Can We Reach Consensus on Census Adjustment?

Thomas R. Belin and John E. Rolph

Abstract. Attempting a complete headcount is an imperfect method for carrying out a census, as is modifying an attempted headcount with sample-based adjustments. It is a mistake to assume that one approach enjoys a scientific presumption over the other. There are important details available from evaluation studies of the 1990 decennial census that reflect upon the accuracy of adjusted and unadjusted census figures. Decisions about adjustment might therefore be based on comparing the accuracy of alternative census-taking strategies at some level of aggregation of the population. In any such comparison, the choices of an appropriate level of aggregation, the factors defining the aggregation, and appropriate loss criteria are important issues to decide in advance. After providing context for decisions about census-taking strategy, we comment on the recent literature on census adjustment, including the papers by Freedman and Wachter and by Breiman contained in this issue; we also discuss the Census Bureau's plans for the year 2000. We conclude that the 1990 approach to summarizing the accuracy of an adjusted census can be improved upon, but that many of the criticisms of census adjustment do not reflect a balanced decision-making perspective. We also conclude that the Census Bureau is pursuing constructive research in evaluating a "one-number census," and we suggest that statisticians have a role to play in avoiding the costly legal battles that have plagued recent censuses by assisting in the process of deciding on a design for the 2000 census.

Key words and phrases: Decision, loss function, census adjustment, one-number census.

For an effective adjustment procedure to be widely accepted, given that not all localities will benefit, it is important that there be as widespread understanding and agreement as possible within the professional community of statisticians that a general reduction in differential coverage error is sufficiently desirable to accept adverse impacts on some individual localities. —Panel on Decennial Census Methodology of the Committee on National Statistics, National Academy of Sciences, 1985.

1. THE PROSPECT OF CONSENSUS IN A CONTENTIOUS DEBATE

That the widespread agreement among statisticians envisioned by the National Academy panel does not exist is clear from the preceding papers, by Freedman and Wachter (1994; hereafter FW) and by Breiman (1994). We are far from sure that consensus is attainable on census adjustment, but we think it should be possible to settle issues that in the past have stood in the way of consensus.

Our primary goal in this article is to set forth the key substantive issues in the debate over statistical adjustment to address differential undercount in the decennial census. To sharpen our focus, we have tried to separate statistical principles from the many relevant technical details that would attend any proposed adjustment. Our hope is that by highlighting

Thomas R. Belin is Assistant Professor in Residence, Department of Biostatistics, UCLA School of Public Health, Los Angeles, California 90024-1772. John E. Rolph is Professor of Statistics, University of Southern California Business School, Los Angeles, California 90089-1421.

points of contention and contributing to a sensible framework for discussion, we might help in putting our profession's collective energies to more constructive use than having us argue about philosophical differences in a courtroom every 10 years from here on in.

Several themes permeate the adjustment debate: the interplay of personal and scientific principles, the idea that census adjustment should be viewed as a decision problem, the role of formal loss functions in such a decision, and the question of what level of aggregation is appropriate for evaluating alternative census numbers. The paper by FW discusses the impact of a particular source of uncertainty on the kinds of evaluations that might be used to decide about adjustment, and the paper by Breiman raises several provocative points in connection with the proposed adjustment of the 1990 census. We discuss the papers by FW and by Breiman with the goal of relating them to the broader context of the adjustment controversy.

Although we wish to promote consensus, not all of our comments, particularly about the latter paper, are conciliatory. Consensus-building sometimes requires criticizing ideas around which consensus will never emerge, and this we have endeavored to do.

The two of us have come at the census undercount problem from different angles. Belin was a graduate student in the Harvard Statistics Department whose thesis work on computer matching of census and post-enumeration survey (PES) data brought him to the Census Bureau, and he worked at the Bureau from mid-1990 to mid-1991 developing programs that were used for the imputation of cases with unresolved enumeration status in the 1990 PES. Rolph is presently an American Statistical Association representative to the Census Advisory Committee of Professional Associations and served as a member of a National Academy of Sciences committee in the mid-1980's that produced the report from which the opening quotation is taken; he also testified in the *New York et al. v. Commerce Department et al.* lawsuit for the plaintiffs on issues pertaining to regression smoothing techniques used in the proposed adjustment procedure.

