A Method for Deciding Whether Adjustment of Census 2000 Improves Redistricting

Bruce D. Spencer[1]

January 21, 2001

Abstract

The Census Bureau must decide by April 1, 2001 whether adjustment for census undercount or overcount improves accuracy. Statistical issues related to this decision are discussed. The primary criterion should be whether adjustment improves accuracy of population numbers for areas the size of Congressional districts, i.e., areas of about 600,000 population. If adjustment does not improve the accuracy of population estimates for such areas, unadjusted numbers should be used. A method for evaluating accuracy is discussed.

1. Introduction

> Representatives . . . shall be apportioned among the several States which may be included within this Union, according to their respective Numbers. . . The actual Enumeration shall be made within three Years after the first Meeting of the Congress of the United States, and within every subsequent Term of ten Years, in such Manner as they shall by Law direct. (Constitution of the United States, Article I, Sec. 2.)

> Every decennial census, from 1790 to 1990, has included in the census count some who should have been excluded and has excluded some who should have been included. . . Every census for which the effect of these errors has been systematically measured has shown that there is a net undercount – that is, the number of residents who were missed was greater than the number of erroneous enumerations. (Census Bureau 2000)

[1]Faculty Fellow, Institute for Policy Research, and Professor and Chair, Department of Statistics, 2006 Sheridan Road, Evanston IL 60208.

file: wp-00-27.wpd

The law requires the Census Bureau to release census-population numbers by April 1, 2001 for states to use in redrawing legislative district boundaries. An important political as well as statistical question is whether the Bureau's numbers ought to reflect adjustments for undercount and overcount in the census. (For brevity, we will refer to the net effect as net undercount, as overcount is numerically equivalent to negative undercount.) Over the last two decades the Bureau has refined its ability to use sampling to estimate the net undercount in the census and to apply adjustments to the census "head counts" for each small area (such as city block). The Bureau expects the adjustment to improve accuracy. It is premature to say now, at the time of this writing, whether adjusted or unadjusted estimates will be more accurate, however, and prior to April the Bureau will conduct an analysis of their accuracy. The Bureau will not release adjusted numbers if its analysis shows that adjustment worsens accuracy. In this article, I review some issues and make some recommendations concerning how such an analysis should be conducted.

Population numbers have many uses, and the measure of accuracy should reflect the most important uses (Spencer 1986). It is possible for the adjusted numbers to be more important for some uses and for the unadjusted numbers to be more important for others. For example, there was little doubt ten years ago that the adjusted numbers for the total (i.e., national) 1990 population were more accurate than the unadjusted 1990 census totals, despite controversy over the relative accuracy of estimates for states and sub-state areas. This implies that the choice of accuracy criterion may be an important part of the decision-making.

The question of which uses are most important is ultimately a policy question, and the answer may reflect the values that the decision-maker is representing. In our democracy, the most important uses of census numbers are to determine the constitutionally mandated apportionment of the House of Representatives and to determine the allocation of the Electoral College. The Supreme Court's 1999 ruling in *Glavin, Barr, et al. v. Clinton, et al.* prohibits use of sampling-based numbers for those purposes, however. I believe that the most important use of the census population numbers to be released in April 2001 is redistricting – the drawing of new within-state boundaries for legislative districts. Indeed, the reason that the Census Bureau has a deadline of producing the numbers by April 1 is to conform with the states' timetables for redrawing congressional district and other legislative boundaries. Most other uses of population data can employ other estimates produced by the Census Bureau, such as postcensal estimates, which account for change since the 2000 census. (The census data will be a year out of date by the time the states begin to use it for redistricting.) Although the Census Bureau has expressed concern for additional important uses (Census Bureau 2000), redistricting is so central to our democracy that it dominates other data uses in importance. If adjustment were known to lead to more accurate redistricting, surely the Census Bureau would be remiss not to release adjusted numbers. On the other hand, if adjustment were known to increase the error in redistricting, would the nation be comfortable with a decision to release adjusted numbers? Such policy questions are complex and may have to be resolved in the courts, but my policy recommendation is that the Census Bureau's decision to adjust should depend on the improvement of redistricting.

Two kinds of districts depend on the population numbers: congressional districts and state legislative districts. In accordance with the "one person, one vote" principle, the Supreme Court ruled in *Wesberry v. Sanders* (1964) that the Congressional districts should be "as nearly as is practicable" equal in size and that "absolute equality" in size is the "paramount objective". In contrast, for state legislative redistricting "the overriding objective must be **substantial** equality of population among the various districts" (*Reynolds v. Sims* 1964, emphasis added).

Absolute equality in state legislative district sizes is still desirable, but moderate inequality in sizes may be acceptable if it advances a “rational state policy” (Reynolds v. Sims 1964). See National Conference of State Legislatures (1999, chapter 3) for an informative review.

It appears, then, that the key consideration is equality in Congressional district sizes, with equality of state legislative sizes a relevant but less critical issue. The primary question is thus: *Do adjusted estimates or unadjusted estimates of census population lead to Congressional districts that are more equal in size within a state?* If (but only if) the answer to that question is that the unadjusted and adjusted estimates lead to similar distributions of deviations among the sizes of Congressional districts, a secondary question becomes key: *Do adjusted estimates or unadjusted estimates of census population lead to state legislative districts that are substantially more equal in size within a state?* For the secondary question to be determinative, several conditions must hold: (i) the first question must be answerable, (ii) the answer must be that the two sets of estimates perform equally in terms of deviations of population sizes of Congressional districts, (iii) the second question must be answerable, and (iv) the answer must be that one of the sets of estimates must perform better in terms of deviations of population sizes of state legislative districts. In this article, we will focus on the primary question.

