvement procedures and led to a remarkable rise in the number of imputations to 5.7 million persons, nearly three times the number in 1990 (U.S. Census Monitoring Board, Presidential Members, 2001). Imputations are removed from the counts in calculating adjustment factors, as described below, thus exacerbating the comparisons between 1990 and 2000.

2.2 Technical Details

With information on erroneous enumerations, matches, imputations, etc., in hand, the Bureau statisticians applied the dual systems estimation approach to each of the 448 poststrata, defined by combinations of

- age,
- gender,
- race,
- Hispanic status,
- housing tenure (owner/renter), and
- geography.

Following completion of ACE, the number of poststrata was reduced from 448 to 416 through collapsing to reduce variances.

For stratum s, the dual systems estimate for the population total is:

$$\hat{\mathrm{N}}_s = \frac{\mathrm{NCE}_s \mathrm{NP}_s}{\mathrm{M}_s}. \tag{1}$$

Here NP_s is the weighted P-sample total for stratum s, M_s is the number of P-sample matches for stratum s, and NCE_s is the census count for stratum s corrected for imputation and erroneous enumerations, i.e.,

$$\mathrm{NCE}_s = (\mathrm{C}_s - \mathrm{II}_s - \mathrm{LA}_s)\left(1 - \frac{EE_s}{\mathrm{NE}_s}\right), \tag{2}$$

where II_s is the number of whole people imputed to match status for stratum s, LA_s is the number of late census additions in stratun s, too late for inclusion in matching and the E-sample, EE_s is the number of E-sample erroneous enumerations for that stratum s, C_s is is the final census household population count for stratum s, and NE_s is the estimated weighted number of people for stratum s. As we noted above, the numbers of imputations in 2000 were about three times those for 1990.

Statisticians familiar with capture-recapture methodology usually think of this result as being described in terms of data organized in the form of a 2×2 contingency table of the form:

Table 1: 2×2 Table Setup for Dual Systems Estimation.

		ACE Survey		
		In	Out	Total
Census	In	M_s Matches	$\mathrm{NCE}_s - \mathrm{M}_s$ Not in Sample	NCE_s Correct Enumeration
	Out	$\mathrm{NP}_s - \mathrm{M}_s$ Not in Census	?? To Be Estimated	
	Total	NP_s In Survey		N_s Population Total

Then formula (1) gives the usual capture-recature population estimate. As we noted above, ACE also allows the bureau to decide if people counted in the census are actually fictitious, duplicates, or some other type of erroneous enumeration. These erroneous counts have to be removed from the E-sample total before the DSE is calculated, because it is theoretically impossible for such people to be matched. If they remained in the E-sample total, they would inflate the number in the non-match (1,2) cell. The calculations of standard errors for DSE population estimates in the census context treat the removal of EEs not as a statistical process (with its own error component) but as a judgment. Assessing biases and added components of variation due to the error in the EEs is deferred to loss function analyses, some of which we outline below.

There is an alternative representation for the estimate in equation (1). If we let

$$\mathrm{DD}_s = \mathrm{C}_s - \mathrm{II}_s - \mathrm{LA}_s \tag{3}$$

be the number of "data-defined" people in the census in stratum s and CE_s be the number of people correctly enumerated in stratum s, then we can rewrite the estimate (1) as

$$\hat{N}_s = \mathrm{DD}_s \times \frac{\mathrm{NP}_s}{\mathrm{M}_s} \times \frac{\mathrm{CE}_s}{\mathrm{NE}_s} \tag{4}$$

(see Davis, 2001, formula (1)).

The version of the DSE formula in (4) is important in that this is how the census Bureau actually does the estimation, i.e., it estimates each of the components of the right hand side of equation (4) separately.

At any rate, this dual systems approach yields "adjustment factors" $\{\mathrm{AF}_s\}$, that are the ratio of the estimated true population for the strata divided by the actual census count for the strata, $\mathrm{C}_s = \mathrm{DD}_s + \mathrm{II}_s + \mathrm{LA}_s$, i.e.,

$$\begin{aligned} \mathrm{AF}_s &= \frac{\hat{\mathrm{N}}_s}{\mathrm{C}_s} \\ &= \frac{\mathrm{DD}_s}{\mathrm{C}_s} \times \frac{\mathrm{NP}_s}{\mathrm{M}_s} \times \frac{\mathrm{CE}_s}{\mathrm{NE}_s}. \end{aligned} \tag{5}$$

The Bureau then applied these adjustment factors to the original raw enumeration counts, block by block, for each of the approximately 8.26 million blocks in the country. Adjusted block population totals come from adding up the corrections for each of the post-strata in the block, and fractional people are avoided by the use of controlled rounding. This adjustment process excluded those living in group quarters, and two other groups of individuals: late census additions and imputations (the latter numbering approximately 5.7 million), and a special subset of suspected duplicates. We discuss some of the implications of these "individuals" in our discussion of gross error in a later section.

If there is a sufficiently large proportion of erroneous enumerations in a post stratum, and more particularly, if the proportion of correct enumerations is less than the P-sample match rate, the adjustment factor ultimately produced by DSE will be less than 1.0. For example, if in a poststratum of 200 people in the E sample, 100 people were determined to be erroneous enumerations, the correct enumeration rate for the post stratum would be 50%. Even if the match rate on the P-sample was perfect, that is 100%, the poststratum adjustment factor would be 0.5 because half of the people originally counted in the poststratum would be erroneous.

For the 1990 census, the Bureau used 1392 poststrata and a special smoothing procedure to shrink extreme adjustments. In 1992, following the discovery of a computer error and additional analyses, the Bureau announced a new set of adjusted counts based on 357 poststrata; these appeared to be sufficient in size to eliminate the need for "smoothing" which was a target for critics of adjustment (e.g., see Brown et al., 1999). For census 2000, the

Bureau continued to shy away from the use of statistical models in the the definition and manipulation of poststrata. Because the sample size of ACE was almost double that of the 1990 post-enumeration survey, the Bureau could afford to increase the number of poststrata somewhat, but in the partisan environment of the late 1990s, Bureau officials felt that they could not defend a position of flexibility in defining post strata after the census results were known. They believed, not unrealistically, that Republicans in Congress would perceive such a stance as partisan manipulation by Democrats. Given this exact prespecification of poststrata, the Bureau was unable to be guided by actual census results as it was in 1992. We return to the issue of the poststrata and their influence on the adjustment results below.

2.3 ACE and DSE in Practice

There are several interesting features of the empirical implementation of the DSE methodology in ACE:

- The mechanics of the estimation process for DSE and the use of sample weighting can produce anomalous results, such as poststrata in which the (1,2) of Table 1 contains a negative value. For example, in 1990 37% of the non-Black 2×2 tables and 26% of the Black 2×2 tables had such negative values (Bell, 1993). The simple "fix" for this phenomenon is to set the values for these cells equal to zero, since the modification of the table does not affect the DSE estimate (4) and hence does not affect the adjustment factor (5). (None of the rates in formula (5) is affected by whether matches exceed the census count reduced by estimated erroneous enumerations, the situation that produces a "negative" (1,2) cell.)

- Most non-technical people describe DSE as producing large numbers of estimated people. But for most post-strata, the number of those added because they were in the survey but not in the census (cell (2,1) in Table 1) was much greater than those estimated as being missed by both the ACE and the census ((cell (2,2) in Table 1). Thus the vast majority of those added as part of the ACE methodology corresponded to people actually reported in the ACE survey.