The rest of the article is organized as follows. Section 2 reviews central elements of the adjustment controversy, with special attention to issues that involve human judgment. In Section 3, we discuss the 1990 adjustment operation, adding details not contained in the FW and Breiman papers, in order to provide the groundwork for our later comments on those papers in Sections 4 and 5. The growing support for a "one-number census" in the year 2000 is the subject of Section 6, and we offer some concluding remarks in Section 7.

2. THE ADJUSTMENT DEBATE

Scientists, politicians, government officials and members of the judiciary, among others, have articulated viewpoints on strategies for census-taking that frequently conflict. There is now a voluminous statistical literature on the subject: for example, Alho, Mulry, Wurdeman and Kim (1993); Belin et al. (1993); Bell (1993); Bryant (1991, 1992, 1993); Citro and Cohen (1985); Cohen (1990, 1991); Cressie (1992); Darroch, Fienberg, Glonek and Junker (1993); Datta et al. (1992); Ericksen and DeFonso (1993); Ericksen and Kadane (1985); Ericksen, Kadane and Tukey (1989); Ericksen, Estrada, Tukey and Wolter (1991); Fay, Passel, Robinson and Cowan (1988); Fay and Thompson (1993); Fienberg (1989–1992, 1992a,b, 1993, 1994a); Freedman (1991, 1994); Freedman and Navidi (1986, 1992); Freedman et al. (1993); Hogan (1992a, 1993); Hogan and Wolter (1988); Isaki, Schultz, Diffendal and Huang (1988); Kruskal (1991); Martin, Brownrigg and Fay (1990); Meyer and Kadane (1992); Mulry and Spencer (1991, 1993); Robinson, Ahmed, Das Gupta and Woodrow (1993); Rolph (1993); Royce (1992); Schirm (1991); Schirm and Preston (1987); Wachter (1991, 1993a, 1994); Wachter and Speed (1991); Wolter (1991); Wolter and Causey (1991); Zaslavsky (1993a); along with the documentation of the 1991 adjustment decision (Mosbacher 1991a); the transcript and exhibits from the *New York et al. v. Commerce Department et al.* lawsuit over the proposed 1990 adjustment (U.S. District Court 1992); the special sections of the *Journal of the American Statistical Association* in September 1993, of *Jurimetrics* in Fall 1993 and of *Survey Methodology* in June 1988, December 1988 and June 1992; and the recent National Research Council reports by the Panel on Census Requirements in the Year 2000 and Beyond and by the Panel to Evaluate Alternative Census Methods (Committee on National Statistics, 1993a,b). Our goal here is to provide some context for the adjustment debate and to do so in a way that makes the "consensus" in our title a plausible objective.

2.1 Accuracy and the Census

Thomas Jefferson's suspicion of an undercount in the 1790 census (Ericksen, Kadane and Tukey, 1989) has often been cited in the adjustment debate as an early sign in our history that the census needs fixing. There was a differential undercount between blacks and whites after the Civil War (Anderson, 1988), and the Census Bureau has published black–white differentials based on demographic analysis from the 1940 census onward. To obtain estimates of differential undercount broken down into subpopulations defined by geography and other characteristics, the

Census Bureau has used "dual system estimates," which are described by Breiman and to which we allude later. The "adjustment debate" surrounding the 1990 decennial census has centered on the comparison of census numbers that include a correction based on estimates of differential undercount and census numbers that contain no such correction. We now describe some of the issues at the heart of the debate.

2.2 "Personal and Political" Considerations in Loss Functions

Formal yardsticks to compare the adjusted to the unadjusted counts in the form of loss functions were used by the Census Bureau in their evaluations of 1990 census results. There has been criticism of the idea of reducing decisions about adjustment to a comparison of "loss estimates" that synthesize large quantities of data with fairly crude statistical measures. Kenneth Wachter, in his report to the Secretary of Commerce as a member of the Secretary's Special Advisory Panel on the 1990 Census, wrote the following (Wachter, 1991, page 5, emphasis in original):

> *I do not believe that any highly aggregated index or loss function is appropriate for summing up overall accuracy.* It is informative to understand how much the outcomes of calculations with different versions of such aggregated indices differ. But the choice among them is not a scientific choice. Each such index involves implicit value judgments about different sorts of error. For example, each index determines whether a few large errors are more serious than a great many smaller errors. *Whether we agree with a particular tradeoff is a matter of personal and political values. It should not be disguised as science.*