2. Accuracy Criterion for Redistricting

Congressional districts are to be constructed so that each of the n_i districts in any state i has the same estimated fraction or share of the state population, $1/n_i$. The courts have relied on various summary statistics to quantify inequality in district sizes (National Conference of State Legislatures 1999, chapter 3). The statistics cited most prominently are (i) a version of the range that is proportional to the difference between the maximum share of the population held by a district and the minimum share held by a district and (ii) the so-called mean deviation, which is proportional to average absolute value of the differences between district shares of population and $1/n_i$. Ideally, the *true* district shares would all be equal within each state. The court is aware that the measures of inequality have been applied to the estimated populations of the districts rather than the "real" sizes, and that the two differ for reasons of postcensal change as well as census data error (*Karcher v. Daggett* 1983).

The preferred means for measuring inequality in true population sizes is to estimate expected values of the measures of inequality, with expectation referring to the distribution of error in the population estimates. Unfortunately, the expected values of the two inequality measures discussed above, the range and the so-called mean deviation, cannot be estimated very well in practice. The estimate of the expected value of the range is too sensitive to the shape of the underlying probability distribution(s) to be trustworthy, and estimates of mean absolute error are known to be troublesome in the census context (Fay 1992, Mulry and Spencer 1993 section 4). What we need is a measure of inequality that is both informative and that can be estimated in practice.

One widely used measure of inequality that can be estimated in practice is the average within-state variance among the true population sizes of the districts. Specifically, let n_i denote the number of Congressional districts seats for state i, $1 \le i \le 51$, let $n = \sum_{i=1}^{51} n_i$ denote the total number of Congressional districts in the states, let P_{ij} denote the population of Congressional district j in state i, let $P_i = \sum_{j=1}^{n_i} P_{ij}$, let $\overline{P}_i = P_i/n_i$, let $\theta_{ij} = P_{ij}/P_i$, and let $\overline{\theta}_i = \sum_{i=1}^{n_i} \theta_{ij}/n_i = n_i^{-1}$. The within-state variance among the true population shares of the districts is $n_i^{-1}\sum_{j=1}^{n_i} (P_{ij} - \overline{P}_i)^2$, which is equal to $n_i^{-1}P_i^2\sum_{j=1}^{n_i} (\theta_{ij} - \overline{\theta}_i)^2$. This variance should be viewed as a descriptive statistic for the population in the same way that the variance of the income distribution is a descriptive statistic. The definition of the variance here does not rely on a notion of randomness. The number of districts in states with more than one district is fairly proportional to the size of the state population, with average Congressional district sizes for such states ranging only from about 500,000 to 620,000. The weighted average of the within-state variances, with weights proportional to the number of districts, is $n^{-1}\sum_{i=1}^{51}\sum_{j=1}^{n_i}(P_{ij}-\overline{P}_i)^2$, or equivalently

$$n^{-1}\sum_{i=1}^{51} P_i^2 \sum_{j=1}^{n_i} (\theta_{ij}-\overline{\theta}_i)^2. \qquad (1)$$

States with only one district will contribute 0 to (1), and other states will contribute as many terms as they have Congressional districts.

The within-state variance among the true sizes is proportional to the average squared error in the estimated shares of the districts; see Appendix A. Specifically, let X_{ij} denote the estimated population share of Congressional district j in state i, let $X_i = \sum_{j=1}^{n_i} X_{ij}$ and let $\overline{X}_i = \sum_{j=1}^{n_i} X_{ij}/n_i = n_i^{-1}$. The average variance given by (1) is equal to a weighted sum of squared errors in the estimated shares of the districts, namely

$$n^{-1}\sum_{i=1}^{51} P_i^2 \sum_{j=1}^{n_i} (X_{ij}-\theta_{ij})^2. \qquad (2)$$

The primary question is thus equivalent to: Is the weighted squared error in the estimated shares of districts (2) smaller when the district boundaries derive from adjusted estimates or unadjusted estimates?

The question cannot be answered prior to April 1, 2001 because the district boundaries will be unknown. Thus, it is necessary to predict the accuracy in two sets of estimates for areas whose boundaries are unknown. One approach to this problem would be to use all available information, both statistical and political, to develop probability distributions for the district boundaries. This does not seem advisable, for considerations of feasibility and of statistical policy. First, there is too much uncertainty concerning how district boundaries will be drawn for the approach to be feasible. Second, it would be catastrophic statistical policy to put a government statistical agency in the business of predicting the political consequences from its estimates in order to choose which set of estimates to produce.

A way to circumvent these problems is to examine the accuracy of the estimates for areas that are about the same size as a Congressional district but whose boundaries are known. We will call such areas "*pseudo-districts*". Accuracy of population estimates for areas typically depends on the size of the area, which is the reason that evaluations of sub-state estimates report accuracy for areas grouped by population size (Census Bureau 1992a, 1992b; National Academy of Sciences 1980). Typically, estimates are more accurate (in percentage terms) for larger areas. One choice for "pseudo-districts" is the set of current Congressional districts, which were based on unadjusted numbers from the 1990 census. Since there is a fair amount of stability in Congressional districts over time, the current Congressional districts would be good choices for pseudo-districts, particularly in states that did not gain or lose a House seat after Census 2000-based reapportionment. An alternative choice for pseudo-districts is areas of about 600,000 population, the approximate average size of the new districts. Such areas could be formed randomly (subject to compactness constraints), and the estimates of accuracy could be calculated multiple times. The variability of the estimates of accuracy would provide information about the dependency of the estimates of relative accuracy on the construction of the pseudo-districts. This choice of pseudo-districts is unrealistic, however, in that districts are not formed randomly. Thus, I recommend that the districts formed after the 1990 census be used as pseudo-districts.

