Differences between the 1990 and 2000 Census Adjustment Plans, and their Impact on Error

P.B. Stark
Department of Statistics
University of California
Berkeley, CA 94720-3860
stark@stat.berkeley.edu

Technical Report 550
16 March 1999
Revised 3 May 2000

Abstract

A revised plan for the 2000 Decennial Census was announced in a 24 February 1999 Bureau of the Census publication [9] and a press statement by K. Prewitt, Director of the Bureau of the Census [39]. Census 2000 will include counts and "adjusted" counts. The adjustments involve complicated procedures and calculations on data from a sample of blocks, extrapolated throughout the country to demographic groups called "post-strata." The 2000 adjustment plan is called Accuracy and Coverage Evaluation (ACE). ACE is quite similar to the 1990 adjustment plan, called the Post-Enumeration Survey (PES). The 1990 PES fails some plausibility checks [4, 12, 44] and probably would have reduced the accuracy of counts and state shares [3, 4].

ACE and PES differ in sample size, data capture, timing, record matching, post-stratification, methods to compensate for missing data, the treatment of movers, and details of the data analysis [6, 10, 11, 25, 26, 27, 28, 31, 32, 35, 41]. ACE improves on PES in a number of ways, including using a larger sample, using a simpler model to assign "match probabilities" to records with insufficient data, and incorporating mail-back return rates into some post-strata. Nonetheless, ACE shares the most serious problems of PES. The "Be Counted" program, census response submission over the internet, computer unduplication of records, the treatment of movers [10, 25], a new definition of "correct address," more limited search for matching records [35], the use of optical character recognition (OCR) to capture data [7, 8, 9, 10, 31], the data collection schedule [10], and the assignment of "residence probabilities" to some sample records [6, 32], are likely to make ACE less accurate than the 1990 PES.

1 Introduction

The census makes mistakes. It misses some people it should count, or finds them in the wrong place. It sometimes counts fictitious people. The principal kinds of census errors are

- Gross omission—failing to count someone where he or she belongs
- Erroneous enumeration—counting someone who does not exist, counting someone more than once, or counting someone where he or she does not belong

Gross omissions reduce the count; erroneous enumerations inflate the count. The same individual can contribute both kinds of error. For example, if a person's address is recorded incorrectly, he can be a gross omission at his correct address and an erroneous enumeration at his incorrect address. The two kinds of errors cancel to some extent, but overall, the census misses some people, resulting in net undercount. The undercount rate is different for different demographic groups, as shown in Table 1.

Group	undercount
Total	1.85%
Non-Black	1.3%
Black	5.7%
Male	2.8%
Female	0.9%

Table 1: 1990 Undercount rate estimated by Demographic Analysis. Source:[40]

The numbers in Table 1 were derived using "Demographic Analysis" (DA) [40], which estimates the total U.S. population by

$$\text{births} - \text{deaths} + \text{immigration} - \text{emigration}.$$

Demographic Analysis estimates the population at the national level, but not more locally, because inter- and intra-state migration are not tracked. Moreover, emigration is not tracked very well, and not every birth or death is recorded, so even the national totals are uncertain.

If the census missed people at the same rate regardless of where they live, the undercount would affect population counts, but not population shares. Geographic variation in the undercount rate produces errors in state population shares, which determine congressional representation and the allocation of Federal funds [4]. Adjustment tries to estimate how many people the census missed in each block in the U.S., to correct the census.

The details of adjustment are quite complex [3, 6, 10, 11, 25, 27, 26, 30, 31, 32, 35, 48], but in sketch, adjustment works like this: The Census Bureau tries to match persons found in a random sample of blocks after census day (post-enumeration, "P-sample") to census records from the same set of blocks (the "E-sample"). Persons found in the P-sample but not in the E-sample (after searching) are concluded to have been missed by the census. Persons found in the E-sample but not in the P-sample (after searching) are concluded to have been counted erroneously by the census. The rates at which persons in different groups (called "post-strata") are missed by the census are extrapolated to all such persons counted in the census in the U.S., block by block.

Typically, sampling is used to make inferences about average properties of the whole from some of its parts, but in census adjustment, sampling is used to make inferences about parts from other parts, in extremely fine geographic detail. This use of sampling in the U.S. Decennial Census is controversial, and arguments have become highly politicised. Generally, Democrats favor adjusting the census and Republicans oppose it. Some think that adjustment would increase the number of seats in the House of Representatives held by Democrats; allocation of federal funds is also affected [4]. On the whole, Democrats and the mass media seem to believe that an adjusted census would be more accurate; Republicans have been more sceptical. Some National Academy of Sciences reports recommend using sampling-based adjustments for the census [15, 50], but they do not examine any particular adjustment plan, nor do they contain theoretical or empirical evidence that sampling-based adjustment improves accuracy.

In 1990, adjustment made so many mistakes that it probably would have made the census worse [3, 4, 36, 17, 18, 20, 22, 23, 29, 36, 44, 45, 46]. The same was true in 1980 [19]. Statistical arguments that adjustment improves accuracy rely on many faulty assumptions [3, 4, 20, 22, 23, 29, 36, 44, 46]. Nonetheless, the Bureau of the Census continues to assert that adjustment will improve census accuracy [38, 39].

The census is not perfect, and statistical sampling is a useful scientific technique; that does not imply that adjustment improves the census. Sampling is only part of the adjustment procedure. The causes of mistakes in the census also produce mistakes in the adjustment, but adjustment has additional sources of error [3, 4, 13], and relies on complicated computations and models [1, 30, 34, 48]. Each person in the sample can have enormous "leverage" on the estimated undercount: each adjustment error can produce the appearance of thousands of mistakes in the census. In contrast, counting is robust.

Adjustment has two kinds of error:

- *sampling error*, from the luck of the draw in selecting the blocks for the P-sample
- *systematic error* or *bias*, from bad data, processing errors, faulty assumptions, and so on.

Sampling errors tend to average out. Bias does not. Increasing the size of the sample tends to decrease sampling error, but not bias. A single programming error, undetected in Census Bureau evaluation studies, was discovered to have added about a million people to the 1990 adjustment [36]; this is an example of bias. In 1990, 60% to 80% of the proposed adjustment was measured bias in the procedure, not measured undercount [3, 4, 36].

The Census Bureau had proposed for 2000 a "one number census" using sampling based adjustments—raw counts would not have been available [7]. The plan also proposed using sampling to select for follow-up only a fraction of households that did not mail back their questionaires. (In the past, attempts were made to follow up all non-responders.) The plan was challenged in two suits against the Department of Commerce [4]. On 25 January 1999 the Supreme Court upheld a lower court decision that sampling could not be used in the census for the purpose of apportionment of seats in the House of Representatives.

In response, the Bureau of the Census modified its plans for 2000; an overview of the revised plan was released in February 1999 [9, 39]. The revision called for a traditional headcount for the state-level apportionment figures to be reported to Congress by 31 December 2000, followed by a set of adjusted numbers which could be used for redistricting and other purposes, to be released by 31 March 2001 [9, 39]. Additional details of the new plan recently became available [6, 10, 11, 25, 27, 26, 28, 31, 32, 35]. The latest adjustment plan for 2000 (ACE) is much closer to the 1990 adjustment scheme (PES) than were previous plans to adjust the 2000 Census [7, 48].

The plan to adjust the 2000 Census (ACE) differs in detail from the 1990 adjustment plan (PES), but basically the two plans are the same. ACE improves on PES in some ways; in other ways, ACE is worse than PES. Section 4 argues that on balance ACE is likely to have larger errors than PES, for reasons that include the "Be Counted" program and internet response submission, the treatment of movers, the use of optical character recognition (OCR) to scan handwritten data, and record matching procedures. The 2000 headcount is likely to be less accurate than the 1990 headcount for a variety of reasons, especially the state of the Master Address File. The 2000 headcount might have such large errors that even a poor adjustment would improve it; that was not true in 1990.

The discussion below relies on information thought to be current as of 2 May 2000 [6, 10, 11, 25, 27, 26, 28, 31, 32, 35, 41]. Some aspects of the census itself are inferred from earlier versions of the 2000 plan [7, 8, 48]. Section 2 explains adjustment methodology in more detail, Section 3 reviews the 1990 PES. Section 4 lists the most important changes between 1990 PES and 2000 ACE and discusses their likely impact on accuracy, along with some differences between the 1990 and 2000 census plans that could affect the accuracy of 2000 headcount data.

2 Background

2.1 Capture-recapture and adjustment

Sampling-based census adjustment uses a technique called "capture-recapture." The idea is quite old. According to [13], it dates back to Laplace in 1783, and it has been used to estimate wildlife populations at least since the late 1800s ([37], cited in the annotated bibliography [16]).