- A substantial proportion of the adjustment factors in both 1990 and 2000 were less than 1. This is because the correction for erroneous enumerations brought the number of census enumerations down substantially, so much so that it exceeded the sum of those added by the survey and estimated as missed in both. These adjustment factors less than 1, when applied at the block level can in fact produce adjusted block totals less than those of the original count. In 2000, when the adjusted data were aggregated from the 417 poststrata to 64 poststrata groups, 10 of the 40 non-minority groups had adjustment factors less than 1 and only 1 of the 24 minority groups did (non-Hispanic Black owners, in large SMSAs where there was a low mail-out mail back return rate). (Davis, 2001)

ACE and dual systems estimation are designed to correct for census error. In the absence of serious measurement error, ACE methodology should improve the raw count. There are several sources for measurement error in ACE and each contributes to both the bias and the variability of ACE adjusted counts:

- *Matching Error.* Dual systems assumes that those in the P-sample can be matched without error to those in the E-sample. In fact, the computerized matching approaches used in 1990 led to the imputation of match status for a substantial fraction of the cases and concern was expressed regarding the implications of this problem. Evidence from ACE evaluations suggests that the matching error rate for 2000 was substantially lower than in 1990.

- *Correlation Bias.* Correlation bias is a much abused term used variously to describe both dependencies between the census and the sample and heterogeneity in response rates. In fact, as Bell (2001) and Darroch, et al. (1993) point out, in the dual system setting the two components are not really distinguishable. (This should be contrasted with Wachter and Freedman, 2001, and Freedman, Stark, and Wachter, 2001, who attempt to do so through some questionable assumptions.) In 1990, it was widely believed that correlation bias was positive and especially strong from Black males. But the net result of correlation bias is the understatement of adjustment, and thus ACE would move the raw counts in the right direction, but not far enough (see Kadane, Meyer, and Tukey, 1999, for a explanation of this phenomenon and Fienberg, Johnson, and Junker, 1999, for related results). In 1990, for a moderate range of values explored, the greater the correlation bias that was assumed the more the adjustment improved upon the unadjusted raw census counts (e.g., see Mulry and Spencer, 1993, and the related discussion in Anderson, et al., 2000).

  - *Dependence.* Dual systems estimation assumes that the responses to the census and the ACE are independent. The ACE survey operations were structurally separate from those of the census but this still does not mean that the the behavioral responses are independent of each other. Intuitively, we expect to see a positive correlation and the dual systems estimate would therefore be biased downward.
  - *Heterogeneity.* Capture probabilities clearly vary among individuals within poststrata and the net effect of this variation (under relatively weak assumptions) is that it induces a positive dependence between the responses to the census and the sample in the aggregate.

- *Missing Data.* The level and pattern of missing data in ACE was determined by the Bureau to be comparable to that in 1990 and not a major problem for the accuracy of adjustment. Missing data contribute largely to the variances of dual system estimates.

- *Balancing Error.* Differences in search strategies for EEs and matches lead to balancing error. We discuss this issue below.

- *Synthetic Error.* The synthetic assumption implies that net census coverage rates do not vary within poststrata. We discuss this issue below.

3 ACE Results and The ESCAP Report

The 2000 ACE included a variety of design improvements that resulted in better quality data. According to the ESCAP Report and Hogan (2001), ACE was carried out as designed, with only minor modifications, and it was an operational success. The net national undercount was reported as 1.2%, with the differential undercount for various minority groups dropping to the 2-3% range from 4-6% in 1990, with the largest drop occurring for non-Hispanic Blacks. What is especially revealing is the shift from 1990 to 2000 in the distribution of the net undercount from minorities to non-minorities and especially from those who own their own households to those that do not. The undercount for males age 18-29 and for both males and females age 30-49 increased, and this was offset by a major drop in the undercount for those age 0-17. These and many other detailed results from ACE, including those for a special set of 64 evaluation poststrata, are reported in Davis (2001).

Virtually all sources of error in the ACE survey and the adjusted estimates produced from it appear to be at the same level as or lower than was the case in 1990, e.g., matching error seemed to be under greater control than was the case in 1990 and sampling error was much reduced as a consequence of the increase in sample size. Nonetheless, the ESCAP report raises several cautions regarding the ACE results and identified four specific concerns that prevented it from recommending the use of ACE-adjusted counts. We consider each in a separate subsection.

3.1 Inconsistency Between ACE and Demographic Analysis

Going back to the 1950s the Census Bureau has produced Demographic Analysis (DA) estimates for the national population based on the accounting identity:

$$\text{Population} = \text{Births} - \text{Deaths} + \text{Immigration} - \text{Emigration}, \qquad (6)$$

and for decades these served as a standard of comparison with the national census count before the full implementation of dual systems methodology. But beginning with the 1980 census, the Bureau needed to confront the inadequacies of its estimates for emigration and immigration, especially for undocumented aliens. These numbers have proven to be especially problematic for those of Hispanic origin. DA had traditionally been used to measure the differential undercount of Blacks versus non-Blacks, but provides no accurate direct measures of the differential undercount for other groups such as Hispanics.

Statisticians and demographers outside the Bureau have expressed growing concerns about the lack of measure of bias and variability associated with its estimates of population totals (see e.g., Anderson, et al., 2000), particularly since disparities between dual systems and DA estimates have often been used to argue against the use of adjustment (e.g., by Darga, 1999), and Brown, et al., (1999) have gone even further and suggested that the DA differentials be carried down to the block level, in place of the dual systems methodology described above.

Reports from the Bureau in early February 2001 showed the following estimates of the total population of the U.S.:

Demographic Analysis	279.6 million
Census 2000	281.4 million
ACE	284.7 million

Taken at face value, these numbers suggest that the census overcounted the population by almost 2 million and that ACE would have added another 3.3 million people. Unfortunately, this is where critics of adjustment like Freedman and Wachter (2001) stop in their analysis. Even the Census Bureau went further, noting considerable agreement between ACE and DA for some population subgroups, and identifying the major problem associated with the immigration values used. The ESCAP Report provides an alternative DA total of 282.3 million which at least moves DA in the correct direction relative to both the unadjusted and ACE adjusted census estimates.

Passel (2001) has used data from the April Current Population survey to derive "quasi-DA" estimates and provided an alternative analysis of the DA results. He reached the following conclusions:

- For the non-Hispanic Black population, ACE and DA are in very close agreement, with both totals exceeding the census count by almost 1 million. A comparison of sex ratios further indicates that ACE undercounts black males (attributable to correlation bias, as it was in 1990; see also Bell, 2001).

- For the non-Black, non-Hispanic population, ACE and DA are also in very close agreement, exceeding the census by 1.5 million and 1 million people, respectively. Thus, here again, DA shows a net undercount in the census.

- For the Hispanic population, ACE and "quasi-DA" estimates diverge, with ACE exceeding the DA estimate by 2.4 million, and the census by only 1 million people. Passel goes on to argue that adjustments to the DA for known sources of underestimation in immigration could reduce the discrepancies between ACE and the DA even further.

The Bureau in its ESCAP Report concluded that "[t]he inconsistency between the demographic analysis estimates and the ACE estimates raises the possibility of an as-yet undiscovered problem in the ACE or census methodology." We disagree. Given the known shortcomings of DA, and the Passel (2001) reconciliation, we find little reason to be concerned with ACE estimates, except for their likely undercount of Black males.

3.2 Balancing Error

Balancing error occurs if the handling of cases differs in the P- and E-samples, i.e., the search area and effort to detect omissions should be the same as to assess erroneous enumerations (EEs). In 1990, the Bureau searched for matches to the P-sample and erroneous enumerations in the E-sample in surrounding blocks in the hopes of reducing the EE rate and increasing the match rate and thus reducing the variances of the dual systems estimates. In 2000, the Bureau altered its search strategy to involve a single ring of blocks on a sample basis. This was known as the Targeted Extended Search (TES), which did what its name suggests, it searched for matches of housing units in the ring of blocks surrounding a sample block-cluster, but only in a targeted form. There was a pre-census housing unit match for the ACE blocks and only those blocks with unit nonmatches were included in the TES. Blocks with a nonmatch rate of at least 45% were included in the sample with certainty, and blocks with errors but at a rate of less than 45% were included with lower probabilities. Again the hope was one of reducing the variances of the ACE estimates, without introducing serious bias. The TES thus produces another set of sample weights which need to be combined with those from the ACE sample design to produce correct estimates of errors, etc.