A measure of improvement from adjustment may be defined as the expected value of (2) for unadjusted estimates minus the expected value of (2) for adjusted estimates, when the units are pseudo-districts. We consider the *expected* squared error in order to account for random error in the undercount adjustments and to account for biases, such as error from imperfect adjustment for missing data, that are modeled as random effects. A simplification that will be useful in practice and should not materially affect the results is to replace the P_i^2 factor in (2) by an estimate, $\tilde{P}_i^2$. Note that whether the estimate $\tilde{P}_i$ in (3) is adjusted or unadjusted should not make a appreciable difference. We denote the measure of improvement from adjustment by Δ, defined as

$$\Delta = n^{-1} \sum_{i=1}^{51} \tilde{P}_i^2 \sum_{j=1}^{n_i} (E_{unadjusted}[(X_{ij} - \theta_{ij})^2] - E_{adjusted}[(X_{ij} - \theta_{ij})^2]). \qquad (3)$$

The notation $E_{unadjusted}$ and $E_{adjusted}$ refer to expected values under the unadjusted and adjusted methodologies. I propose that Δ be used to quantify improvement from adjustment and that adjustment should be judged more accurate if and only if $\Delta > 0$.

Evaluation data can be used to estimate the expected values in (3) and produce an estimate of Δ, say $\hat{\Delta}$ (Mulry and Spencer 1993). Section 4, below, provides more details. One can then use the evaluation data to formally judge adjustment as improving the accuracy of redistricting if, for some threshold τ, $\hat{\Delta} > \tau$. The role of τ is analogous to a presumption of innocence in a trial, except here the presumption is of greater accuracy. For the adjustment decision for the 1990 census, the Department of Commerce, having announced its strong prior expectation that adjustment would not improve accuracy of population estimates, set $\tau > 0$ (Fay 1991). The effect was to increase the burden of proof on adjustment to overturn the presumption that adjustment does not improve accuracy. Now, a decade later, the Census Bureau (2000) expects adjustment to improve accuracy for Census 2000. Should τ now be negative? Not in my view: a non-zero value of τ is unjustified now and was unjustified for 1990. Not everyone agrees with the Census Bureau's expectation today or in 1990. Given the lack of consensus on what null hypothesis or prior expectations to use, and without assigning differential costs to adjusting or not adjusting when the opposite would be more accurate, it is best to take a neutral approach and, without casting either hypothesis as the null, set $\tau = 0$. This effectively sidesteps the issues arising from a forced choice of null hypothesis. In this case, the evaluation data would *formally* dictate the decision for Census 2000. Depending on the outcome of that decision, different decision criteria might apply for adjusting the 2010 census. (In addition, other, less formally stated considerations, could also enter into the decision process.)

There is a regrettable set of circumstances in which the formal outcome of the evaluation would not dictate the decision. One would not simply follow the prescription, adjust if $\hat{\Delta} > 0$ and do not adjust if $\hat{\Delta} < 0$, if there is too much uncertainty concerning $\hat{\Delta}$, for example $\hat{\Delta}$ has too much variance or demographic analysis or other evaluation results suggest severe problems with one estimator or the other that are not reflected in the estimation of $\hat{\Delta}$. In such cases, high level professional judgement should be exercised rather than a default acceptance of a null hypothesis.

3. Undercount Adjustments and Their Errors

The Census Bureau constructs adjustment factors applicable to individual enumerations. The census count for any arbitrary group or area is a sum of census counts of individuals (i.e., individual enumerations) in the group or area, and the adjusted estimate for the group or area is the sum of the adjustment factors corresponding to the census enumerations. To account for the fact that different types of people have different rates of being missed in the census, the Bureau calculates a separate adjustment factor for each of four hundred or so “post-strata”, which cross-classify census enumerations by age, sex, race, Hispanic origin, tenure (owner/renter), region, metropolitan statistical area size, type of enumeration area, and tract-level mail return rate. Each factor may be interpreted as the erroneous enumeration rate for the post-stratum divided by the probability that a person in the post-stratum was enumerated in the census. To estimate the erroneous enumeration rate, the Census Bureau selects a sample of census enumerations and conducts interviews to learn whether those enumerations were valid or not. To estimate the probability of enumeration, the Bureau attempts to interview a random sample of housing units and obtain data that will allow the Bureau to classify each person in the sampled housing units as enumerated or not. The classification of persons as enumerated or not, and the classifications of enumerations as erroneous or not, depend on matching of interview records with census forms. The proportions of matches are used to estimate the fraction of persons in each post-stratum in the sample that were enumerated in the census as well as the erroneous enumeration rate in each post-stratum. The sample survey providing the data is called the Accuracy and Coverage Evaluation (A.C.E.) and the estimator is called a dual system estimator (DSE) (Census Bureau 2000, Hogan 1993). The 1990 version of the A. C. E. was called the Post-Enumeration Survey (P.E.S.).

The DSE is subject to a variety of errors (Census Bureau 2000, Anderson et al 2000, Brown et al 2000, Mulry and Spencer 1993).

- Sampling error gives rise to random error, quantified by sampling variance, and to a systematic error known as ratio-estimator bias, which arises because even if X and Y are unbiased estimators, X/Y typically is biased.