Suppose we seek to estimate the number of fish in a pond. The capture-recapture approach is to

- catch some fish, tag them, and let them go
- wait for the tagged fish to mix with the others
- catch another lot of fish, and count the number with tags

The fraction of fish with tags among the second catch estimates the fraction of the whole population caught the first time, so an estimate of the total number of fish in the pond is

$$\frac{\text{number in the first catch}}{\text{fraction of fish in the second catch with tags}}. \tag{1}$$

This estimate is reasonable if certain conditions hold. For example:

1. We must know how many fish we tagged.
2. The number of fish in the pond must be the constant.
3. The tags must be impossible to misread.
4. All the fish need to have the same propensity to be caught the second time (the second catch needs to be like a random sample of the fish in the pond).

The analogous approach to census adjustment is to

- Take the census.
- Take a random sample of people, and determine how many of them were also found by the census.
- Estimate the total number of people by

$$\frac{\text{number found by the census}}{\text{fraction of people in the sample who were also found by the census}}. \qquad (2)$$

Such a simple approach will not work for people, however. The census contains some erroneous enumerations: in the analogy, we do not know exactly how many fish were tagged. It is quite hard to determine whether someone found in the P-sample was also found by the census: the census cannot record unique individual identifiers, such as social security numbers or fingerprints, because of privacy issues. Instead, the Bureau relies on matching records between the census and the P-sample using names and addresses, which can be missing entirely or recorded with error, both in the census and in the sample. In the analogy, we cannot always tell whether or not a fish in the second catch was tagged.

One cannot take a random sample of people in the U.S.; there is no list from which to draw a random sample (if there were, the census would be unnecessary). Instead, the Census Bureau takes a stratified random sample of blocks, and surveys the residents of the housing units in each block. Because address lists within blocks have errors and omissions, the sampling frame (the universe from which the P-sample is drawn) does not include the entire population, nor even every housing unit.

People do not have equal propensities to be caught by the census, or by the sample survey. Some people are hard to find; others deliberately avoid the census. Some of the difficulty is associated with demography and geography, which leads to geographic differences in the undercount rate. The adjustment procedure tries to circumvent this problem by dividing people into post-strata with similar demographic characteristics, on the assumption that the undercount rate depends only on those characteristics. For example, one 1990 PES post-stratum consisted of all black male renters age 30-44 living in the central city of major metropolitan areas in New England. The 1990 PES had 1,392 post-strata. The current plan for 2000 ACE has 448 post-strata for the U.S. and 84 for Puerto Rico (see tables 3, 4 and 5); some of the post-strata might be combined if they contain too few people [26]. In the analogy, the second catch is not like a simple random sample of fish from the pond.

In 1990, about 7% of the U.S. population moved between census day and the end of the PES, roughly four times the 1990 estimated undercount. The Census Bureau tries to identify people in the census who have moved from the P-sample blocks since census day, and to determine where people in the P-sample blocks were living on census day. In 1990, weighted to the nation, people identified by the PES as having moved into the sample blocks after census day contributed about 5 million to the estimated number of gross omissions ([4], p. 358). Identifying and tracking the movers and accounting for them in the estimate adds complexity and error to the adjustment. In the analogy, the number of fish in the pond is changing.

Here is a slightly more realistic sketch of the adjustment strategy:

- Take the census.
- Take a stratified random sample of block groups after the census is taken (the P-sample).

  For each block group in the P-sample:

  - Take a survey.
  - For each person found in the survey, determine whether the census has a record of that person in the same block group.
  - For each census record in the block group, determine whether the person is in the survey, or was erroneously omitted from the survey, or was enumerated erroneously in the census.
- Pool results for the blocks in the P-sample to get sample fractions omitted and erroneously enumerated for each post-stratum.
- Use the counts for each post-stratum, the estimated erroneous enumerations for each post stratum, and the estimated gross omissions for each post stratum, to construct an adjustment factor for each post stratum:
  - The fraction of people in a post-stratum in the P-sample blocks who were not recorded in the census is an estimate of the fraction of census gross omissions in the post-stratum.
  - The fraction of people in the census in a post-stratum in the sample blocks who were not recorded in the survey is an estimate of the fraction of census erroneous enumerations in the post-stratum.
  - The adjustment factor for the post-stratum is

$$\left(1 - \frac{\#\text{erroneous enumerations}}{\text{census count}}\right) \times \frac{\text{survey count}}{\#\text{matching records}}. \quad (3)$$

- Adjust the count in every block in the country according to the number of members of each post-stratum it contains, using the adjustment factors determined from the P-sample. That is, for each block in the country,

– For each post-stratum, multiply the number of people the census found in the block in the post-stratum by the adjustment factor for the post-stratum.

Then add the products to obtain the adjusted population estimate for the block.

This is still just a sketch. It omits details of determining whether or not someone in the survey matches a census record, and *vice versa*; following up unmatched records to check whether they are errors in the census, the survey, or neither; searching nearby blocks for matching records; imputing missing data; accounting for movers; following up nonresponders to the survey; weighting records to account for uneven sampling rates; combining numbers from different blocks to estimate fractions in post-strata; and rounding to obtain integer-valued estimates.

The final estimate relies on some dubious assumptions about how people behave, including:

Independence. The probability of being counted in the survey is the same for every individual in a post-stratum, whether or not the individual was enumerated in the census.

Failure of the independence assumption leads to "correlation bias."

Perfect Matching. One can tell definitively whether or not a person in the P-sample was counted in the census, and *vice versa*.

Violation of this assumption leads to "matching error."

The Survey is Perfect. The adjustment process tacitly assumes that the survey has no fabricated interviews or address errors. If a survey record is not matched by any census record (possibly after follow-up), the record is concluded to be a gross omission from the census. In contrast, if a census record is not matched by any survey record, more effort is expended in follow-up to determine whether or not the record is an erroneous enumeration. This tends to produce "balancing error." (Another sort of "balancing error" has to do with the definitions of the "correct address" in the survey and the census; see [31].)

Synthetic Assumption (Homogeneity). The census undercount rate is constant within post-strata across blocks.

Failure of the synthetic assumption leads to "heterogeneity bias."

The first three assumptions are needed to estimate census undercount for post-strata straightforwardly from the P-sample, with relatively small bias. The synthetic assumption allows the undercount estimate to be extrapolated to blocks outside the sample. Unfortunately, none of these assumptions holds, and violations can bias the adjustment up or down; see [24] for a probability model for census adjustment that clearly separates correlation bias, heterogeneity, and ratio-estimator bias.

3 What happened in 1990?

The 1990 Census counted about 248.7 million people in the U.S. The Census Bureau used a sample (the Post-Enumeration Survey or PES [30]) to estimate the 1990 undercount. The PES comprises the sample survey data collection, the post-stratification, the rules for matching and follow-up, and the data processing, which is based on the Dual System Estimate (DSE). The sample included about 380,000 people in 166,000 households in 5,290 block groups.

Secretary of Commerce Mossbacher decided not to adjust the official 1990 census numbers using the PES. The City of New York, et al., subsequently sued the Federal government to try to compel adjustment. In July 1991 (when Secretary Mossbacher had to decide whether or not to adjust the official numbers), the adjustment would have added about 5.3 million people, more than demographic analysis indicated the census had missed (see Table 2).

3.1 The 1990 adjustments do not make sense

3.1.1 Who gets added?

Darga [12] questions the plausibility of some details of the 1990 adjustment. One of the most stable demographic parameters across geography and culture is the gender ratio of children under the age of 10: about 51% boys to 49% girls. This is essentially what was observed in the 1990 census head count. The fraction of boys among the 284 metropolitan statistical areas and consolidated metropolitan statistical areas ranged from 50.3% in Pine Bluff, Arkansas, to 52.1% in Topeka, Kansas—even for relatively small geographical units, the unadjusted gender ratios made sense. Darga argues that adjustment moves the gender ratios to highly implausible extremes (48% to 56% boys); some of the variation could be sampling error.

3.1.2 Comparison with Demographic Analysis.

Population estimates from the 1990 PES and Demographic Analysis (DA) are in rough agreement. The Census and the PES miss many of the same people, including homeless, and people who do not want to be found; such persons are less likely to be missed by DA. The only way that the PES can approach DA totals at the national level, therefore, is if it misclassifies "matches" as "non-matches." Record matching is hardest in areas that are hardest to enumerate [51], so adjustment is most likely to have the largest errors where it is needed most. Breiman [3] argues that many of the non-matches are errors; see section 3.2 below. There is no reason to think that the geographical or demographic distribution of the classification errors mirrors the true undercount.

Population estimates from Demographic Analysis and the 1990 PES do not agree in detail. The 1990 PES estimate is *higher* than the DA estimate: according to the PES, there are over a million more women, and about half a million fewer men (see Table 2). The PES estimate of the number of women missed by the census is about twice the DA estimate. The CAPE report notes the discrepancy:

> ... there was concern that the PES estimated a higher population than DA and estimated about a million more women than DA. ([36] p. 27)

Group	PES	DA	Difference
Black Male	804,000	1,338,000	-534,000
Black Female	716,000	498,000	+218,000
Other Male	2,205,000	2,142,000	+63,000
Other Female	1,544,000	706,000	+838,000

Table 2: National-level estimates of the 1990 census undercount from the 1991 revision of the Post Enumeration Survey and Demographic Analysis (DA). Table from [4].