What evidence do we have that balancing error is a problem for the ACE? The TES strategy in 2000 was highly successful, and produced important reductions in variances (Navarro and Olson, 2001). But the evaluation of TES also found that the rate at which matches from the P-sample were found in the surrounding blocks exceeded that rate at which EEs from the E-sample were found there by just under 1%, which corresponds to approximately 2.9 million people (with a standard error of 734 thousand) when weighted up to the population. There was a similar level of "imbalance" in 1990 but then the Bureau attributed it to the treatment of "movers" and a small number of blocks which had high leverage on the results. The changes in the treatment of movers and greater care in some other features of the operation suggested that there must be a different explanation in 2000. The most reasonable explanation relates the "fuzzy" boundaries for the blocks, the understatement of E-sample geocoding errors, and possible errors in the blocks that were not included in the TES. The Bureau is doing an additional field followup as we write to explore these issues and get revised estimates of the possible level of balancing error. Nonetheless, the best current number still show that balancing error was lower in 2000 than it was in 1990.

3.3 Uncertainty About Effects of Synthetic Error

Synthetic error occurs as a result of applying the poststrata-level adjustment rates computed from a sample of blocks to all of the other blocks in the nation. The extent to which blocks within poststrata are heterogeneous induces synthetic error. Clearly blocks within poststrata do differ in terms of what their "true" adjustment factors and thus there is synthetic error in ACE. The problem is that we can't measure the extent of synthetic error directly.

The Census Bureau used four artificial populations to "simulate" the effects of synthetic error associated with applying the ACE adjustment factor to every block in the nation. The artificial populations used data from the following surrogate variables in different ways to get block-level surrogates for omissions and erroneous enumerations.

- Number of allocations–Households (HHs) with more than a specific amount of item nonresponse for race, Hispanic origin, relationship to head of HH, sex, and age.
- Number of non-mail returns.
- Number of substitutions–whole-HH imputations and/or partial-HH substitutions.
- Number of duplicates added back (late adds).
- Number of units at basic street addresses.

The underlying methodology for these simulations is described in Griffin and Malec (2001) and was influenced by the surrogate variables and methodology used by Freedman and Wachter (1994). Related results and methodology used to study synthetic error in 1990 can be found in Fay and Thompson

(1993), Kim, Zazlavsky, and Blodgett (1995), and Wachter and Freedman (2000). The surrogate populations used by Griffin and Malec turn out to have correlations with EEs and omissions on the order of 0.25 (with R^2s of about 0.08). Thus we worry that the proxies constructed in these analyses may not behave enough like the omissions and EEs in terms of heterogeneity to be truly useful (see the related discussions in Belin and Rolph, 1994, and Ericksen, Fienberg, and Kadane, 1994).

The Bureau's loss function analyses did not include a component for synthetic error and thus the various analyses reported by Griffin and Malec (2001) attempt to assess what its effects would have been had it been included. In general, their analyses supported the superiority of the ACE-adjusted counts, but they were done using primarily assumptions of a high level of correlation error. If the effects of synthetic error scaled down proportionally at lower levels of correlation bias, then there would be no serious concern. Because these analyses had not been completed by the time the ESCAP decision was made, the committee expressed reservations about this source of error.

There are some instructive alternative ways to consider this issue. For example, Spencer (2001) develops a total error model for census 2000, and shows how the model bias decomposes the sum of the correlation bias and the synthetic estimation bias such that the sum of the synthetic bias over the whole nation is zero.

We find the evidence supporting concern about synthetic error to be weak and derivative in nature. But even if one were to take the problem seriously, the Griffith and Malec (2001) report suggests that synthetic error was no greater a problem in 2000 than it was in 1990 and that, were it included in the loss function analyses, it would not be expected to change the observed superiority of the ACE-adjusted results.

3.4 Loss Function Analyses

Loss Function Analysis is a way to take measurement errors and combine them into a number that compares accuracy between raw and ACE adjusted data. The ESCAP looked at data processing errors, data collection errors and correlation bias. In the ACE evaluation reports by Navarro and Asiala (2001) and by Mulry and Spencer (2001a), the reported loss function analyses of the 2000 ACE showed that the ACE was superior in terms of accuracy compared with the unadjusted census results.

Given that these analyses seem to overwhelmingly favor adjustment, why did the ESCAP rule against adjustment and cite reservations about the loss function analyses? It turns out that there were five "computer runs" (out of about 200) that raised questions. These runs concerned counties with populations under 100,000 people. While 100,000 people sounds like a lot, these counties comprise a very small percentage of the nation's population. In California, for example, while 25% of counties contain less than 100,000 people, only 2% of the state's population reside in those areas. Furthermore, the ESCAP reports do not tell us the effects of variability. Overall, with this one strange exception involving a small portion of the population of the country, we have yet to find any other analysis that concludes that the unadjusted census data are superior.

Counties with populations under 100,000 vary widely in population size. For example, Loving County, Texas has a population of 67 people; Clay County, West Virginia has 10,330 people; and Navajo County, Arizona has just under 100,000 (97,470). The population varies such that it is not surprising that the loss function analysis showed disparaging results. A better comparison would have been to check places with populations of 100,000. Indeed, minutes from ESCAP meetings show that the initial loss function analysis comparing places with populations of 100,000 showed the ACE adjusted counts to be more accurate than the raw census counts.

The loss function analyses conducted after the 1990 census were not without their critics. For example, Freedman, et al. (1994) and Brown, et al. (1999) attempted to cast doubt upon the accuracy of the Census Bureau's analyses and especially the approach of Mulry and Spencer (1993). Anderson, et al. (2000), however, have exposed the the sleight of hand in their algebraic manipulations. For the 2000 results, Freedman and Wachter (2001) have taken a somewhat different tack, dismissing the analyses by noting that "the apparent gains in accuracy ... tend to be concentrated in a few geographic areas" and claiming that "[a]t a deeper level, loss function analysis turns out to depend more on wishful assumptions than on data." We disagree.

The critics of census adjustment just mentioned, and others, often cite concerns with adjusted data at lower levels of geography. Neither data set is conclusively more accurate than the other at small areas of geography. But when aggregated to higher levels of geography, however, the adjusted numbers demonstrate improved accuracy over the unadjusted census counts.

The ESCAP was correct in flagging runs regarding counties under 100,000, at least because these areas differed from 1990 post-enumeration survey comparisons. However, we see no reason to conclude that the problem should be a basis for discarding the clear indications of improvement (compared to the post-enumeration survey) of loss function analysis. Our conclusion, and that of the Bureau, is that for geographic areas that matter (Congressional Districts and areas of that size), the loss function analysis evidence strongly supports the superiority of ACE-based adjustment.

4 Observations and Issues

4.1 The Census 2000 Gross Error Rate

What is the current best estimate for the gross error in the 2000 census, i.e., the sum of omissions and erroneous enumerations? Recall that, in 1990, gross error was estimated to be on the order of 10% of the population, or 25 million people (e.g., see Bryant, 1993). This number was later revised downward through corrections and the use of a different poststratification scheme developed in 1992, which reduced the number of poststrata from 1392 to 357 (see Hogan, 1993).

Table 2 reports the Census Bureau's breakdown of the erroneous enumeration rate for 2000 by type and contrasts it with the figures for 1990 (after the corrections of 1992). Several changes from 1990 to 2000 are worthy of note. The overall decline in EE rate is due primarily to the decline in duplicates and "other residence." Duplicates may be underestimated here, however, because the Bureau examined several million suspected duplicates for a separate operation in Fall 2000 and these households were not included in the ACE operations. We comment further on this below.