- Data errors, including reporting errors and processing errors, such as false matches between census enumerations and interview records, also affect the DSE. False matches bias the estimate of erroneous enumerations downward and bias the estimate of census misses downward.
- Another source of error is missing data for match classification – some A.C.E. interviews fail to take place, some households provide incomplete data on questionnaire items, and in some cases the information for classification as a match or nonmatch is ambiguous. Such cases are called “unresolved matches”, and a match status is imputed. Methods are used to compensate for missing data, but they effectively assume that the match status for the case with missing data is equal on average to the status for cases that are similar except that they have complete data (Belin et al 1993). If the missing-data cases have more errors in enumeration (either non-enumeration or erroneous enumeration) than complete-data cases, the imputation method will tend to overstate the match rate, biasing upward both the numerator and denominator of the adjustment factor.
- Correlation bias, which likely pushes the DSE to understate the population size, arises from a lack of independence between enumeration in the census and in the A.C.E., e.g., people less likely to be enumerated in the census may also be less likely to be found in the A.C.E. (Kadane, Meyer,and Tukey 1999, Bell 1993, Wolter 1986).
- The DSE relies on a method called synthetic estimation to provide the same adjustment factors for all enumerations in a given post-stratum, regardless of whether the enumerations are from the same geographic area. Synthetic estimation bias arises when

enumerations from different areas should have different adjustment factors. (Census Bureau 2000, Anderson et al 2000, Brown et al 2000, Freedman and Wachter 1994, Fay and Thompson 1993).

Of the various errors, sampling error and data error are most easily estimated (Census Bureau 2000, Anderson et al 2000, Brown et al 2000, Mulry and Spencer 1993).

The Census Bureau analyzed the accuracy of the 1990 population numbers and found $\hat{\Delta}$ to be positive for (i) places with 100,000 or more population, and (ii) counties with 200,000 or more population (Census Bureau 1992a, 1992b). Their analysis found that the improvement in accuracy was positively related to the size of the place, and it is reasonable to infer that the improvement from adjustment was even stronger for areas with about 600,000 population. It is reasonable to conclude that *if the relative accuracy of the adjusted and unadjusted estimates for 2000 is similar to the Bureau's estimate for 1990, then adjusting Census 2000 will improve accuracy of redistricting.*

4. Estimation of Improvement from Adjustment, Δ

The Census Bureau considered the following sources of error (as well as others deemed negligible) in the adjustment of the 1990 census for the postcensal base (Census Bureau 1992a, 1992b).

- Sampling variance and ratio estimator bias were estimated by jackknife methods.
- Bias from data collection and data processing operations (e.g., bias from false non-matches or false matches) was estimated from data in evaluation studies based on

subsamples of the 1990 P.E.S. Despite disagreements over estimates of data bias (Anderson et al 2000, Brown et al 2000, Belin and Rolph 1994, Breiman 1994), this source of bias is more easily estimated than the others.

- The bias from missing data can in principle be estimated from intensive followup of cases with missing data, but in practice the fraction completed by followup is too low. Although one can consider the range of effects on the DSE by considering extreme alternatives – e.g., all unresolved matches truly are matches or truly are non-matches – the range is too wide to be informative about the likely bias. As noted earlier, the missing data might well cause an upward bias in both the numerator and denominator of the adjustment factor. The Census Bureau analyzed the missing-data bias by looking at the changes in the DSE when alternative methods were used to compensate for missing data. The Bureau modeled the bias as a random effect whose variance could be estimated from the changes observed in the DSE when alternative imputation methods were applied. Missing data error was believed to be small in 1990, especially compared to correlation bias (Census Bureau 2000, Anderson et al 2000, Mulry and Spencer 1993).
- Evidence of correlation bias in national estimates is provided by sex ratios (males to females) for adjusted numbers that are low relative to ratios derived from demographic analysis of data on births, deaths, and migration (Robinson, Ahmed, DasGupta, Woodrow 1994). The demographic analysis is unreliable at the subnational level, however, and allocation of correlation bias sub-nationally has relied upon alternative parametric statistical models (Bell 1993, Wolter 1990, 1986).
- To assess synthetic estimation bias for a given area one needs to develop an estimate based on data from the area alone, which is rarely possible. Attempts to estimate

synthetic estimation bias in undercount estimates from analysis of "surrogate" variables, whose geographic distributions are known, are unconvincing (Freedman and Wachter 1994, Ericksen, Fienberg, and Kadane 1994, Fay and Thompson 1993). The Census Bureau did not include an explicit error component for synthetic estimation bias.

Consider now the use of the evaluations of error components to estimate Δ. Let E(.) and V(.) denote expectation and variance, respectively. Let U_{ij} denote the effect of undercount (for population total or share) for area ij (i.e., pseudo-district j in state i) and let its estimate, the difference between the DSE and the census count, be denoted by $\hat{U}_{ij}$. Jackknife methods are used to estimate the variance $V_{ij} = V(\hat{U}_{ij})$ by $\hat{V}_{ij}$. Write the expected value of the undercount estimate as $E(\hat{U}_{ij}) = U_{ij} + B_{ij}$. Denote the estimate of B_{ij} by $\hat{B}_{ij}$. The Census Bureau constructs $\hat{B}_{ij}$ by analyzing the propagation of the diverse sources of error in the DSE; see Mulry and Spencer (1993, 1991) for details. The biases were estimated directly for groups of poststrata and then were distributed among individual poststrata under assumptions of proportionality of errors to the DSE ("GRODSE") or proportionality to the absolute value of the net undercount ("GROSUC"); see Mulry (1992) for more details. It is reasonable to assume that $\hat{B}_{ij}$ is biased, say $E(\hat{B}_{ij}) = B_{ij} - \beta_{ij}$, and the sampling designs are such that the correlation between $\hat{B}_{ij}$ and $\hat{U}_{ij}$ is negligible. (We discuss the bias β_{ij} in section 5.) The measure of improvement of area ij from adjustment is $\Delta_{ij} = \tilde{P}_i^2(U_{ij}^2 - (B_{ij}^2 + V_{ij}))$, and the estimate of Δ_{ij} is taken to be $\hat{\Delta}_{ij} = \tilde{P}_i^2((\hat{U}_{ij} - \hat{B}_{ij})^2 - \hat{B}_{ij}^2 - 2\hat{V}_{ij})$. The within-state average of Δ_{ij} is $\Delta_i = n_i^{-1}\sum_{j=1}^{n_i} \Delta_{ij}$ and the weighted average across states is $\Delta = \sum_{i=1}^{51} (n_i/n)\Delta_i$. Similarly, the estimate for state i is $\hat{\Delta}_i = n_i^{-1}\sum_{j=1}^{n_i} \hat{\Delta}_{ij}$ and the weighted average across states is $\hat{\Delta} = \sum_{i=1}^{51} (n_i/n)\hat{\Delta}_i$.