3.1.3 Where do the people go?

Figure 1 of [4] illustrates the effect of the proposed 1990 adjustment on State shares. The share changes are miniscule: from -695 parts per million (Pennsylvania) to +1,956 parts per million (California), with typical changes of tens to hundreds of parts per million. Estimating such small changes accurately is extremely difficult.

New York, Pennsylvania, and Illinois lose shares in the 1990 adjustment. Texas and Arizona gain shares. Arguably, it is easier to count in Dallas and Phoenix than in the Bronx, Philadelphia, and Chicago; if so, adjustment should move population shares northeast, not southwest.

3.2 Error in the 1990 PES

This section summarizes some earlier studies of the accuracy of the 1990 PES, including work by the Census Bureau, The Committee on Adjustment of Postcensal Estimates Asessment of Accuracy of Adjusted Versus Unadjusted 1990 Census Base for Use in Intercensal Estimates (CAPE Report [36]), and Breiman [3]. Breiman combines information from Bureau of the Census evaluation studies to estimate various biases in the PES. Breiman considers bias caused by fabrications by interviewers, matching errors, census day address errors, bias in the ratio estimator, people discovered to be out-of-scope in reinterview, late census data, and the computer coding error. See also [4].

The Census Bureau rematched records in the 104 block groups with the most extreme match rates, which decreased the adjustment by about 300,000 people. In the process of rematching, the Census Bureau found a programming error that had inflated the undercount estimate by about 1,000,000 people. It is now acknowledged that about 2-3 million of the PES adjustment is error in the adjustment, not error in the census [3, 36].

> As a result of an error in computer processing, the estimated national undercount rate of 2.1% was overstated by 0.4%. After correcting the computer error, the national level of undercount was estimated to be about 1.7%. After making other refinements and corrections, the national undercount is now estimated to be about 1.6% [the figure is 1.58% in attachment 3, Table 2] ... The level of total bias, excluding correlation bias, on the revised estimate of undercount is negative 0.73 (-0.73%). ([36] p. 15)

Thus about (2.1 - 1.58 + 0.73)/2.1 = 60% of the original undercount estimate of 2.1% is bias. The report continues, evaluating the "revised" estimates, which correct the coding error and use different post-strata:

> Therefore, about 45% (0.73/1.58) of the revised undercount is actually measured bias and not measured undercount. In 7 of the 10 evaluation strata, 50% or more of the estimated undercount is bias. ([36] p. 15)

According to then Director of the Bureau of the Census, even after correcting the programming error and re-matching some records,

> A significant amount of bias remains. The research estimates that, at the national level, removing all biases from the PES estimates would lower the estimated undercount from 1.6 to 1.3 percent. When the effect of correlation bias is not taken into account ... the estimated undercount would fall to 0.9 percent. [43]

3.2.1 Matching error.

Search Region. The match rate depends strongly on the geographic search area, because addresses are sometimes recorded and/or "geocoded" incorrectly. In 1990, the search area was a ring of one or two blocks around the target address [3, 30]. Increasing the search area tends to increase the match rate, decreasing both the estimated erroneous enumerations and the estimated gross omissions. If the Census Bureau had not searched a ring of blocks, the undercount estimate would have been about double, roughly 4% ([3], p. 468). In the 1986 Los Angeles PES rehearsal, of those individuals who matched outside their target block, 38% matched more than 5 blocks away ([3] pp. 473–4). The Census Bureau estimated that 75% of the non-matching records in the 1990 PES could have been converted to matches had the search area been expanded ([3] p. 473). Breiman ([3], p. 473) argues that if the 1990 search area had been expanded to 6-8 rings of blocks, the undercount estimate would have dropped by over 1,000,000. Together, these facts strongly suggest that most of the non-matching records were false non-matches.

Personnel. Production matching began with a computer matching program; cases the computer could not resolve (about 25%) were sent to human matchers. Two production teams worked on the cases not matched by the computer. The two teams disagreed in 10.7% of the matched cases, 6.6% of the unmatched cases, and 31.2% of the unresolved cases ([3] Tables 5–7, p. 466). Census Bureau study P7 [14] compares production record matching with expert record rematching (results are summarized in [3]). The match and rematch were dependent—the experts saw the production classifications. Among the records determined in production not to match, the experts disagreed with the classification assigned in production in about 8% of cases. Among records determined in production to match, the experts disagreed with the production classification in about 0.5% of cases. Among records whose match status was unresolved in production, experts disagreed with the classification assigned in production in about 23.8% of cases. The overall match/rematch discordance rate was 1.8%.

If the experts' match determinations are taken as ground truth, this would mean that 8% of the non-matching cases really should be matches, and 0.5% of the matching cases should be non-matches. These errors add bias to the adjustment.

Modeling unresolved cases. In 1990, about 2% of the records lacked sufficient information to determine whether or not there was a match. Weighted to the nation, this amounts to roughly 4,000,000 people in the census, and 4,000,000 in the sample survey [4]. Depending on how the cases with unresolved match status are treated, the undercount estimate can be anywhere between about 9,000,000 (undercount) to −1,000,000 (overcount) ([3], p. 469). See also [45] for "half-high" and "half-low" estimates. The Census Bureau handled unresolved cases in 1990 by assigning a "match probability" using two hierarchical logistic regression models (one for unresolved census cases, and one for unresolved PES cases) [1]. The models are *ad hoc*. They are calibrated on cases whose match status was resolved; those cases are likely to be quite different from the unresolved cases to which the models are applied. Moreover, at least one explanatory variable in the model is missing for 28% of the unresolved PES cases, and 38% of the unresolved census (E-sample) cases; those missing variables were imputed using another model ([3], p. 469). A critique can be found in [46].

3.2.2 Survey Errors.

The DSE assumes that the survey is ground truth against which to compare the census; tacitly this requires that the survey be completely accurate—no fabricated interviews, no address errors, *etc.* In fact, fabrications are unavoidable. The rate of fabricated interviews in 1990 was estimated to be from about 0.5% to 1.5% [49]. Just 13 (detected) fabrications identified in 1990 added about 50,000 to the undercount estimate. A 1% fabrication rate would have inflated the undercount by about 1.7 million ([3], p. 467).

3.2.3 Correlation Bias.

Correlation bias is hard to estimate in detail. One source of correlation bias is people unreachable by any survey; if they are unreachable by the census and the PES, determining their locations and effect on the estimate must depend crucially on strong assumptions. The Census Bureau estimated correlation bias in the 1990 PES by disaggregating demographic analysis estimates from national to local levels using a statistical model [2]. The model implies that the 1990 census missed nearly 900,000 white males overall, but only 13 between 20 and 30 years old; and that the 1990 census missed more than 750,000 black males, but counted almost 30,000 too many black males under age 10 ([21], p. 533). Wachter and Freedman [47] estimate the net correlation bias of the 1990 PES to be about 3.04 million by subtracting the PES from the sum of the demographic analysis estimate and the measured bias, using 3.63 million as the measured bias, midway between the Census Bureau's and Breiman's figures [3].

CAPE expressed concern about the treatment of correlation bias in the 1990 proposed adjustment:

> The fourth cell in the DSE is an estimate of the number of people missed in both the PES and the census ... Both the Committee and the Panel of Experts were very

> concerned about the negative values in the fourth cell ... correlation bias should be a component of total error. However, there was concern about our method of estimating it and very serious concern about the method of allocating it. ([36] pp. 22–3)

3.2.4 Heterogeneity Bias.

In 1990, heterogeneity within post-strata was significant, and had a large effect on the PES ([29] and [20]; see also [44] and [4] for discussion). According to then Director of the Bureau of the Census, "... it is possible that errors due to heterogeneity in fact could be larger than all other sources of error in the adjustment" [43]. CAPE also studied heterogeneity:

> The Panel cautioned that artificial population analysis ... was inconclusive about whether the homogeneity assumption held. ([36], p. 30)
>
> A first analysis showed similar homogeneity for the 1,392 design as well as the 357 design as well as for a design with only 2 strata." ([36], p. 26).
>
> The level of bias in the PES was close to the point where artificial population analysis shows that homogeneity assumption fails to hold. ([36], p. 26)

3.2.5 Discussion.

CAPE conclude that the bias cannot be removed at a level of geographic detail that would improve the accuracy of state shares:

> The Census Bureau ... knew of no adequate methodology to remove the bias by state, city, etc. ([36] p. 30)
>
> ... there is no intention to adjust the 1990 census because research shows insufficient technical justification. ([36] p. 33)

According to then Director of the Bureau of the Census,

> ... no survey – either the high quality, well controlled and interviewed PES of 170,000 households or a larger one – can be used to make post-census fine tuning of an average undercount as small as 1.6 percent in all types of places, counties, and states at a level of accuracy beyond that by which surveys are usually judged. ... there is little or no evidence adjustment would improve the quality of substate estimates ... [43]

In summary, faulty assumptions, poor data quality, and *ad hoc* choices in the data analysis, such as the choice of search area for matching records and models for the "match probability" of cases with unresolvable match status [1], produced large biases in the PES. The biases are unrelated to true undercount. On the order of 60% [36] to 80% [3] of the 1990 PES was bias: adjustment adds more error than it fixes.