The ACE population weighted estimate for erroneous enumerations corresponding to the 4.7% figure in Table 2 is 12.5 million. Given a net undercount of 3.3 million this implies that the number of omissions is approximately 15.8 million, and thus the gross error total should be 28.3 million. The comparable figures from Bryant (1993) for 1990 were 16.3 million EEs and 20.3 million for a gross error total of 36.6 million.

One can reasonably argue, as does the Bureau, that we should not include those with insufficient information and the imputed EEs in these numbers (which are a subset of the estimated 5.7 million imputations). If we exclude them, then for 2000 the number of EEs goes down to 5.9 million and the

Table 2: Breakdown of Erroneous Enumeration Rates for 1990 and 2000. Source: U.S. Census Bureau and Hogan (1993).

Type	1990	2000	Estimated 2000 Persons
insufficient information	1.2%	1.8%	4.8 million
duplicates	1.6%	0.7%	1.9 million
fictitious	0.2%	0.3%	0.7 million
geocoding error	0.3%	0.2%	0.6 million
other residence	2.2%	1.0%	2.7 million
imputed	0.3%	0.6%	1.8 million
Total	5.8%	4.7%	12.5 million

number of omissions reduces to 9.2 million, with the gross error total being 15.1 million. The comparable figures for 1990 are 11.2 million EEs and 15.2 million omissions for a gross error total of 26.4 million. Thus, the number of gross errors appears to have declined from 1990 to 2000. These are the figures we list in Table 3 below.

Freedman and Wachter (2001) suggest that there were only 1 million erroneous enumerations, and 4.3 million omissions but this is incorrect. The numbers they cite is the number of *net* erroneous enumerations and the number of *net* omissions carried down to the block level as a result of the poststrata used by the Bureau. The numbers for 1990 were comparable: 0.9 million erroneous enumerations and 6.2 million omissions using the original adjusted figures, and 1.5 million erroneous enumerations and 5.5 million omissions using the adjusted figures "corrected" figures from 1992 (see Hogan, 1993).

There is yet a third version of gross errors based on calculation by the Bureau omitting "other residence" and "imputed" errors (Killion, 1997). This produces 3.1 million erroneous enumerations and an implied omission count of 6.4 million, for a gross error of 9.5 million in 2000, compared with 4.4 million EEs and 8.4 million EEs for a gross error of 12.8 million in 1990 (using the revised postratification values). Finally if one adds in to the EEs one-half of the reinstated 2.4 million "potential duplicates" that were excluded from the ACE process in the Fall of 2000 (the fraction suggested by the Bureau as possibly erroneous; see Miskura, 2000), as well as an equivalent number of implied omissions, one gets a gross error rate of 11.9 million. It is in this sense that we agree with the Bureau that the level of error in 2000 was lower than that in 1990, even though the level of gross error in 2000 was high and problematic. But, there remain issues of comparability such as those for group quarters discussed in the next subsection. And one could go even further and reconsider the implications of the 5.7 million imputations discussed earlier for the count of gross errors (e.g., see the discussion in U.S. Census Monitoring Board, Presidential Members, 2001), but we have refrained from doing so here.

We summarize all of these values on omissions and erroneous enumerations in Table 3 so that others can keep track of what quantities are being cited in other references discussing these issues. The only thing that is held constant in each block of Table 3 is net undercount, and it is combined with the relevant numbers of estimated EE's to get the estimated number of omissions.

There is enormous confusion regarding what various documents mean regarding erroneous enumerations, omissions, and gross error, and the Census Bureau has compounded matters be selectively citing different numbers in different documents. For example, the press release by Barron (2001) misleadingly states: that "initial estimates ... indicate that there were 4.3 million persons undercounted by Census 2000 and 1.0 million persons overcounted ..." These numbers simply tell us that ACE, however much it improved upon the raw census counts, was only able to correct for a small proportion of the gross errors in the census which it detects! Moreover, these numbers disregard the differential gross error at low levels of geography because the errors net out!

What can we learn from this information on gross error and the comparisons with figures from 1990? First, it is difficult to make sense of the comparisons because of differences in procedures, e.g., in connection with last census additions, imputations and group quarters. Nonetheless, we agree with comments of others that the Bureau did a valiant job in 2000, and made many improvements in both its census enumeration and ACE processes. Had they not done so, the census and ACE results would clearly have been much worse. We also note that it is not easy to reconcile this level of gross error with the even more substantially diminished levels of differential undercount for minority groups, especially at the national level. But we do not see this difficulty as an argument against adjustment. Indeed the level of gross error in 2000 is, for us, a strong argument in favor of the use of ACE-adjusted data.

Table 3: Various "Versions" of Estimated Numbers of Omissions, Erroneous Enumerations, and Gross Errors for 1990 and 2000 Censuses (in millions).

Year	Source	Omissions	EEs	Net Error	Gross Error
1990	Original PES	20.6	15.3	5.3	35.9
	Original PES minus "unmatchables"	15.5	10.2	5.3	25.7
	Original PES as applied	6.2	0.9	5.3	7.1
	Corrected PES	20.3	16.3	4.0	36.6
	Corrected PES minus "unmatchables"	15.2	11.2	4.0	26.4
	Corrected PES minus additional EEs*	8.4	4.4	4.0	12.8
	Corrected PES as applied	5.5	1.5	4.0	7.0
2000	ACE	15.8	12.5	3.3	28.3
	ACE minus "unmatchables"	9.2	5.9	3.3	15.1
	ACE minus additional EEs*	6.4	3.1	3.3	9.5
	ACE minus additional EEs* plus potential duplicates	7.6	4.3	3.3	11.9
	ACE as applied	4.3	1.0	3.3	5.3

*also omitting "other residence" and "geocoding" errors.

4.2 Group Quarters

In April and May 2001, a new concern arose over the accuracy of the block level census counts. Local officials and demographers in state agencies began to notice sharp spikes or precipitous declines in reported populations of some blocks from 1990 to 2000, particularly in rural areas where sharp population changes are more easily spotted in block level data. The problem seems to be in group quarter geocoding errors, namely that some of the population of institutions such as college dormitories, prisons, or jails are being put in the wrong census block. The group quarters population was not included in the ACE program, so these errors may or may not carry over to the non group quarters population.

In New York State, for example, as reported in the *Buffalo News* (May 22, 2001), the "Green Haven Correctional Faculty in Beekman had a population of 2,203 on Census Day. But, according to the census, the block where the prison is located had 19 people.... Meanwhile, 25 miles and four towns to the north, the rural town of Milan, population 1,895 in 1990, more than doubled to 4,459 in 2000." Similar anomalies have been reported in Wisconsin, Arizona, Connecticut, and California. (*The Associated Press State & Local Wire*, May 28, 2001, "Flagstaff sees possible university undercount"; *The Hartford Courant*, March 27, 2001, "Census May Have Misplaced College Students; Hartford Officials Say 2,500 University of Hartford Students Were Mistakenly Counted as West Hartford Residents." By Mike Swift And Dan Haar).

The Bureau has to this point refused to change any of the data released in the PL 94-171 file on the basis of the challenges to the numbers, and has proposed that local areas use the count resolution program to identify them. If errors are confirmed, they have offered local areas "certificates that could be used in government grants in aid requests to reflect the actual population." (*Buffalo News*, May 22, 2001). As of this writing, it is not clear how extensive the problem is, whether the Bureau will issue revised files correcting the errors, or whether geocoding error affected ACE as well.