5. Accuracy of Estimates of Improvement

The estimates $\hat{\Delta}$ are subject to random error and to bias. The random error could shift $\hat{\Delta}$ up or down. (The usual justification for taking $\tau \neq 0$ is protection against such random error.) Overstatement of the variance of the DSE shifts $\hat{\Delta}$ down, favoring non-adjustment, and conversely for understatement of variance. The sampling variance of the DSE will be approximately unbiasedly estimated for the A.C.E. The variance of the DSE may be overstated, however, by the treatment of missing data bias as a random effect.

The estimates $\hat{\Delta}$ will be biased if there is unmeasured bias, denoted by β_{ij}, in the estimates of undercount. Unmeasured bias will tend to lead to overstatement of the accuracy of the adjusted estimates, except to the extent that the unmeasured bias cancels out other biases. At the same time, the unmeasured bias affects the estimates of accuracy of the unadjusted estimates. To understand the net effect on the *comparative* accuracy, it is useful to observe (see Appendix B) that the bias in $\hat{\Delta}_i$ is

$$E(\hat{\Delta}_i - \Delta_i) = 2n_i^{-1}\tilde{P}_i^2 \sum_{j=1}^{n_i} \beta_{ij} E(\hat{U}_{ij}), \qquad (4)$$

which is proportional to the cross-area correlation between β_{ij} and $E(\hat{U}_{ij})$ in state i, and that the bias overall is

$$E(\hat{\Delta} - \Delta) = 2n^{-1} \sum_{i=1}^{51} \tilde{P}_i^2 \sum_{j=1}^{n_i} \beta_{ij} E(\hat{U}_{ij}). \qquad (5)$$

Concerns have been expressed that in the context of the 1990 census the Bureau may have underestimated the magnitudes of both correlation bias and bias from imperfect adjustment for missing data and that it has not quantified synthetic estimation bias; see Brown et al (2000) for critique and Anderson et al (2000) for rebuttal. To understand the implications of these concerns for bias in $\hat{\Delta}$ when population shares are being estimated, think of unmeasured bias as the sum of components due to correlation bias, synthetic estimation, missing data.

- The Census Bureau (1992a, 1992b) observed in its analyses of accuracy that if no adjustment were made for correlation bias, the unadjusted census fared better in accuracy comparisons than when Bell's (1993) adjustments were made. Although the Bureau did not use the same loss function as (3), the results suggest negative correlations between $E(\hat{U}_{ij})$ and estimates of correlation bias. If the unmeasured correlation bias is small relative to the measured correlation bias, the empirical finding suggests a negative correlation between $E(\hat{U}_{ij})$ and unmeasured correlation bias in state i, which would lead to a component of bias in $\hat{\Delta}_i$ (and hence in $\hat{\Delta}$) favoring non-adjustment.
- It is logically possible for synthetic estimation bias to cause either a positive or a negative bias in $\hat{\Delta}_i$ (and in $\hat{\Delta}$), depending on the sign of cross-area correlation between $E(\hat{U}_{ij})$ and the unmeasured bias from synthetic estimation in state i. The Census Bureau (1992b) discussed some analyses indicating that the effect of ignoring synthetic estimation bias is quite possibly to overstate the accuracy of both the census and the DSE and to favor the census relative to the DSE. The Census Bureau's analyses were based on modeling for artificial populations, however, and thus are not definitive.
- Missing data bias was treated as a variance component, rather than as a bias. Inclusion of extra variance causes an downward bias in $\hat{\Delta}$, thus favoring non-adjustment. Whether this is more than balanced out by the effect of the unmeasured bias from missing data

depends on the correlations between $E(\hat{U}_{ij})$ and the actual missing-data bias.

One approach to estimating the effect of unmeasured bias is to speculate about the unmeasured bias directly, as in artificial population analysis for synthetic estimation error. We may write $\beta_{ij} = \alpha_i E(\hat{B}_{ij}) + \epsilon_{ij}$, and unless $\sum_j E(\hat{B}_{ij}) = 0$ but $\sum_j \beta_{ij} \neq 0$, we may assume that $\sum_j \epsilon_{ij} = 0$. It follows from the definitions that

$$\sum_j (E\hat{U}_{ij})\beta_{ij} = \alpha_i \sum_{ij} (E\hat{U}_{ij})(E\hat{B}_{ij}) + \sum_j (E\hat{U}_{ij})\epsilon_{ij}. \tag{6}$$

For a simple example, suppose that the unmeasured biases are nearly proportional to the measured biases. Then the ϵ_{ij}'s will be small and hence the last term on the right in (6) will be small, and thus the bias in $\hat{\Delta} - \Delta$ will be approximately equal to $2n^{-1}\sum_i \alpha_i \tilde{P}_i^2 \sum_j (E\hat{U}_{ij})(E\hat{B}_{ij})$. For further simplicity, suppose the α_i's are all equal to a common value α. The data used in Census (1992a, 1992b) were reanalyzed for Congressional districts based on the 1990 census, and $\hat{\Delta}$ and $\tilde{P}_i^2 \sum_j \hat{U}_{ij} \hat{B}_{ij}$ were calculated as estimates of Δ and $\tilde{P}_i^2 \sum_j (E\hat{U}_j)(E\hat{B}_j)$. Unadjusted census counts were used to calculate $\tilde{P}_i$. The estimates of Δ were 2.7×10^{13} under either the GRODSE or GROSUC method for estimating B, whereas the estimates of approximate bias in $\hat{\Delta}$ were much smaller, -8.3×10^6 and -8.6×10^6, respectively, times α. (These calculations are based on data from Starsinic (2000). Under this simplified model, the effect of unmeasured bias on the estimate of improvement from adjustment would be minor unless α were huge (i.e., the measured bias was incredibly small compared to the unmeasured bias).