For adjustment to make the census better, biases need to cancel. Random errors tend to cancel; biases do not. Arguments that biases in the cancel depend on *ad hoc*, counterfactual statistical models with bizarre consequences. For a description of some of the models, see [2, 34]; see [21] for a critique. Removing the bias at useful levels of geography (states and smaller) is impossible [43].

4 Differences between 1990 and 2000 Plans

4.1 The Plan for 2000 ACE

The 2000 adjustment procedure was originally called Integrated Coverage Measurement (ICM), but has been renamed Accuracy and Coverage Estimate (ACE). ACE is largely the same as the 1990 PES. Some of the differences that affect accuracy are listed below.

The original 2000 plan used sampling for non-response follow-up (SNRFU) [7]. In addition to contributing sampling error to the census itself, SNRFU required using a more complicated and error-prone method to account for movers [48], entailing additional assumptions and degrading data quality. Following the Supreme Court decision on 25 January 1999, the Census Bureau dropped SNRFU from its plans, but kept the more complicated and error-prone treatment of movers ([25], pp. 6–7; also see section 4.9 below).

The following subsections discuss some differences between the 1990 and the 2000 plans, and their anticipated impact on accuracy. The differences are listed alphabetically, not in order of importance. The differences include advertising 4.2, the "Be Counted" program 4.3, computer-aided telephone interviews 4.4, demographic analysis and racial categories 4.5, evaluation studies 4.6, internet response submission 4.7, the state of the Master Address File 4.8, the treatment of movers 4.9, optical character recognition for data capture 4.10, post-stratification 4.11, record matching (including the definition of correct location, minimum information requirements, search region, personnel, and modeling "residence probability" and "match probability") 4.12, sample size 4.13, and timing 4.14. Some of these affect primarily the accuracy of the adjustment; some have more impact on the accuracy of the census headcount.

4.2 Advertising

The 2000 Census is the first to use paid advertising [7, 8, 9, 39]. This seems to have stemmed the deterioration in mail-back response rate; higher mail-back rates could increase the accuracy of the census headcount and reduce costs. If advertising also improves cooperation with the ACE survey beyond the level of cooperation with the 1990 PES, it could improve the accuracy of adjustments.

4.3 Be Counted

The "Be Counted" program is new for Census 2000 [7]. "Be Counted" forms are versions of the short form made available in public places, instead of being delivered to specific residence addresses. They are intended to reach persons who did not receive a mail-back form, or think they might not have been included in the response from their residence [7, 8, 9]. Making forms available in more than one way to the same people (mail-out, "Be Counted," and internet) will result in some duplications, which the Bureau plans to eliminate by "computer unduplication" [7, 8]. The computer "unduplication" of records will be complicated by OCR data capture of the names and addresses in "Be Counted" forms, and the names in mail-back forms (see section 4.10). Unduplication is subject to two kinds of errors: incorrectly concluding that distinct records are duplicates, and failing to detect a duplication. The first kind of error nullifies the potential benefit: the "Be Counted" form is ignored. The second kind of error produces erroneous enumerations, for example, when one family member returns the mail-back form, and another submits a "Be Counted" form. Unduplication errors are likely to be distributed unevenly with geography and demography.

"Be Counted" forms do not have pre-printed addresses, so address errors will be more common than for mail-out/mail-back forms. It seems unlikely that propagating "Be Counted" forms will elicit accurate responses from those disinclined to respond to a mail-back questionnaire. Moreover, they invite abuse.

It is plausible that the "Be Counted" will be most effective in areas with a high response rate: where people are most inclined to cooperate with the census, they might also be most inclined to obtain and return "Be Counted" forms. If so, errors in unduplicating "Be Counted" forms could increase the differential undercount by introducing a disproportionate number of erroneous enumerations in well counted areas.

The "Be Counted" program is targeted at traditionally undercounted groups; presumably, the "walk-in centers" at which the forms are to be made available will be concentrated where people are expected to be harder to enumerate. Some of the "Be Counted" records that survive "computer unduplication" will have incorrect names and/or addresses, possibly because of OCR errors (section 4.10). Some subset of those records will correspond to real people living in the correct block, who really are not duplicates. If the "Be Counted" records are merged with mail-back records before ACE (Census Bureau documentation is unclear on this point, but merging before ACE is consistent with what I have read), the ACE survey will not find a person with the name/address as scanned. If an incorrectly scanned address in a block in the ACE P-sample does not correspond to a real physical address, ACE will infer that the record is an erroneous enumeration, and subtract from the count in that block, and all others that intersect the corresponding post-stratum. This would offset the benefit of the "Be Counted" program.

4.4 CATI

The Census Bureau might use computer-aided telephone interviews (CATI) for the ACE interviews with households that responded to the census by mailing back their forms; they tried CATI for the first time in the dress rehearsal for Census 2000 [48]. The effect of CATI on accuracy is hard to guess, but it could reduce the cost of ACE.

4.5 Demographic Analysis (DA) and Racial Categories

Racial categories in the census do not correspond directly to racial categories in DA, partly because individuals assign their own racial labels in the census, while DA is based on administrative records, such as birth and death certificates, which are filled out by others. Racial categories in DA are limited to "black" and "non-black." In 1990, for the purpose of comparing DA with the census and the PES, it was assumed that if an individual identified himself as "black" in the census, his administrative records (such as birth certificate) would also list him as "black," and that if he did not identify himself as "black" in the census, his administrative records would not list him as "black." The mismatch between the definition of race in the census and in DA was not as obvious in 1990 as it will be in 2000, because in the 2000 census, individuals can identify themselves as belonging to more than one race.

This will make it much harder to compare DA estimates with the 2000 Census or ACE; any solution will involve some sort of modeling, entailing additional errors and uncertainties. For example, one might construct two sets of census forms, one with only "black" and "non-black" as racial categories, and the other allowing the full set of combinations. These two forms could be administered to a randomly selected group of individuals to try to calibrate the 2000 self categorization against DA categories. No matter how large the P-sample, there will remain uncertainty arising from the difference between race as identified by the individual, and race as identified by administrative records.

4.6 Evaluation Studies

The evaluation studies carried out by the Census Bureau as part of the 1990 Census and PES provided invaluable information about the accuracy of the 1990 census and the proposed 1990 adjustment. The evaluation studies covered a variety of sources of error, including

- distribution of missing data
- error in the imputation models
- address errors
- P-sample fabrications
- matching error
- balancing error
- correlation bias

Indeed, the evaluation studies gave some of the strongest evidence that the proposed 1990 adjustment was mostly bias [3]. As of February 2000, the Census Bureau had not planned evaluation studies for the 2000 Census and ACE (Howard Hogan, personal communication, 3 February 2000). If the Bureau does not perform an adequate collection of evaluation studies, there will be insufficient scientific evidence to judge whether the 2000 ACE would improve or degrade the accuracy of the 2000 Census. The 1990 studies do not apply directly to 2000, because of differences in personnel, timing, and procedures, including the wide use of OCR 4.10, reliance on computer unduplication (sections 4.3,4.7), new search rules and models for match status 4.12, the treatment of movers 4.9, post-stratification 4.11, *etc.*

4.7 Internet Response Submission

For some (as yet unspecified) households, 2000 census data can be submitted over the internet [7, 8, 9]. This will reduce OCR errors for those forms (section 4.10), but will also tend to produce duplicate submissions and invite abuses, as the "Be Counted" program will (section 4.3). Whether internet submissions are to be merged with mail-back responses and follow-up before ACE is not clear. It seems likely that internet submissions are most likely to come from middle-income and more affluent households. This could increase the differential undercount in the census headcount.

4.8 Master Address File

The Master Address File (MAF) is the frame for the census. It is the list of housing units to which the Census Bureau will deliver forms (by mail or otherwise), and perform follow-up. The Census Bureau evaluated the MAF as of 1 April 1998, using a two-stage cluster sample and independent address listings. At the divisional level (the U.S. is comprised of 9 divisions), they found that the MAF was missing about 5-16% of addreses, and that about 8-17% of the addresses in the MAF were erroneous [5]. In late 1997 the Census Bureau abandoned its plan to rely on the Postal Service and its own 1990 address file [7, 42]; this change to the method of compiling the MAF occurred too late to test in the Dress Rehearsal.

An accurate, complete MAF is crucial to obtaining an accurate count:

> ... the Census Bureau must identify all living quarters and locate these living quarters ...
>
> For Census 2000 to be as accurate, complete, and cost effective as possible, the address list that serves as the basic control for the census must be as accurate and complete as possible. If an address is not on the list, then its residents are less likely to be included in the census. [7].

The state of the MAF as of April 1998 seems worse than it was before the 1990 Census, which could decrease the accuracy of the 2000 Census relative to 1990; the Census Bureau worked to improve the MAF before Census Day and might have succeeded.

An inaccurate MAF affects census headcounts directly, but does not directly impact the error of census adjustments. If census accuracy is compromised sufficiently by a poor MAF, ACE is more likely to help.