Unlike in 1990, group quarters were not part the ACE adjustment in 2000. Thus the type of geocoding error described here is simply one form of census error for group quarters not corrected by ACE. Duplicates for children away at colleges and universities and omissions from group quarters are all sources of error that would swell the numbers of errors for 2000 that we reported in Table 3.

4.3 Strategic Errors in the Planning for Census 2000 and the ACE-Based Adjustment

In our review of the events leading up to and including the 2000 census, as well as the ESCAP evaluation of ACE and the adjusted data, we have identified a series of what might be considered to be strategic errors that contributed to the Bureau's decision to recommend against the use of adjusted counts.

1. For good reasons and bad, the Bureau always seems to be planning to do the last census better instead of developing methodologies that will adapt to unexpected changes in the future. The most striking illustration of this was the choice of a post-stratification scheme for census 2000, which was modeled after a "model-free" post-hoc choice made in

1992 for the 1990 census. Some statistical observers expressed concern about this strategy as the final pieces of it were put into place and suggested many more strata and the use of a logistic regression or some related model that would have adapted itself to the changes manifested in the census and ACE results. Then, instead of expressing surprise when the poststrata net undercounts were smaller than expected for minority groups, the Bureau might have been able to reconcile the high gross error rates with the low differential undercount rates.

2. We don't mean to imply that that the Bureau failed to make major changes as it developed the plans for census 2000. Many of these changes were critical to the success of the census and had they not been made the accuracy of 2000 would have been far less that for 1990. In fact the changes were so numerous and extensive that it has been difficult to make comparisons between 1990 and 2000 on key criteria such as gross error.

3. The focus at the Bureau in 1990 was on reducing the differential undercount. But after the evaluations were complete, it became apparent that erroneous enumerations and not just omissions were a major concern. Slowly over the decade, Bureau documents began to refer to both components of error and their implications for lower levels of geography where they do not readily cancel out. The failure to address publicly the level of omissions and erroneous enumerations as part of the ESCAP Report and decisions appears to have been a major oversight.

4. In 2000, the Bureau continued to take the demographic analysis estimates as a form of gospel, when critics had warned that their accuracy which has never been properly assessed, would in all likelihood deteriorate. Thus when the ACE and DA estimates for the national totals disagreed in certain details, instead of questioning the DA, the Bureau pointed its finger at ACE and suggested that it might contain an as-yet undiscovered major source of error. This reliance on DA is part of a bureau culture that has yet to embrace the census results as estimates instead of counts.

5. Time after time, outside groups such as committees and panels from the National Academies of Sciences have noted that programs introduced in the census without careful testing and evaluation inevitably add more

error than they take out. And the Bureau has in general been a zealous advocate of such careful planning through its mid-decade test and full-scale pre-census dress rehearsal. In 2000, as in the past, seemingly ad-hoc decisions introduced at the last moment have had surprisingly major impact either on the census results or our ability to evaluate them. Thus the level of imputations and late-census additions startled many observers, especially since these counts were subject to possible high levels of error *and were excluded* from the ACE matching and adjustment procedures. Similarly, a group of several million suspected duplicates were taken out of the census for special examination, and while this process may have reduced the real level of erroneous enumeration substantially, it led to yet another group of records that were not part of the ACE adjustment procedures.

6. The ESCAP and its report appear not to have taken seriously the mandate from former Bureau Director Prewitt to use ACE to adjust the raw census counts unless there were problems with ACE and the raw census enumeration counts could be shown to be superior. The Bureau properly probed the quality and accuracy of the ACE results, but whenever there was any unresolved issues, however minor or speculative, they were treated as arguments against adjustment. Thus the Bureau's implicit decision analysis could be viewed as requiring the rejection of the "null hypothesis" that the raw census counts were more accurate, as was the case in 1990 (e.g., see Fay, 1991). We think that that approach is at best backwards statistically.

7. Clearly the Bureau was under considerable time pressure to complete the analyses of its evaluation of the ACE-adjusted counts. By law the data for redistricting needed to be produced by the end of March, 2001. But the ESCAP decision was announced at the beginning of March, thus denying the Bureau of extra time to reconsider the real evidence against adjustment. We think that this early decision was in retrospect a mistake.

8. The Bureau statisticians appear to have developed a bad case of *Freedmanitis*, and many of the questionable choices it made in the planning of ACE and its subsequent evaluation can be traced to criticisms emanating from David Freedman and his colleagues, largely those in the Department of Statistics at the University of California, Berkeley (e.g.,

see Freedman et al., 1993, Freedman and Wachter, 1994, Freedman, et al. 1994, Stark, 2000, and the summary of these criticisms in Brown et al., 1999). There were others who have written detailed assessments with regard to the adjustment methodology for 1990 and 2000, but the Bureau chose not to react to them in the same fashion. We hope that the Bureau recovers from this affliction and cast its purview vis a vis non-bureau technical commentators somewhat wider as it plans ahead to 2010.

4.4 Multiple List Approaches to Census-taking and Adjustment

New methodological approaches to multiple recapture problems (involving 3 or more as opposed to only 2 lists) offer the prospect of measuring dependence and heterogeneity (departures from two key dual systems assumptions) and separating out their effects. They have not been brought to bear on decennial census data, however, except for a triple system estimation experiment carried out in conjunction with the 1988 census dress rehearsal (Zaslavsky and Wolfgang, 1993, Darroch, et al., 1993).

The Census Bureau as part of its planning for the 2010 census is sponsoring research:

1. to use multiple system estimation (MSE) to measure the bias in the DSE estimation approach as a result of departures from the assumptions of independence and homogeneity;

2. to gauge the extent to which it improves upon DSE both in terms of accuracy and precision;

3. to extend the DSE and MSE models to allow for multiple categories and possibly even continuous models for capture rates and thus allow for departures from the perfect matching assumption as well.

As part of its 2000 decennial census assessment activities, the Census Bureau has collected administrative data from six administrative record sources prepared a composite administrative list called the Statistical Administrative Records System (StARS). The Bureau plans to match this composite list against the census enumeration counts and the results of the ACE survey to produce a three-dimensional cross-classification for each of a selection of census blocks, and these three-dimensional cross-classifications could be used directly in triple systems estimation of the populations for those blocks. One research project currently underway at Carnegie Mellon University plans to extend the Bureau approach to DSE to this triple systems setting, using traditional log-linear models of Fienberg (1972) to account for dependence, and Rasch-like models developed in Darroch, Fienberg, Glonek, and Junker (1992) and in Fienberg, Johnson, and Junker (1998) for heterogeneity. To date, Asher and Fienberg (2001) have used the six administrative record sources to model heterogeneity and dependence across administrative lists for five selected counties. Other projects will be looking at these and related data using alternative methods.

5 Learning From and Implications For Other Countries

In the United States, we appear to have turned the corner and, more and more statisticians and the educated public have begun to think of census results as estimates instead of simply a headcount. The problems and statistical issues faced by the Census Bureau in the 2000 decennial census are not unique to the U.S. In some ways we appear to be ahead of other countries and in other ways we are behind.

The difficulties associated with omissions and erroneous enumerations, and undocumented aliens in particular, are not only an American phenomenon. Other countries, such as Australia (e.g., see Choi, Steel, and Skinner, 1988, and Paice and Steel, 2000) and Canada (e.g., Statistics Canada, 1991; Gambino and Dick, 2000), have directly incorporated statistical corrections into their enumeration results with very little public outcry.

Canada's example is especially instructive because it demonstrates how the technical activities of estimation link to the public and legal arenas. Statistics Canada publishes the " census population counts" (unadjusted for coverage error) about nine months after census day - February 2002 for the 2001 Census taken in May 2001. It plans to release "final population estimates" for 2001 (incorporating 2001 census figures and adjustments for net census coverage) separately, but not until September 2003. In the meantime, the population estimates carried forward from the 1996 Census (adjusted for net coverage in 1996), continue to be the "official" population estimates. A particular change that had to be made to accommodate this approach was to ammend the Federal-Provincial Fiscal Arrangements Act (an Act that is renegotiated and passed every five years, and which defines the formula for fiscal equalization payments to the provinces). This Act originally defined the population to be used in the formula as the census population in census years and the official population estimate as certified by the Chief Statistician in other years. Now the second part of that definition applies in all years.