Further work could be done to illuminate the effect of unmeasured bias on the estimates of improvement from adjustment. The simple model (6) could be modified, for example, by (i) decomposing β into a sum of components reflecting synthetic estimation bias, missing data bias, and correlation bias, (ii) modeling $\sum_j (E\hat{U}_{ij})\beta_{component,i}$ for each component, and (iii) summing over the components.

An alternative approach is to consider models for the error in the DSE. A very simple example is the following. Let C_{ij} denote the census count, let D_{ij} denote the DSE, and as before let P_{ij} denote the true population for area j in state i. Write the expected value of the DSE, $E(D_{ij})$ as satisfying $E(D_{ij}) - C_{ij} = \lambda_{ij}(P_{ij} - C_{ij})$. For insight, consider the case when all λ_{ij} are equal to λ; although equality is undoubtedly a gross oversimplification, data (Bell 1993, Table 2) are compatible with a range $.4 < \lambda < 2.0$. Then (see Appendix C) we have the approximation

$$\frac{1}{n}\sum_{i=1}^{51} \tilde{P}_i^2 \sum_j (E\hat{U}_{ij})\beta_{ij} \approx (1 - \lambda^{-1})\frac{1}{n}\sum_{i=1}^{51} \tilde{P}_i^2 \sum_j (E\hat{U}_{ij})^2 - \frac{1}{n}\sum_{i=1}^{51} \tilde{P}_i^2 \sum_j (E\hat{U}_{ij})(E\hat{B}_{ij}). \qquad (7)$$

As noted earlier, the estimate of the second term on the right in (7) is about -8.3×10^6 or -8.6×10^6 for 1990 data. The data used in Census (1992a, 1992b) indicate a value of 3.0×10^7 for $n^{-1}\sum_{i=1}^{51} \tilde{P}_i^2 \sum_j (E\hat{U}_{ij})^2$ and hence the first term on the right in (7) is in the range -1.5×10^7 to 0.5×10^7. The sign of the expression given by (7) is could be either positive or negative, depending on which value of λ is chosen, and thus the model does not tell us in this case whether the bias favors adjustment or non-adjustment. The magnitude is only about 10^7, however, which is trivial compared to the magnitude of the estimate of Δ, 10^{13}. Although the model leading to (7) is overly simple, Appendix C provides a generalization of (7) allowing for unequal λ_{ij}'s.

The numerical results, interpreted in light of (6) and (7), suggest that unmeasured bias did not appreciably bias the estimate of Δ. The results are only suggestive, however, and not conclusive, as their interpretations rest on highly simplified assumptions. Unmeasured bias will inevitably be a source of uncertainty in the estimate of Δ. Unmeasured bias is a third-order error

– the error in estimate of error in estimate of error in the census – and it is elusive. Developing more realistic models for unmeasured bias is an interesting topic for future research. Unless there is compelling evidence that the effect of unmeasured bias is large enough to invalidate the estimate of Δ, however, $\hat{\Delta}$ should be relied upon.

6. Real-Time Assessment of Accuracy

The calculation of $\hat{\Delta}$ requires estimates of variances and biases of the DSEs. Unfortunately, the full set of evaluations of the A.C.E. and Census 2000 are not scheduled to be completed until 2003. Thus, the *real-tim*e estimation of accuracy – to be completed before April 1, 2001 – must use some rough approximations to develop $\hat{\Delta}$. A reasonable strategy is to use the method for estimation of Δ in 1990 and modify parts of it to reflect known differences between the DSEs for 1990 and 2000, such as sampling variances, missing-data rates, and sex-ratios. In addition to $\hat{\Delta}$, the decision should be based on indicators of data quality that may not be fully taken into account in the formal estimation of Δ. As the discussion of unmeasured bias suggests, some high-order technical judgement will be required as well. Given the politicized nature of the decision, such uses of judgement should be transparent and as restricted as possible.

References

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**, 341-356.

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

Belin, T. R. , Diffendal, G., Mack, S., Rubin, D., Schafer, J., and Zaslavsky, A., (1993) "Hierarchical Logistic Regression Models for Imputation of Unresolved Enumeration Status in Undercount Estimation". *Journal of the American Statistical Association*, **88**, 1149-1159.

Belin, T. R. and Rolph, J. E. (1994) "Can we reach consensus on census adjustment?" *Statistical Science*, **9**, 486-508.

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

Breiman, L. (1994) "The 1991 census adjustment: Undercount or bad data," *Statistical Science*, **9**, 458-475.

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, Summer 1999.

Census Bureau (2000) *Accuracy and Coverage Evaluation. Statement on the Feasibility of Using Statistical Methods to Improve the Accuracy of Census 2000.* United States Bureau of the Census, Washington, D. C., June 2000.

Census Bureau (1992a) "Additional Research on Accuracy of Adjusted Versus Unadjusted 1990 Census Base for Use in Intercensal Estimates". Report of the Committee on Adjustment of Postcensal Estimates, Bureau of the Census, Department of Commerce, November 25, 1992.

Census Bureau (1992b) "Assessment of on Accuracy of Adjusted Versus Unadjusted 1990 Census Base for Use in Intercensal Estimates". Report of the Committee on Adjustment of Postcensal Estimates, Bureau of the Census, Department of Commerce, August 7, 1992.