4.9 Movers

Some people move between census day and the time the ACE P-sample is taken, which complicates census adjustment substantially. As mentioned above, the number of movers during the four months allocated for ACE is about 7% of the population. If someone living in a P-sample block moves out after census day (an "out-mover"), she will not be found by ACE at her census-day address. Such a person will be concluded to be an erroneous enumeration, unless determined from proxy data to be an out-mover. Weighted up to the nation, each false erroneous enumeration can have a large impact on the estimated undercount.

Someone who moves into a block between census day and the time of the ACE P-sample (an "in-mover") will be inferred to be a gross omission unless it is known that the person did not live in the area on census day. Such a failure to match can inflate the undercount enormously. For example, a single unmatched family of five in the 1990 PES P-sample contributed 45,000 to the undercount estimate.

In-movers are inherently easier to identify and interview than out-movers, because they are present in the P-sample blocks at the time of the PES. Getting information about out-movers relies much more heavily on proxy information, which tends to be less accurate. ACE assumes that in every P-sample block, the number of out-movers is equal to the number of in-movers ([25], p. 7).

Even though SNRFU has been abandoned [9, 39], the Bureau plans to use "Procedure C" to account for movers ([25], pp. 7, 13–14) rather than using "Procedure B," which was used in the 1990 PES [30, 48]. Procedure C requires that the post-enumeration interview identify for each address in a P-sample block (a) all current residents, and (b) all former residents who moved out since census day. In an earlier version of the ACE plan, information about individual movers would be obtained by proxy interviews, but for entire households that move, the plan called for tracing the household to its new address and obtaining information by CATI [48]. In the latest version of the 2000 ACE plan, it seems that out-movers will not be traced: all information about out-movers will be obtained by proxy ([25], pp. 13–4; [10], pp. 76–7). The P-sample is matched against the census for residents as of census day, and undercount rates are determined separately for out-movers and non-movers. In-movers are not tracked specifically. The undercount rate estimated for out-movers is applied to in-movers as well.

Procedure C is likely to have more data quality problems than Procedure B, because Procedure C relies more on proxy information. In Procedure B, the in-movers are interviewed. In-movers' records are traced back to their census-day addresses, which are obtained by interviewing the in-movers themselves. In contrast, with Procedure C, information about outmovers comes from proxy interviews. The assumption that in-movers and out-movers have similar undercount rates is questionable, especially for neighborhoods that are changing rapidly [4].

Using Procedure C for 2000 is likely to reduce the accuracy of ACE compared with the 1990 PES because data quality is likely to be worse, and because it applies adjustments derived for out-movers to in-movers, implicitly assuming that they are commensurable. Moreover, using Procedure C involves assigning a "residence probability" to P-sample records with inadequate address information [6, 32]; see section 4.12.

4.10 OCR

The plan for Census 2000 calls for optical character recognition (OCR) to scan handwritten responses from mail-back forms and, apparently, "Be Counted" forms [7, 8, 9]. Handwriting OCR is quite difficult, and has a high error rate (see below). In preliminary tests at census data centers, approximately 86% of scanned forms had machine-detected OCR errors in at least one field, and approximately 20% of fields had at least one machine-detected error (W. Eddy, personal communication, 3 February 2000). Uncorrected OCR errors will make it harder to match records between the census and the survey, increasing the rate of false non-matches.

Typically, addresses are bar-coded on mail-back forms, but not on "Be Counted" forms. Names are among the written responses in both cases. Names, captured by OCR, are key to the 2000 record matching procedure; indeed, records for which the names are not defined will not be eligible for matching ([10], pp. 29–31). "Unduplication" of multiple responses (a mail-back form plus one or more "Be Counted" forms from the same household) will be complicated by OCR errors. The recent success of hand-held handwriting recognition devices depends on "stroke recognition:" real-time tracking of the order and direction of the strokes (R. Fateman, personal communication, 1999). Information about the time-order of strokes is lost on paper forms. For hand-printed upper case characters in boxes, with the boxes printed in a color of ink that is invisible to the scanner, the error rate of the best handwriting recognition algorithms is currently 2-3%. For digits, the figure is on the order of 0.5%. For mixed upper and lower case letters and digits, the error rate is 8% or more (Y. LeCun, personal communication; see [33]). Moreover, the internally generated measures of uncertainty in character recognition do not correlate well with the error rate, so the software does not tend to ask for operator intervention when it is most likely to be in error. Assuming that errors in scanning different characters in a document are independent, the chance that an address with 25 characters and digits would contain at least one OCR error is about 88%. Even at the extremely optimistic error rate of 0.5% per character (appropriate for characters that are known to be digits), the chance that an address with 25 characters and digits would contain at least one OCR error is about 12%. Of course, the chance of OCR error depends on the document—some people have handwriting that is hard for OCR, some forms are damaged or mutilated, *etc.*

Many handwriting recognition algorithms produce weighted graphs corresponding to different interpretations of the writing, and scores for those interpretations; the final interpretation is the highest scoring path through the graph. If the entire interpretation graph for each document were kept and used for record matching in ACE, the accuracy of matching could be higher than the raw accuracy of handwriting recognition (Y. LeCun, personal communication, 1999). The 2000 plan [7, 8] seems to indicate that only the final interpretation will be stored and used to match records, so many spurious non-matches are likely—producing spurious erroneous enumerations and spurious gross omissions in ACE.

For ACE, the accuracy of names and addresses is critical. It is important for names and addresses to match between the census response and the P-sample; otherwise, ACE generates spurious gross omissions and spurious erroneous enumerations. OCR is likely to increase the number of matching errors. OCR is unlikely to be a serious problem for the headcount data, because the accuracy of names does not affect the accuracy of the count, and addresses will be bar-coded on most forms. (OCR might contribute to geocoding error in Be Counted forms, and it is important for the census to scan the number of people residing in the household correctly, *etc.*)

4.11 Post-stratification

The proposed number of post-strata for 2000 ICM was about 12,600 [7, 48]:

$$50 \text{ states } \times 6 \text{ race/origin } \times 7 \text{ age/sex } \times 2 \text{ tenure } \times 3 \text{ place type.} \tag{4}$$

	High Return Rate				Low Return Rate			
	Northeast	Midwest	South	West	Northeast	Midwest	South	West
Non-Hispanic White& "Some other race"								
Owner								
Large MSA MO/MB	1	2	3	4	5	6	7	8
Medium MSA MO/MB	9	10	11	12	13	14	15	16
Small MSA&Non-MSA MO/MB	17	18	19	20	21	22	23	24
All other TEAs	25	26	27	28	29	30	31	32
Non-Owner								
Large MSA MO/MB		33				34		
Medium MSA MO/MB		35				36		
Small MSA&Non-MSA MO/MB		37				38		
All other TEAs		39				40		
Non-Hispanic Black								
Owner								
Large& Medium MSA MO/MB		41				42		
Small MSA&Non-MSA MO/MB, & Other		43				44		
Non-Owner								
Large& Medium MSA MO/MB		45				46		
Small MSA, Non-MSA MO/MB, & Other		47				48		
Hispanic								
Owner								
Large&Medium MSA MO/MB		49				50		
Small MSA, Non-MSA MO/MB, & Other		51				52		
Non-Owner								
Large&Medium MSA MO/MB		53				54		
Small MSA, Non-MSA MO/MB, & Other		55				56		
Native Hawaiian or Pacific Islander								
Owner				57				
Non-Owner				58				
Non-Hispanic Asian								
Owner				59				
Non-Owner				60				
American Indian or Alaska Native on Reservation								
Owner				61				
Non-Owner				62				
Off-Reservation American Indian or Alaska Native								
Owner				63				
Non-Owner				64				

Table 3: Geographic and ethnic groups for post-stratifying the 2000 ACE P-sample within the 50 states. These 64 groups each are divided into seven age and gender categories to yield the post-strata; see table 5. Low return rate means 25th percentile or below, by census tract, and high return rate is all other tracts. MO/MB means "mail-out, mail-back" delivery of Census forms. Off-Reservation American Indian or Alaska Native includes persons who either (1) are in Indian Country but not on reservations who identify with a single or many races, or (2) non-Hispanic American Indians not in Indian Country who identify with only one race.

Some of the post-strata with few members would have been consolidated; the ultimate number might have been closer to 6,000. The first 2000 ICM plan used post-strata that did not cross state boundaries [7].

In the latest 2000 ACE plan, there are 448 post-strata for the U.S. and 84 post-strata for Poerto Rico. Some post-strata might be combined. Every post-stratum for the U.S. crosses state boundaries. The following tables are reproduced from [26], and are supposed to be the final plan for ACEpost-stratification. Table 3 gives the racial/ethnic/tenure/MSA/geographic/mail-back breakdown for the U.S.; table 4 gives the tenure/MSA/mail-back breakdown for Puerto Rico; table 5 gives the age/gender breakdown for the 50 states and Puerto Rico. "Tenure" denotes whether the dwelling is rented or owned by the occupant. "MSA" stands for Metropolitan Statistical Area, and "TEA" stands for type of enumeration area, such as mail-out/mail-back ("MO/MB"), update/leave, *etc.* "Return rate" is the rate at which households mail back their census forms, measured at the tract level. Low return rate tracts are those in the 25th percentile and below of mail-back return rate; high return rate tracts are the rest.