Yet other countries, such as the United Kingdom, have plans to implement sample-based adjustment methods similar to those used in the U.S. (e.g., see Brown, Diamond, Chambers, Buckner, and Teague, 1999, and the related discussion by Chambers and Cruddas, 2000), and are currently encountering opposition from those who do not understand that census-taking is a statistical activity, and one best carried out using the full array of tools developed by the statistical profession, including sampling and statistical models and estimation methods.

We believe that thinking about censuses in terms of quality assessment and improvement will help statistical agencies in other countries move census-taking into the new millennium. Further, we think that the U.S. could benefit from a system similar to those in Australia and Canada, both of which separate the census counts from the official population figures that incorporate statistical adjustment.

6 Concluding Observations

To date, the Census Bureau has declined to release the 2000 ACE-adjusted census results at the block level, in part because it claims that the results are still being assessed and rechecked and may yet be revised (Barron, 2001). Thus we have not seen the actual numbers that would have been released in place of the block level counts even though we have seen adjustment factors and other quantities computable from ACE. Nonetheless, we would tentatively suggest a number of issues that have emerged about censuses with sample surveys designed for adjusting local area data.

1. We know that adjusting the April enumeration counts for erroneous enumeration and undercoverage can shift seats in Congress (e.g., see Schirm, 1991, and Anderson and Fienberg, 2000). It did so in 1990,

and it is likely to do so in future censuses. The Supreme Court has ruled that current statute prevents the use of such adjusted data, but it is likely that Congress will face the issue at some point in the next several decades.

2. We suspect that adjusting the April enumeration counts for erroneous enumeration and undercoverage before the states and local legislatures redraw district boundaries, would result in politically salient changes in legislative districts around the country. But we also know that the districting process is a sufficiently large 'variable' on its own to make prediction of a particular partisan, regional, racial, or socio-economic impact of census adjustment highly problematic.

3. We know that the experience of conducting post-enumeration surveys with dual systems estimation has produced a new understanding of the importance of erroneous enumerations for the evaluation of accuracy in the census. Data of increasing sophistication on net undercoverage by age, sex and race has been available for the U.S. decennial census since the late 1940s. Data on the character and import of over-enumeration, or erroneous enumeration (EEs), and its impact on the uses of census data had not been measured systematically before 1990, nor we suspect, do we know the basic scope of the impact of EEs on accuracy yet for 2000.

4. Much of the public criticism of adjustment leveled by politicians and other non-technical experts has argued that the PES in 1990 and ACE in 2000 did not improve block level data (e.g., see U.S. Census Monitoring Board, Congressional Members,1999, 2001; see also the discussion of the latter report in the appendix). While there is no evidence to show that that the adjusted block counts are any less accurate that the unadjusted ones, it is important to remember that the PES and ACE block level data are "building blocks" for larger geographical units of interest, and it is for these aggregates that adjustment provides demonstrable improvement over raw census counts. Nonetheless, the Bureau would do well to rethink the way it organizes and releases census data, and perhaps move to higher geographic units, e.g., census tracts, for future censuses. This would not only ameliorate the concerns about accuracy at the lowest levels of geography, both for census counts and

corrections to them via adjustment methodology, but it would also address ongoing concerns about confidentiality of census data (see the related discussion in Anderson and Fienberg, 2001b).

5. We know that the controversies about accuracy of the census have been driven by both political and methodological issues, including whether there is a problem with accuracy, whether it can be fixed, and whether it should be fixed. We also know that while the accuracy debate has raged, other census innovations and uses have changed the playing field, so to speak. OMB and the Congress, for example, have taken an interest in race classifications used in the census, which in turn affects the data file produced by the Bureau for redistricting. Congress, to our knowledge, has not systematically evaluated what level of geographic or demographic detail would be sufficient to produce the best redistricting file, nor has the Bureau inquired whether 8.26 million blocks, and 126 racial and ethnic categories are appropriate across the entire nation.

We don't expect the controversy about census accuracy and sample-based adjustment to disappear, either as a result of the March 2001 decision by the Bureau to recommend the use of unadjusted data, or by the fall 2001 forthcoming decision on which data (adjusted or unadjusted) to use for intercensal estimation. Indeed lawsuits underway and the political ramifications of adjustment for redistricting will surely keep the issue alive politically.

As we look ahead to the ongoing discussions of census errors, we expect them (1) to make the conceptual breakthrough of recognizing the difference between "counting" and "estimation," and (2) to be refocused in terms of the impact of differential gross error instead of differential net error. We also hope that the methodological debates that will ensue will address the uses of the census in more focused fashion, e.g., as in Mulry and Spencer (2001b).

Appendix A: The Census Monitoring Board

In October, 1997, Congress set up a eight-person Census Monitoring Board (CMB) which was to provide oversight for activities associated with the 2000 census. The CMB has four members appointed by the then-Democratic president and four appointed by the then-Republican Congress. Over the past two and one-half years, the Board has issued a series of reports on the issue of adjustment, typically separately as the presidentially appointed members (CMBP) or the congressionally appointed members (CMBC). The former tend to support adjustment and the latter tend to oppose it. Thus the CMBP, in its semi-annual report to Congress issued on April 11, 2001, discussed the gross error rates measured by ACE and attempted to reconstruct state level adjusted totals (figures that had not been released by the Bureau) and it included a version of Passel (2001).

In its June 7, 2001 report to Congress, the CMBC drew the following broad conclusions with regard to ACE and the ACE adjusted figures. Our response to each of the conclusions is included in italics:

- Despite the diverse racial and ethnic makeup of communities in the United States, the methodology assumes certain similarities in behavior within demographic groups, while ignoring critical sociological distinctions such as income, literacy and educational levels.

  It is widely recognized by now that socio-demographic factors explain much of the variation in omissions and erroneous enumerations captured by the ACE process. The census short form, however, measures only age, race, Hispanic status, sex, and housing tenure. Therefore these are the only variables that can be used to capture the socio-demographic effects. Those who criticize ACE in this regard do not take the next step and propose changing the short form questionnaire to improve the measurement of census error.

- When counts differ between the actual census data and the A.C.E. survey results, the methodology assumes the A.C.E. counts are always correct.

  This statement is demonstrably false and echoes mistaken interpretations of the PES process by those without a careful understanding of DSE methodology (e.g., Chief Justice Rehnquist of the U.S. Supreme Court, see Anderson and Fienberg, 2001a, pp. 163-164) as well as *more technically-minded critics (e.g., see Stark, 2000). First we note that ACE consists of several components, beginning with the 310,000 person survey. The ACE methodology then independently checks the census counts for the ACE blocks for erroneous enumerations. These numbers are correctly subtracted out from the census counts because they are errors. At that point, the DSE methodology includes people counted in the census but missed in the ACE survey ($NCE_s - M_s$), and estimates those missed by both the survey and the census from formula (1) above).*

We also note that the labels in Figure 3 of the CMBC report, depicting the cross-classification of the ACE and census results, are similarly incorrect (c.f., our tabular representation in Table 1).

- Using adjustment factors to address undercounts, millions of people, virtual people who were not contacted during the actual census, would be added or hundreds of thousands who chose to participate would be eliminated from the census figures by statistical adjustment. Preliminary analysis in 2000 indicates that approximately one million people who returned their census forms would have their records nullified if statistical adjustment were employed - an ironic reward for cooperating with the census. Many of the nullified would come from groups that defy commonly held notions of the overcounted - such as Asian children and African American and American Indian homeowners.