Ericksen, E. P., Fienberg, S. E., and Kadane, J. B. (1994) “Comment on three papers on census adjustment,” *Statistical Science,* **9**, 511-515.

Fay, R. E. (1992) “Absolute Values in PES Loss Function Analysis”. Memorandum for the Committee on Adjustment of Postcensal Estimates (CAPE), United States Bureau of the Census, Washington, D. C., June 29, 1992.

Fay, R. E. (1991) “Hypothesis Tests for Improvement”. Memorandum for the Undercount Steering Committee, United States Bureau of the Census, Washington, D. C., June 28, 1991.

Fay, R. E. and Thompson, J. (1993) “The 1990 Post-Enumeration Survey: Statistical Lessons in Hindsight,” *Proceedings of the 1993 Annual Research Conference of the U.S. Bureau of the Census,* 71-91.

Freedman, D. and Wachter, K. (1994) “Heterogeneity and Census Adjustment for the Inter-Censal Base,” *Statistical Science,* **9**, 476-485.

Hogan, H. (1993) “The 1990 Post-Enumeration Survey: Operations and Results,” *Journal of the American Statistical Association,* **88**, 1047-1060.

Kadane, J. B. Meyer, M. M., and Tukey, J. W. (1999) “Yule's Association Paradox and Stratum Heterogeneity in Capture-Recapture Studies,” *Journal of the American Statistical Association*, **94**, 855-859.

Karcher v. Daggett (1983) 462 U.S. 725.

Mulry, M. H. (1992) “Loss Function Analysis for the Post Census Review (PCR) Estimates,” July 2, 1992. Decennial Statistical Studies Division, U. S. Bureau of the Census.

Mulry, M. H. and Spencer, B. D. (1993) “Accuracy of the 1990 Census and Undercount

Adjustments," *Journal of the American Statistical Association,* **88**, 1080-1091.

Mulry, M. H. and Spencer, B. D. (1991) "Total Error in PES Estimates of Population: The Dress Rehearsal Census of 1988". *Journal of the American Statistical Association*, **86**: 839-854 (with discussion 855-863).

National Academy of Sciences (1980) *Estimating Population and Income of Small Areas.* Report of Panel on Small Area Estimates of Population and Income. National Academy Press, Washington, D.C.

National Conference of State Legislatures (1999) *Redistricting Law 2000.* Report of the Redistricting Task Force of the National Conference of State Legislatures. Printed report completed January 1999. Online version at time of this writing at: http://www.senate.leg.state.mn.us/departments/scr/redist/Red2000/ch2equal.htm.

J. G. Robinson, B. Ahmed, P. DasGupta, and K. A. Woodrow (1993) "Estimation of population coverage in the 1990 United States Census based on demographic analysis," (with discussion) *Journal of the American Statistical Association*, **88**, 1061-1073.

Spencer, B. D. (1986) "Conceptual Issues in Measuring Improvement in Population Estimates." pp. 393-407 in *Second Annual Research Conference*. Washington, D.C.: U.S. Bureau of the Census.

Spencer, B. D. (1980) "Implications of Equity and Accuracy for Undercount Adjustment: A Decision-Theoretic Approach." Pp. 204-216 in U.S. Bureau of the Census, *Conference on Census Undercount: Proceedings of the 1980 Conference*. Washington, D.C.: U.S. Department of Commerce.

Starsinic, M. D. (2000) personal communication, 12/7/00.

Wesberry v. Sanders (1964) 376 U.S. 1

Wolter, K. (1990) "Capture-Recapture Estimation in the Presence of a Known Sex Ratio," *Biometrics,* **46**, 157-162.

Wolter, K. (1986) "Some Coverage Error Models for Census Data," *Journal of the American Statistical Association*, **81**, 338-353.

Appendix A

We prove that the variance in the true shares of the districts within a state is equal to the average squared error in the estimated share of each district for any fixed values of the true shares and the estimates. Notice that, by construction of Congressional districts, $X_{ij} = 1/n_i$. To show that (1) equals (2) it suffices to observe that

$$\begin{aligned}
\sum_j (X_{ij} - \theta_{ij})^2 &= \sum_j X_{ij}^2 - 2\sum_j X_{ij}\theta_{ij} + \sum_j \theta_{ij}^2 \\
&= n_i^{-1} - 2\bar{\theta}_i + \sum_j \theta_{ij}^2 \\
&= \sum_j \theta_{ij}^2 - n_i^{-1} \\
&= \sum_j \theta_{ij}^2 - n_i\bar{\theta}_i^2 \\
&= \sum_j (\theta_{ij} - \bar{\theta}_i)^2.
\end{aligned}$$

Appendix B

Observe that

$$\tilde{P}_i^{-2} E(\hat{\Delta}_{ij} - \Delta_{ij}) = E[(\hat{U}_{ij} - \hat{B}_{ij})^2 - \hat{B}_{ij}^2 - 2\hat{V}_{ij}] - [U_{ij}^2 - B_{ij}^2 - V_{ij}]$$

$$= [E(\hat{U}_{ij}) - E(\hat{B}_{ij})]^2 + V(\hat{U}_{ij} - \hat{B}_{ij}) - E(\hat{B}_{ij})^2 - V(\hat{B}_{ij}) - V_{ij} - U_{ij}^2 + B_{ij}^2 - V_{ij}$$

$$= (U_{ij} + \beta_{ij})^2 - (B_{ij} - \beta_{ij})^2 - U_{ij}^2 + B_{ij}^2$$

$$= 2\beta_{ij}U_{ij} + 2\beta_{ij}B_{ij}$$

$$= 2\beta_{ij}E(\hat{U}_{ij}).$$

Thus, $E(\hat{\Delta}_{ij} - \Delta_{ij}) = 2\tilde{P}_i^2 \beta_{ij} E(\hat{U}_{ij})$. It follows by definition that $E(\hat{\Delta}_i - \Delta_i) = 2n_i^{-1} \tilde{P}_i^2 \sum_{j=1}^{n_i} \beta_{ij} E(\hat{U}_{ij})$ and $E(\hat{\Delta} - \Delta) = 2n^{-1} \sum_{i=1}^{51} \tilde{P}_i^2 \sum_{j=1}^{n_i} \beta_{ij} E(\hat{U}_{ij})$.