Using the mail-back return rate to define post-strata is new for 2000, and seems like an

Tenure	MSA	High Return Rate	Low Return Rate
Owner	San Juan CMSA	1	2
	Other MSA	3	4
	Non-MSA	5	6
Non-owner	San Juan CMSA	7	8
	Other MSA	9	10
	Non-MSA	11	12

Table 4: Geographic groups for post-stratifying the 2000 ACE P-sample in Puerto Rico. These 12 groups each are divided into seven age and gender categories to yield 84 post-strata; see table 5. Low return rate means 25th percentile of mail-back return rate and below, by census tract; high return rate means all other tracts.

Age	Male	Female
Under 18	A	
18-29	B	C
30-49	D	E
50+	F	G

Table 5: Age/Sex groups for post-stratifying the 2000 ACE P-sample. Each of the 64 racial/ethnic/regional/tenure/return-rate groups in table 3 is divided into these seven age/sex groups, giving a total of $7 \times 64 = 448$ post-strata for the U.S.; similarly, each of the 12 Tenure/MSA/return-rate groups in table 4 is divided into these seven age/sex groups, giving a total of 84 post-strata for Puerto Rico. Some post-strata might be combined.

improvement. Post-stratification has important effects on bias, primarily through heterogeneity (see, *e.g.*, [24]. The 2000 post-stratification contains geography explicitly only for non-hispanic white and other owners; for some other groups, geography enters implicitly through MSA size or rural area and mode of census form delivery. Some post-strata have no geographical information at all. For example, non-hispanic male asian homeowners age 50 and above are assumed to have the same propensity to be missed by the census, regardless of where in the U.S. they live.

Some experts believe that in the 1990 PES, there were apparently large differences between the response rates for children under age 10 by gender [12]. However, the 2000 post-stratification assumes that the response rates of children under age 18 do not depend on gender. Intuition suggests that there is a particularly strong tendency to count college students (mostly age 18-22) living away from home in the wrong place—their parents' address rather than their college address. If so, grouping ages 18-29 would increase heterogeneity bias.

In 1990, geographic heterogeneity was significant [4, 20, 29]; the 2000 ICM plan, with state-specific post-strata, might have had less heterogeneity bias than the current 2000 ACE plan. On the other hand, the increased number of post-strata in the ICM plan probably would have increased sampling error (see section 4.13); to combat the increase in sampling error, the results might have been smoothed, which can introduce additional biases [22].

4.12 Record Matching

As described in section 3.2, adjustment figures are extremely sensitive to details of the matching procedure. There are several changes to the matching procedure between 1990 PES and 2000 ACE.

Correct Location. In the 1990 PES, an E-sample person was deemed to be at the correct address if he was found in the P-sample within the search region around that block (one ring of adjacent blocks in urban areas, two rings in rural mail-out/mail-back areas, and the entire address register area in list/enumerate areas). A P-sample record was deemed to match an E-sample record if the records matched according to various identity criteria and the P-sample record was found in the search region around the E-sample address. The notion of the "correct location" of a person is more restrictive for 2000 ACE than it was for the 1990 PES.

> The person is correctly included in the census if he/she is included at the location where the person believes, at the time of PES interview, to have been his/her usual residence as of April 1. ([31], p. 9)

This choice seems to be biased against matches ([10], p. 100), which would increase the estimates of gross omissions and erroneous enumerations.

Minimum Information. The minimum amount of information needed about a P-sample or E-sample record to attempt matching is different for 2000. For a P-sample or E-sample record to be eligible for matching and follow-up in 2000, it must contain

- two letters in the first name and two letters in the last name, or
- first initial, middle initial, and two letters in the last name

and at least two of the following:

- relationship to head of household
- sex
- race
- Hispanic origin
- age or year of birth

See [10] pp. 30–2 for more detail. This is more than was required in 1990 to attempt a match, and could increase the fraction of records whose match status is unresolved, increasing the sensitivity of the adjustment to the model for "match probability." See below.

According to the current ACE plan ([10], pp. 30–3), if a person in a sample block filled out his census form completely, but omitted his name, and gave ACE complete information, his census record would not be eligible for follow-up and matching, but his ACE record would. ACE would find him to be a gross omission (unless his P-sample record were incorrectly matched to someone else's census record). Because his census record was incomplete, it would have unresolved match status, and a model would be used to assign it a match probability. The match probability would be less than one, so the census record would generate a fraction of an erroneous enumeration. Net, the person would contribute to the ACE undercount estimate. In 1990, names were not required for records to be eligible for matching ([10], p. 41), so this sort of error was less likely.

The same sort of error occurs whenever an individual's census record has insufficient information for matching but the individual's survey record has sufficient information for matching. If the survey tends to be more successful than the census in obtaining the minimal amount of information needed for matching, this will bias the ACE undercount estimate upward.

Search Region. As noted in section 3.2, address geocoding errors caused the match rate in 1990 to depend strongly on the choice to search a ring of one or two blocks around the target address; using a smaller search area would have increased the undercount estimate by about a factor of two, and most of the undercount estimate might have evaporated had the search region been expanded to a ring of 6-8 blocks ([3], p. 473). As of 4 February 2000, the planned search area for 2000 ACE is much smaller than for the 1990 PES: only the block of residence is searched, unless there is no match in that block for the entire housing unit. In that case, the immediately adjacent blocks (a single ring of blocks) is searched, but only on a sample basis: the search is not performed for every such household, only those in a subset of ACE blocks. The new search rules are called "targeted extended search," (TES) described in detail in [35]; see also [31, 11, 10]. With TES, the search will be performed whole-household mismatches within a subset of sample blocks, selected as follows ([35], pp. 4–5):

- Certainty sample
  - 5% of block clusters found to have the most census geocoding errors and address non-matches with the independent address listing

  - 5% of block clusters with the most weighted census geocoding errors and address mis-matches with ACE
  - all block clusters relisted in the P-sample.

- Systematic sample
  - A systematic sample from those remaining block clusters that have one or more detected census geocoding errors or one or more census addresses that do not match the independent listing. The number of such block clusters will not be determined ahead of time; it will be chosen so that the total number of block clusters in the certainty sample and systematic sample comprise 20% of the ACE block clusters.

These rules do not apply to list/enumerate block clusters in ACE, which will be handled by procedures that are not yet specified. TES has not been tested with the new procedure (Procedure C) for treating movers [35].

As noted above, match rates are quite sensitive to search area, because of geocoding errors. Using a smaller search area will probably inflate the number of false non-matches, increasing the rate of spurious erroneous enumerations and spurious gross omissions compared with 1990. Also, for 2000 ACE there is a double extrapolation from the TES blocks to the ACE blocks, then from the ACE blocks to the nation ([35], pp. 7–8); this will tend to increase error.

Matching Personnel. Match-status determinations depend on the expertise of the matching personnel: as cited in section 3.2, the discordance of match status assignments by production personnel and experts was about 1.8% in 1990, about 8% for non-matches, 0.5% for matches, and 23.8% for unresolved cases (see [14, 3]). The clerical staff for the 2000 matching might be better trained than the production matchers of 1990 [31], which could reduce the rate of false non-matches. Even if false non-matches are balanced by false matches nationally, the geographical distribution of the false non-matches and false non-matches is unlikely to be identical, which leads to geographical bias in the adjustment.

In 1990, the matching was done at seven Census Bureau centers; in 2000, the plan had called for matching to be done partly on-site during the ACE interviews by field personnel [7], and partly by clerical staff in a single Census Bureau center. In the latest operational plan [31, 10], matching will be performed in a single Census Bureau center. The 2000 ACE plan has greater reliance on computer matching than did the 1990 PES [31]. Only records from those sample block-clusters expected to benefit from clerical matching will be sent for clerical matching; records from some block clusters will be sent directly for follow-up interviews without clerical matching. The criteria for skipping clerical review are as follows ([10], p. 12):

- For all urban and rural block clusters
  1. E-sample nonmatches, but no P-sample nonmatches; all P-sample addresses match a census address
  2. P-sample nonmatches, but no E-sample nonmatches; all census addresses match a P-sample address

- For all urban areas, and only those rural areas with 100% address matches between the E-sample and P-sample.
  1. All the following conditions hold:
     (a) the sum of the numbers of unresolved matches, unmatched P-sample addresses, and unmatched E-sample matches is at most fifteen
     (b) the sum of the numbers of unmatched P-sample addresses and unmatched E-sample addresses is greater than zero
     (c) if the number of unmatched P-sample addresses is less than six, it cannot equal the number of unmatched E-sample addresses
  2. Alternatively, both of the following conditions hold:
     (a) the absolute difference between the numbers of unmatched P-sample addresses and E-sample addresses is at least eleven
     (b) the sum of the numbers of unresolved cases, unmatched P-sample addresses, and unmatched E-sample addresses, minus the absolute difference between the numbers of unmatched P-sample addresses and E-sample addresses, is at most fifteen

Match probability modeling. As discussed in section 3.2, the 1990 undercount estimate depended crucially on a model for imputing the "match probability" of cases with unresolved match status. The 1990 model used hierarchical logistic regression; the model currently planned to be used for the 2000 ACE is simpler; however, the choice of models is not yet final, except that movers and non-movers will be treated separately ([6], pp. 8; [32], p. 10). Regardless of the structural form of the model, of necessity the parameters in the model must be estimated from cases where the match status is resolved, and those cases are likely to differ in important ways from the unresolved cases; see [46] for discussion.