  This is a stock "anti-statistical" critique. And it is a series of misrepresentations. First, the ACE process is not a counting process but an estimation one. Second, the claim ignores the fact that ACE counts real people missed by the census in the ACE blocks. Third, ACE removes real duplicates, fabrications, and other types of census errors in the ACE blocks. Then the process generalizes from the ACE sample blocks as organized by the post-strata to the population of all blocks. At this level there is also the estimation of the counts for the missing cells in the table in Table 1 using formula (1). Finally the adjustment factors are carried back down to the blocks.

  We also note that no one is removed from or added to the census files through ACE. Rather ACE produces adjusted counts at the block level, not because blocks are of interest in their own right, but so that they can be aggregated to for legislative and congressional districts. The figure *of 1 million corresponds to the number of net erroneous enumerations measured at the block level once the ACE adjustment factors are applied (see the related discussion in Section 4.1). It dramatically understates the level of erroneous enumerations at the block level in the census which range up to 12 million.*

- With statistical adjustment, most of the data used to adjust census figures in a particular state, city or town are from other states, other counties and other cities. This "synthetic estimation" process would apply individual data taken from distant, disparate parts of the country to other geographic locations. For example, it could adjust the profiles of some Mississippians with data taken from people in Alaska.

  The first sentence is correct in the sense that ACE borrows strength for its estimation by using data across geographic boundaries, but the claim that this is true for most of the data suggests that for big states such as California, the state data is swamped by data from remote locations in other states across the country. This could not be further from the truth. Further the majority of the ACE poststrata have a geographic component, and the collapsing of the number of poststrata from 448 to 417 further reduces the numbers of non-geographic-specific ones. The reason these poststrata were not broken down further by geography was their small sizes!

  We agree with the CMBC members that the results would be improved through more and better uses of geography and this is why we have concluded that it would have been preferable to use many more strata and a logistic regression or some other model to carry the results down to the block level.

- The A.C.E. survey is the largest post-enumeration study ever done and is a useful evaluative tool to judge the overall accuracy and effectiveness of Census 2000. Despite its record size, however, survey data cannot be reliably used to actually adjust census counts at the block or neighborhood level with statistical validity as its supporters argue.

  ACE is certainly the largest post-enumeration study ever done and its 98%+ response rate is remarkable. As we have noted repeatedly here and in our previous writings (e.g. Anderson and Fienberg, 2001), ACE adjusted-data were never intended to be used at the block level, even

> *though we now believe that there is evidence showing their possible superiority at such levels of geography. Our evaluation, which pays close attention to the real levels of gross error, makes clear that adjustment appears to work well for precisely those geographic areas we care about the most, e.g., at the size of congressional districts.*

The Census Monitoring Board is scheduled to end its oversight activities in September, 2001, and thus we can expect final reports from the two components of the board at that time.

Acknowledgments

The preparation of this paper was facilitated by early access to information on ACE through the U.S. Census Monitoring Board, but virtually all of the materials we have drawn upon are now publicly available mainly through the Census Bureau website. We are especially grateful to John Thompson and Howard Hogan for clarifying some points and directing us to specific details or providing us with information that was difficult to extract directly from the reports. William Bell, Frances Bourne, Gordon Brackstone, Mary Mulry, and Bruce Spencer provided helpful comments and references. Eugene Ericksen and Jeffrey Passel shared various draft memorandum relating to the accuracy of the 2000 census and ACE and demographic analyses. Our earlier work evaluating the 1990 census adjustment decision was supported by the Russell Sage Foundation and the Donner Foundation. Fienberg has received support from the Census Bureau in the form of contracts for research on confidentiality and disclosure limitation and on multiple system estimation methodology. Neither of us has accepted any other grants or contracts from or served as paid consultants to any party involved in the litigation surrounding census adjustment issues.

References

[1] Anderson, M. and Fienberg, S.E. (2000). History, myth making, and statistics: A short story about the reapportionment of Congress and the 1990 census. *PS: Political Science and Politics*, **XXXIII**, No. 4 (2000), 783–792.

[2] Anderson, M. and Fienberg, S.E. (2001a). *Who Counts? Census-Taking in Contemporary America.* Revised Paperback Edition. Russell Sage Foundation, New York.

[3] Anderson, M. and Fienberg, S.E. (2001b). U.S. census confidentiality: Perceptions and reality. *Bull. Int Statist. Inst.*, (2001), in press.

[4] Anderson, M., Daponte, B.O., Fienberg, S.E., Kadane, J.B., Spencer, B.D., and Steffey, D.L. (2000). Sampling-based adjustment of the 2000 census–A balanced perspective. *Jurimetrics*, **40(3)**, 341–356.

[5] Asher, J. and Fienberg, S.E. (2001). Variations on the administrative record census. Paper presented at the Joint Statistical Meetings, Atlanta, GA.

[6] Barron, Wm.G., Jr. (2001). Statement by William G. Barron Jr. on the Current Status of Results of Census 2000 Accuracy and Coverage Evaluation Survey. U.S. Bureau of the Census Press Release. `http://www.census.gov/Press-Release/www/2001/cb01cs06.html`

[7] Belin, T.R. and Rolph, J.E. (1994). Can We Reach Consensus on Census Adjustment? *Statistical Science*, **9**, 486–508.

[8] Bell, W.R. (1993). Using information from demographic analysis in post-enumeration survey estimation. *Journal of the American Statistical Association*, **88**, 1106–1118.

[9] Bell, W.R. (2001). *Accuracy and Coverage Evaluation Survey: Correlation Bias.* Report B-12, U.S. Bureau of the Census.

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

[11] Brown, L.D., Eaton, M.L., Freedman, D.A., Klein, S.P., Olshen, R.A., Wachter, K.W., Wells, M.T., and Ylvisaker, D. (1999). Statistical controversies in Census 2000. *Jurimetrics*, **39**, 347–375.

[12] Brown, J.J., Diamond, I.D., Chambers, R.L., Buckner, L., and Teague, A.D. (1999). A methodological strategy for a one-number census in the UK. *Journal of the Royal Statistical Society, Series A*, **162**, 247–267.

[13] Bryant, B.E. (1993). Census-taking for a litigious, data-driven society. *Chance*, **6** No. 4, 44-49.

[14] Chambers, R.L. and Cruddas, M. (2000). A one-number census for the United Kingdom. In "Window on the World," (C.F. Citro, ed.), *Chance*, **13**, No. 3, 38–41.

[15] Choi, C.Y., Steel, D.G., and Skinner, T. (1988). Adjusting the 1986 Australian census count for inder-enumeration. *Survey Methodology*, **14**, 173–189.

[16] Citro, C.F. and Cohen, M.L., eds. (1985). *The Bicentenial Census: New Directions for Methodology in 1990.* Panel on Decennial Census Methodology. National Academy Press, Washington, DC.

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

[18] Darroch, J.N., Fienberg, S.E., Glonek, G.F.V., and Junker, B.W. (1993). A Three-Sample Multiple-Recapture Approach to Census Population Estimation With Heterogeneous Catchability. *Journal of the American Statistical Association*, **88**, 1137–1148.

[19] Davis, P.P. (2001). *Accuracy and Coverage Evaluation Survey: Dual System Estimation Results.* Report B-9, U.S. Bureau of the Census.

[20] Edmonston, B. and Schultze, C., eds. (1995). *Modernizing the U.S. Census.* Panel on Census Requirements in the Year 2000 and Beyond. National Academy Press, Washington, DC.

[21] Ericksen, E.P. and DeFonso, T.K. (1993). Beyond the net undercount: How to Measure Census Error. *Chance*, **6** No. 4, 38–43.