Note: The preceding analysis applies to estimates of either population totals or shares. In the case of redistricting, $\sum_j E(\hat{U}_{ij}) = 0$, which implies that $\sum_j \beta_{ij} E(\hat{U}_{ij}) = n_i \gamma_i$ with γ_i the covariance of β_{ij} and $E(\hat{U}_{ij})$ among pseudo-districts in state i. It follows that $E(\hat{\Delta} - \Delta) = 2n^{-1} \sum_i n_i \tilde{P}_i^2 \gamma_i$.

Appendix C

Let P_{ij} denote the true population of area j within a state i, let C_{ij} denote the census count, and let D_{ij} denote the DSE. We will express D_{ij} as

$$D_{ij} = C_{ij} + \lambda_{ij}(P_{ij} - C_{ij}) + \delta_{ij},$$

where $\lambda_{ij} \neq 0$ denotes a shrinkage factor indicating how much the expected value of D_{ij} underestimates the true number of persons undercounted and δ_{ij} is random with mean $E(\delta_{ij}) = 0$. Define $P_i = \sum_j P_{ij}$, $C_i = \sum_j C_{ij}$, $D_i = \sum_j D_{ij}$, and $\lambda_i = \sum_j \lambda_{ij}(P_{ij} - C_{ij})/(P_i - C_i)$. The true undercount in the share is $U_{ij} = P_{ij}/P_i - C_{ij}/C_i$. A first-order Taylor approximation to U_{ij} is

$$U_{ij} \approx \frac{C_{ij}}{C_i}\left(\frac{P_{ij}}{C_{ij}} - \frac{P_i}{C_i}\right).$$

The DSE estimates the population share by

$$\frac{D_{ij}}{D_i} = \frac{C_{ij} + \lambda_{ij}(P_{ij} - C_{ij}) + \delta_{ij}}{C_i + \lambda_i(P_i - C_i) + \delta_i},$$

with $\delta_i = \sum_j \delta_{ij}$. The DSE-based estimate of undercount in the population share, $\hat{U}_{ij}$, is equal to

$$\hat{U}_{ij} = \frac{D_{ij}}{D_i} - \frac{C_{ij}}{C_i} = \frac{C_{ij} + \lambda_{ij}(P_{ij} - C_{ij}) + \delta_{ij}}{C_i + \lambda_i(P_i - C_i) + \delta_i} - \frac{C_{ij}}{C_i}.$$

A first-order Taylor approximation is

$$\hat{U}_{ij} \approx \left(\frac{C_{ij}}{C_i}\right)\left[\frac{\lambda_{ij}(P_{ij} - C_{ij}) + \delta_{ij}}{C_{ij}} - \frac{\lambda_i(P_i - C_i) + \delta_i}{C_i}\right]$$

$$= \left(\frac{C_{ij}}{C_i}\right)\left[\lambda_{ij}\left(\frac{P_{ij}}{C_{ij}} - \frac{P_i}{C_i}\right) + (\lambda_{ij} - \lambda_i)\left(\frac{P_i}{C_i} - 1\right) + \frac{\delta_{ij}}{C_{ij}} - \frac{\delta_i}{C_i}\right].$$

Taking expected values, we obtain the approximations

$$E\hat{U}_{ij} \approx \left(\frac{C_{ij}}{C_i}\right)\left[\lambda_{ij}\left(\frac{P_{ij}}{C_{ij}} - \frac{P_i}{C_i}\right) + (\lambda_{ij} - \lambda_i)\left(\frac{P_i}{C_i} - 1\right)\right]$$

and

$$B_{ij} = E\hat{U}_{ij} - U_{ij} \approx (\frac{C_{ij}}{C_i})[(\lambda_{ij} - 1)(\frac{P_{ij}}{C_{ij}} - \frac{P_i}{C_i}) + (\lambda_{ij} - \lambda_i)(\frac{P_i}{C_i} - 1)]$$

$$= (1 - \frac{1}{\lambda_{ij}})E\hat{U}_{ij} + (1 - \frac{\lambda_i}{\lambda_{ij}})(\frac{P_i}{C_i} - 1).$$

Note that $\beta_{ij} = B_{ij} - E\hat{B}_{ij}$, and so

$$\sum_i \sum_j (E\hat{U}_{ij})\beta_{ij} \approx \sum_i \sum_j (E\hat{U}_{ij})^2(1 - \frac{1}{\lambda_{ij}}) + \sum_i (\frac{P_i}{C_i} - 1)\sum_j (E\hat{U}_{ij})(1 - \frac{\lambda_i}{\lambda_{ij}}) - \sum_i \sum_j (E\hat{U}_{ij})(E\hat{B}_{ij}).$$

This is a general result that does not depend on a model for the measurement processes except to assume that $P_{ij} \neq C_{ij}$ and $\lambda_{ij} \neq 0$. The result is an approximation, however, which depends on small errors, although that the regularity conditions under which the approximations are valid have not been verified for our application. If $\lambda_{ij} = \lambda_i \neq 0$, then

$$\sum_i \sum_j (E\hat{U}_{ij})\beta_{ij} \approx \sum_i (1 - \lambda_i^{-1})\sum_j (E\hat{U}_{ij})^2 - \sum_i \sum_j (E\hat{U}_{ij})(E\hat{B}_{ij}).$$