The current ACE plan calls for grouping people into 22 "cells" with similar characteristics, measured by racial and operational variables ([32], p. 11). The variables include the number of imputed characteristics for the individual, whether the individual is a non-Hispanic white, whether only part of the household failed to match, *etc.* ([32], pp. 10–12). The match probability of a record with unresolved match status is estimated by the rate of matches in the cell to which the individual belongs ([6], p. 8) among cases with resolved match status. The final choice of cells is quite different from the cells used in the dress rehearsal ([32], p. 11).

New for 2000, a "probability of residence" will be assigned to some P-sample individuals whose residence status in the block cluster cannot be resolved ([6], p. FIX ME!). The model divides such individuals into 32 groups and estimates the probability of residence by the fraction of individuals in the group with resolved residence who were found to be residents ([32], pp. 8–9).

Overall, the rate of false non-matches is likely to be higher in the 2000 ACE than in the 1990 PES, primarily because of the smaller search region and the more restrictive definition of "correct address."

4.13 Sample size

The 2000 ICM P-sample was planned to be about 1.7 million people in 750,000 households in 60,000 block groups [7]. The ACE P-sample is planned to comprise about 300,000 housing units [9, 39, 31]; the 1990 PES P-sample was about 166,000 housing units.

The ACE P-sample is roughly twice as large as the 1990 sample. It is difficult to assess the effect of the increased sample size, because it interacts with the choice of post-strata and other aspects of modeling and data processing. All other things being equal, increasing the sample size would decrease sampling error. However, (1) bias, not sampling error, was the larger problem in 1990, and (2) everything else is not equal. The 1990 PES average number of data per adjustment parameter was $380{,}000/1{,}392 \approx 273$. The 2000 ACE plan calls for at most 448 post-strata for the U.S. and 84 for Puerto Rico [26]; no smoothing is planned.

4.14 Timing

The most recent 2000 ACE operational plan allows approximately four months for census follow-up and sample data collection, with follow-up interviews continuing through late November 2000 for individuals, and housing-unit followup continuing through mid May, 2001 ([10], pp. 105–6). (The announced timing of the release of population estimates seems inconsistent with the end of follow-up: the Director of the Bureau of the Census stated that the final adjusted figures would be released by April, 2001 [39], but according to [10], follow-up will not be complete by then.) Some operations are allocated less time than in 1990, and some are allocated more time.

Generally, collecting data closer to Census Day is preferable, because memories are fresher and fewer people will have moved. However, if data collection is rushed, the error rate and fabrication rate are likely to increase, and there might not be enough time for adequate quality control. If collecting the data quickly requires using poorly trained staff, data quality will suffer. ACE operations begin before census follow-up is finished ([31], p. 6) which could exacerbate correlation bias.

5 Conclusions

Open Issues. As mentioned in section 4.6, it is crucial that the Census Bureau perform evaluation studies to measure components of the errors in the adjustment. This is especially important because of the large number of substantive changes from the 1990 plan. I am not aware of any planned evaluation studies.

A number of potentially important details have not yet been decided, including how demographic analysis population estimates will be compared with census and ACE estimates, given the new racial categories; and which data will be released to the public, and when.

Headcounts. The accuracy of 2000 headcount data is likely to be worse than that of 1990, in part because of accuracy and completeness of the Master Address File [5]. The address errors in the early mailings (a "1" was pre-pended to millions of addresses) indicates that logistic problems might be worse than in 1990. Paid advertising seems to have stemmed the trend towards lower mail-back response rates (as of this writing, about 66% of households have mailed back their census forms, more than were expected). The mail-back response rate affects both cost and accuracy. The use of multiple modes of data submission (mail-back, "Be Counted," and internet), followed by computer unduplication [7, 9], seems likely to increase the rate of erroneous enumerations unevenly. The new multi-racial categories will complicate comparing census counts with demographic analysis population estimates.

Adjustments. The 2000 ACE is likely to be less accurate than the 1990 PES. The principal causes of the expected decrease in accuracy are

1. the smaller geographic search region used to match records [11, 35], and searching on only a sample basis (targeted extended search, [35]), will tend to increase the rate of false non-matches and increase sampling error
2. using out-mover characteristics rather than in-mover characteristics increases the reliance on proxy interviews [48, 25], which degrades data quality
3. the post-stratification [26] could increase heterogeneity
4. errors in optical character recognition, which will be used to capture handwritten information [8, 10], will tend to inflate the apparent rate of non-matches
5. the "Be Counted" program and internet response submission, and the reliance on "computer unduplication" [7, 8, 9, 31] are likely to increase the rate of erroneous enumerations
6. requiring names for records to be eligible for matching [10] could increase the reliance on models for "match probability" and "residence probability" [6, 32] and could add bias if the survey is more successful than the census in collecting the minimum information needed for matching
7. the more restrictive definition of "correct address" [31, 35] could inflate the apparent rate of non-matches
8. imputing the "residence probability" of some P-sample records [6, 32]
9. the earlier start of ACE data collection [10, 31] increases interference between the census and ACE

Some factors will tend to improve the accuracy of ACE versus the 1990 PES:

1. the use of better-trained clerical staff for record matching [31]
2. collecting sample data closer to census day [10]
3. increasing the sample size, from approximately 166,000 housing units in 1990 PES to about 300,000 housing units in 2000 ACE ([10], p. 4)
4. the simpler model for imputing match probability [6, 32]

5. including mail-back response rate in the post-stratification [26]

On balance, systematic errors are likely to be larger than in 1990, decreasing the accuracy of the adjustment.

6 Acknowledgments.

I am grateful to D.A. Freedman (Berkeley) for numerous helpful conversations and comments on earlier drafts, and to R. Fateman (Berkeley) and Y. LeCun (ATT) for helpful conversations. This work was supported in part by the William H. Donner Foundation, and was performed in part while the author was on appointment as Miller Research Professor in the Miller Institute for Basic Research in Science.

References

[1] T.R. Belin, G.J. Diffendal, S. Mack, D.R. Rubin, J.L. Schafer, and A.M. Zaslavksy. Hierarchical logistic regression models for imputation of unresolved enumeration status in undercount estimation. *J. Am. Stat. Assoc.*, 88:1149–1166, 1993.

[2] W.R. Bell. Using information from demographic analysis in post-enumeration survey estimation. *J. Am. Stat. Assoc.*, 88:1106–1118, 1993.

[3] L. Breiman. The 1991 Census adjustment: Undercount or bad data? *Stat. Sci.*, 9:458–537, 1994.

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

[5] J.A. Burcham and D.F. Barrett. Assessing the quality of the initial master address file for Census 2000. 1999. Paper presented at the 1999 Joint Statistical Meetings, Baltimore, MD.

[6] Patrick Cantwell. Accuracy and coverage evaluation survey: Missing data procedures. Technical Report DSSD Census 2000 Procedures and Operations Memorandum Series #Q-19, U.S. Department of Commerce, Economics and Statistics Administration, Bureau of the Census, Washington, D.C., January 2000. Memorandum From Donna Kostanich to Howard Hogan; presented to the February 2-3, 2000 DSE Workshop of the National Academy of Science Panel to Review the 2000 Census.

[7] *Census 2000 Operational Plan.* U.S. Department of Commerce, Economics and Statistics Administration, Bureau of the Census, Washington, D.C., November 1998. Revised.

[8] *Census 2000 Operational Plan: Using Traditional Census-Taking Methods.* U.S. Department of Commerce, Economics and Statistics Administration, Bureau of the Census, Washington, D.C., January 1999.

[9] *Updated Summary: Census 2000 Operational Plan.* U.S. Department of Commerce, Economics and Statistics Administration, Bureau of the Census, Washington, D.C., February 1999.

[10] Danny R. Childers. The design of the census 2000 accuracy and coverage evaluation. Technical Report DSSD Census 2000 Procedures and Operations Memorandum Series, Chapter S-DT-1, U.S. Department of Commerce, Economics and Statistics Administration, Bureau of the Census, Washington, D.C., January 2000. Memorandum to Magdalena Ramos; presented to the February 2-3, 2000 DSE Workshop of the National Academy of Science Panel to Review the 2000 Census.