[22] Ericksen, E.P., and Fienberg, S.E., and Kadane, J.B. (1994). Comment on three related papers on census adjustment. *Statistical Science*, **9**, 511–515.

[23] Fay, R.E. (1991). Hypothesis tests for improvement. Memorandum for the Undercount Steering Committee. U.S. Bureau of the Census, Washington, DC, June 28, 1991.

[24] Fay, R.E. and Thompson, J. (1993). The 1990 Post Enumeration Survey: Statistical lessons, in hindsight, *Proceedings of the Bureau of the Census Annual Research Conference*, U.S. Bureau of the Census, Suitland, MD, 71–91.

[25] Fienberg, S.E. (1972). The multiple recapture census for closed populations and incomplete 2^k contingency tables. *Biometrika*, **59**, 591–603.

[26] Fienberg, S.E., Johnson, M.S., and Junker, B.W. (1999). Classical multilevel and Bayesian approaches to population size estimation using multiple lists. *Journal of the Royal Statistical Society, Series A*, **162**, 383–405.

[27] Freedman, D.A., Stark, P.B., and Wachter, K.W. (2001). A probability model for census adjustment. *Mathematical Population Studies*, **9**, 165–180.

[28] Freedman, D.A. and Wachter, K.W. (1994). Heterogeneity and census adjustment. *Statistical Science*, **9**, 476–485.

[29] Freedman, D.A. and Wachter, K.W. (2001). Census Adjustment: Statistical promise or statistical illusion? Technical Report No. 596. Department of Statistics, University of California, Berkeley. May 30, 2001. `http://www.stat.berkeley.edu/ census/596.pdf`

[30] Freedman, D.A., Wachter, K.W., Coster, D.C., Cutler, D.R., and Klein, S.P. (1993). Adjusting the census of 1990: The smoothing model. *Evaluation Review*, **17**, 371–443.

[31] Freedman, D.A., Wachter, K.W., Cutler, D.R., and Klein, S.P. (1994). Adjusting the U.S. census of 1990: Loss functions. *Evaluation Review*. **18**, 243–280.

[32] Gambino, J. and Dick, P. (2000) Small area estimation practice at Statistics Canada. *Statistics in Transition*, **4**, 597–610.

[33] Griffith, R.A. and Malec, D.J. (2001). *Accuracy and Coverage Evaluation Survey: Assessment of Synthetic Assumption.* Report B-14, U.S. Bureau of the Census.

[34] Hogan, H. (1993). The 1990 Post-Enumeration Survey: Operations and Results. *Journal of the American Statistical Association*, **88**, 1047–1061.

[35] Hogan, H. (2001). *Accuracy and Coverage Evaluation: Data and Analysis To Inform the ESCAP Report.* Report B-1, U.S. Bureau of the Census. Also included as an attachment to the ESCAP Report, *Federal Register*, Vol. 66, No. 46, March 8, 2001, 14021-14046.

[36] Kadane, J.B. and Meyer, M.M. and Tukey, J.W. (1999). Corelation Bias in the presence of stratum heterogeneity. *Journal of the American Statistical Association*, **94**, 855–859.

[37] Killion, R.A. (1997). *Memorandum for John Thompson on Gross Errors and Erroneous Enumerations in the 1990 Decennial Census.* DSSD Briefs, information, and Topics Memorandum Sereis F-2. U.S. Bureau of the Census.

[38] Kim, J.J. and Zaslavsky, A. and Blodgett, R. (1995). Between-state heterogeneity of undercount rates and surrogate variables in the 1990 U.S. Census. *Survey Methodology*, **21**, 55–62.

[39] Miskura, S.M. (200). Results of reinstatement rules for the housing unit duplication operations. Memorandum for P.J. Waite, U.S. Bureau of the Census. November 21, 2000.

[40] Mulry, M.H. and Spencer, B.D. (1993). Accurracy of the 1990 census and undercount adjustments. *Journal of the American Statistical Association*, **88**, 1080–1091.

[41] Mulry, M.H. and Spencer, B.D. (2001a). *Overview of Total Error Modeling and Loss Function Analysis.* Report B-19, U.S. Bureau of the Census.

[42] Mulry, M.H. and Spencer, B.D. (2001b). Uncertainty and ambiguity in the 2001 decision against census adjustment for redistricting. Institute

for Policy Research Working Paper WP-01-8, Northwestern University, Evanston, IL.

[43] Navarro, A. and Asiala, M. (2001). *Accuracy and Coverage Evaluation Survey: Comparing Accuracy.* Report B-13, U.S. Bureau of the Census.

[44] Navarro, A. and Olson, D.B. (2001). *Accuracy and Coverage Evaluation Survey: Effect of Targeted Extended Search.* Report B-18, U.S. Bureau of the Census.

[45] Paice, J. and Steel, D. (2000). Census adjustment in Australia. In "Window on the World," (C.F. Citro, ed.), *Chance*, **13**, No. 3, 41–42.

[46] Passel, J.S. (2001). Comparison of Demographic analysis, A.C.E., and Census 2000 Results by Race. Unpublished memorandum prepared for the U.S. Census Monitoring Board, Presidential Members, February 27, 2001.

[47] Prewitt, K. (2000). Accuracy and coverage evaluation: Statement on the feasibility of using statistical methods to improve the accuracy of the census. U.S. Bureau of the Census, June, 2000.

[48] Schirm, A.L. (1991). The effects of census undercount adjustment on congressional apportionment. *Journal of the American Statistical Association*, **86**, 526–541.

[49] Skerry P. (2000). *Counting on the Census? Race, Group Identity, and the Evasion of Politics.* Brookings Institution Press, Washington, DC.

[50] Spencer, B.D. (2001). A method for deciding whether adjustment of census 2000 improves redistricting. Institute for Policy Research Working Paper WP 00-27. Northwestern University, January 21, 2001. `http://www.nwu.edu/IPR/publications/papers/wp-00-27.pdf`.

[51] Stark, P.J. (2000). The 1990 and 2000 census adjustment plans, Technical Report 550. Department of Statistics, University of California, Berkeley.

[52] Statistics Canada (2000). *Annual Demographic Statistics 2000.* (Statistics Canada Catalog No. 91-213-XPB), Ottawa, Canada.

[53] U.S. Bureau of the Census Executive Steering Committee for Accuracy and Coverage Evaluation Policy (2001). *Report of Tabulations of Population to States and Localities. Federal Register*, Vol. 66, No. 46, March 8, 2001, 14004-14046. `http://www.census.gov/dmd/www/EscapRep.html`).

[54] U.S. Census Monitoring Board, Congressional Members (1999). *Unkept Promise: Statistical Adjustment Fails to Eliminate Local Undercounts, as Revealed by Evaluation of Severely Undercounted Blocks From the 1990 Census Plan.* Report to Congress, September 30, 1999. (Available from http://www.cmbc.gov/)

[55] U.S. Census Monitoring Board, Congressional Members (2001). *A Guide to Statistical Adjustment: How It Really Works.* Report to Congress, June 7, 2001. (Available from http://www.cmbc.gov/)

[56] U.S. Census Monitoring Board, Presidential Members (2001). Report to Congress, April 11, 2001. (Available from http://www.cmbp.gov/)

[57] Wachter, K.W. and Freedman, D.A. (2000). Measuring local heterogeneity with 1990 U.S. census data. *Demographic Research*, **3**. `http://www.demographic-research.org/Volumes/Vol3/10`

[58] Wachter, K.W. and Freedman, D.A. (2001). The fifth cell. *Evaluation Review*, **24**, 191-211.

[59] Zaslavsky, A.M. and Wolfgang, G.S. (1993). Triple-system modeling of census, post-enumeration survey, and administrative-list data. *Journal of Business & Economic Statistics*, **11**, 279–288.