[11] Danny R. Childers and Debbie Fenstermaker. The design of the census 2000 accuracy and coverage evaluation. Technical Report DSSD Census 2000 Procedures and Operations Memorandum Series, Chapter S-DT-2, U.S. Department of Commerce, Economics and Statistics Administration, Bureau of the Census, Washington, D.C., January 2000. Memorandum to Howard Hogan; presented to the February 2-3, 2000 DSE Workshop of the National Academy of Science Panel to Review the 2000 Census.

[12] K. Darga. Quantifying measurement error and bias in the 1990 undercount estimates. `www.stat.berkeley.edu/~stark/census/camel.pdf`, 1998.

[13] K. Darga. Sampling and the Census: Scientific folly and scientific sense (manuscript), 1999.

[14] M.C. Davis and P. Biemer. 1990 Post-Enumeration Survey Evaluation Project P7 Series #H-2, Estimates of P-sample clerical matching error from a rematching evaluation. Technical report, U.S. Department of Commerce, Economics and Statistics Administration, Bureau of the Census, 1991.

[15] B. Edmonston and C. Schultze, editors. *Modernizing the U.S. Census.* National Academy Press, Washington, D.C., 1995.

[16] S.E. Fienberg. Bibliography on capture-recapture modelling with application to census undercount adjustment. *Survey Methodology*, 18:143–154, 1992.

[17] D. Freedman. Adjusting the 1990 Census. *Science*, 252:1233–1236, 1991.

[18] D. Freedman. Adjusting the Census of 1990. *Jurimetrics*, 34:99–106, 1993.

[19] D. Freedman and W.C. Navidi. Should we have adjusted the U.S. Census of 1980? *Survey Methodology*, 18:3–74, 1992.

[20] D. Freedman and K. Wachter. Heterogeneity and Census adjustment for the intercensal base. *Stat. Sci.*, 9:476–485, 527–537, 1994.

[21] D. Freedman and K. Wachter. Rejoinder. *Stat. Sci.*, 9:527–537, 1994.

[22] D. Freedman, K. Wachter, D. Coster, R. Cutler, and S. Klein. Adjusting the Census of 1990: The smoothing model. *Evaluation Review*, 17:371–443, 1993.

[23] D. Freedman, K. Wachter, R. Cutler, and S. Klein. Adjusting the Census of 1990: Loss functions. *Evaluation Review*, 18:243–280, 1994.

[24] D.A. Freedman, P.B. Stark, and K.W. Wachter. A new model for census adjustment. Technical Report 557, Dept. Statistics, Univ. of Calif., Berkeley, 2000.

[25] Richard Griffin. Accuracy and coverage evaluation survey: Dual system estimation. Technical Report DSSD Census 2000 Procedures and Operations Memorandum Series #Q-20, U.S. Department of Commerce, Economics and Statistics Administration, Bureau of the Census, Washington, D.C., January 2000. Memorandum from Donna Kostanich to Howard Hogan; presented to the February 2-3, 2000 DSE Workshop of the National Academy of Science Panel to Review the 2000 Census.

[26] Richard Griffin and Dawn Haines. Accuracy and coverage evaluation survey: Final post-stratification plan for dual system estimation. Technical Report DSSD Census 2000 Procedures and Operations Memorandum Series #Q-24, U.S. Department of Commerce, Economics and Statistics Administration, Bureau of the Census, Washington, D.C., April 2000. Memorandum from Donna Kostanich to Howard Hogan.

[27] Richard Griffin and Dawn Haines. Accuracy and coverage evaluation survey: Post-stratification for dual system estimation. Technical Report DSSD Census 2000 Procedures and Operations Memorandum Series #Q-21, U.S. Department of Commerce, Economics and Statistics Administration, Bureau of the Census, Washington, D.C., January 2000. Memorandum from Donna Kostanich to Howard Hogan; presented to the February 2-3, 2000 DSE Workshop of the National Academy of Science Panel to Review the 2000 Census.

[28] Dawn Haines. Accuracy and coverage evaluation survey: Synthetic estimation. Technical Report DSSD Census 2000 Procedures and Operations Memorandum Series #Q-22, U.S. Department of Commerce, Economics and Statistics Administration, Bureau of the Census, Washington, D.C., January 2000. Memorandum from Donna Kostanich to Howard Hogan; presented to the February 2-3, 2000 DSE Workshop of the National Academy of Science Panel to Review the 2000 Census.

[29] N. Hengartner and T.P. Speed. Assessing between-block heterogeneity within post-strata of the 1990 post-enumeration survey. *J. Am. Stat. Assoc.*, 88:1119–1129, 1993.

[30] H. Hogan. The 1990 post-enumeration survey: Operations and results. *J. Am. Stat. Assoc.*, 88:1047–1060, 1993.

[31] Howard Hogan. Accuracy and coverage evaluation: Theory and application. Technical report, U.S. Department of Commerce, Economics and Statistics Administration, Bureau of the Census, February 2000. Prepared for the February 2-3, 2000 DSE Workshop of the National Academy of Science Panel to Review the 2000 Census.

[32] Michael Ikeda. Accuracy and coverage evaluation survey: Specifications for the missing data procedures. Technical Report DSSD Census 2000 Procedures and Operations Memorandum #Q-25, U.S. Department of Commerce, Economics and Statistics Administration, Bureau of the Census, Washington, D.C., April 2000. Memorandum from Patrick J. Cantwell through Donna Kostanich.

[33] Y. Le Cun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proc. IEEE*, 86:2278–2324, 1998.

[34] M. Mulry and B. Spencer. Accuracy of the 1990 census and undercount adjustments. *J. Am. Stat. Assoc.*, 88:1080–1091, 1993.

[35] A. Navarro. Accuracy and coverage evaluation survey: Targeted extended search methodology. Technical Report DSSD Census 2000 Procedures and Operations Memorandum Series #Q-18, U.S. Department of Commerce, Economics and Statistics Administration, Bureau of the Census, Washington, D.C., 2000.

[36] Committee on Adjustment of Postcensal Estimates. Asessment of accuracy of adjusted versus unadjusted 1990 Census base for use in intercensal estimates (C.A.P.E. Report). Technical report, U.S. Department of Commerce, Economics and Statistics Administration, Bureau of the Census, 1992.

[37] C.G.J. Petersen. The yearly immigration of young plaice into Limfjord from the German sea. *Report of the Danish Biological Station*, 6:1–48, 1896.

[38] K. Prewitt. Census 2000 – Science meets politics. *Science*, 283:935, 1999.

[39] K. Prewitt. Statement, National Press Club, 24 February, 1999.

[40] J.G. Robinson, B. Ahmed, P. Das Gupta, and K.A. Woodrow. Estimation of population coverage in the 1990 United States Census based on demographic analysis. *J. Am. Stat. Assoc.*, 88:1061–1079, 1993.

[41] Eric Schindler. Accuracy and coverage evaluation survey: Post-stratification preliminary research results. Technical Report DSSD Census 2000 Procedures and Operations Memorandum Series #Q-23, U.S. Department of Commerce, Economics and Statistics Administration, Bureau of the Census, Washington, D.C., January 2000. Memorandum from Donna Kostanich to Howard Hogan; presented to the February 2-3, 2000 DSE Workshop of the National Academy of Science Panel to Review the 2000 Census.

[42] L.N. Stevens. Testimony before the Subcommittee on the Census, Committee on Government Reform and Oversight, House of Representatives, DECENNIAL CENSUS: Preparations for Dress Rehearsal Underscore the Challenges for 2000. Technical Report GAO/T-GGD-98-84, U.S. General Accounting Office, Washington, D.C., 1998.

[43] U.S. Department of Commerce, Economics and Statistics Administration, Bureau of the Census. Decision of the Director of the Bureau of the Census on whether to use information from the 1990 Post-Enumeration Survey (PES) to adjust the base for the intercensal population estimates produced by the Bureau of the Census. ACTION: Notice of final decision. *Federal Register*, 58 FR 69, 1993.

[44] K. Wachter. The Census adjustment trial: An exchange. *Jurimetrics*, 34:107–115, 1993.

[45] K.W. Wachter. Recommendations on 1990 Census adjustment, report to the Secretary of Commerce as a member of Special Advisory Panel. Technical report, U.S. Department of Commerce, 1991.

[46] K.W. Wachter. Comment: Ignoring nonignorable effects. *J. Am. Stat. Assoc.*, 88:1161–1163, 1993.

[47] K.W. Wachter and D.A. Freedman. The fifth cell: Correlation bias in U.S. census adjustment. *Evaluation Review*, 24:191–211, 2000.

[48] P.J. Waite and H. Hogan. Statistical methodologies for Census 2000. Manuscript, U.S. Bureau of the Census, 1998.

[49] K.K. West. 1990 Post-Enumeration Survey Evaluation Project P6 Series #G-2, fabrication in the P-sample: interviewer effect. Technical report, U.S. Department of Commerce, Economics and Statistics Administration, Bureau of the Census, 1991.

[50] A.A. White and K.F. Rust, editors. *Preparing for the 2000 Census*. National Academy Press, Washington, D.C., 1997.

[51] K. Wolter. Some coverage error models for census data. *J. Am. Stat. Assoc.*, 81:338–346, 1986.