An example that illustrates what Wachter has in mind as a relevant "personal and political" consideration is his concern for future funding of censuses (Wachter, 1991, page 42, emphasis in original):

> *I believe that Census adjustment would reduce the stake that individuals, civic leaders, and Congressional representatives would have in coverage improvement efforts. If coverage improvement efforts suffer, our knowledge of the characteristics of the hard-to-count population will suffer too. Adjustment would also increase the political leverage of technical decisions. Extra efforts to guarantee the Census Bureau's independence and objectivity would be required.*

Wachter (1991, pages 42–43) further argues that the "most serious consequence" of adjustment on future census efforts is that politicians would be likely to reduce funding for data collection, notably the expensive efforts to collect information on especially hard-to-count individuals, which he describes as a "tragedy" because the information lost at the margin is "especially precious as background for policy." Although information on hard-to-count people may be precious, data from the "coverage improvement" efforts Wachter defends do not seem to have an especially prominent role in current policymaking; also, their cost-effectiveness is questionable (Ericksen and DeFonso, 1993). In any event, we do not endorse Wachter's speculative predictions about the consequences of using estimation techniques in census-taking, but we do agree that there are a variety of relevant nonstatistical issues that should be aired and discussed by all parties, statisticians and nonstatisticians alike.

One reason it has been so difficult to build consensus on census adjustment is that the "rules of the game" do not seem to be fixed. For example, Wachter (1991, page 4) commented (emphasis in original):

> I have heard versions of an *a priori* argument contending that we should always be able to do better by using all our data, combining all the information we have, than by restricting ourselves to one part or another. This is meant to be an argument that we can do better by combining the Census enumeration with the PES than by using the Census alone.... The assumptions of this argument....do not hold for the Census and the PES. In ordinary life, we all often make wrong decisions figuring in new information, when we would have been better off to stay with what we thought at first. *The question of whether counts can be improved is an empirical question, not an* a priori *one.*

When combined with his comment that "I do not believe that any highly aggregated index or loss function is appropriate..." one page later in his report, this passage presents us with a philosophical quandary. On the one hand, Wachter argues that adjustment is an empirical question and that it is only from evidence in the data that we can reliably infer that adjusted counts are more accurate. On the other, however, Wachter argues that no index for summarizing the evidence from data is an unambiguous measure of whether one estimate is better than

another. How he would have us proceed is unclear. Our view is that it would be helpful, at the very least, to agree upon what type of evidence from the data would be viewed as favorable for adjustment.

Some may accept Wachter's position on both fronts and may then reason that adjustment is philosophically untenable. We reject this argument, as it fails to heed the warning cited by several commentators on the adjustment debate (e.g., Tukey, 1991; Fienberg, 1992a), namely, "Don't let the perfect be the enemy of the good." Substantial gains in accuracy have been shown to be possible through adjustment by research such as Schirm and Preston (1987), Isaki, Schultz, Diffendal and Huang (1988) and Zaslavsky (1993a), and several investigators have found that comparisons between unadjusted and adjusted counts are apt to agree across reasonable loss function choices (see Citro and Cohen, 1985; Schirm and Preston, 1987; Mulry, 1992a; Mulry and Spencer, 1993; Zaslavsky, 1993a). The issue of how to balance evidence from the data with *a priori* considerations is admittedly difficult, but we see no reason to ignore the evidence from research on loss function analysis.

2.3 "Validation" of Adjustment Methods

In responding to the argument "Nothing is perfect, and don't let the best be the enemy of the good," Freedman and Navidi (1992) write the following:

> We think the census is imperfect, but good. We think the smoothing models are quite questionable, and the arguments to defend them are bad. Proponents of adjustment have an obligation to state their assumptions and produce data to validate them. The models don't have to hold perfectly, but departures from assumptions and their impacts need to be studied. Otherwise, the algorithms have no justification except familiarity.

Of course, the adjustment process should be fully documented with its assumptions spelled out, and indeed that happened. But did proponents of the change that was made early in 1994 in the way unemployment figures are produced (Plewes, 1994) "validate" their assumptions? What would it take to "validate" such assumptions, and who gets to decide whether the assumptions are "validated?"

Freedman offered a hint of what kind of validation he had in mind in some personal correspondence with Belin (Freedman, 1993a). The discussion focused on results presented in Belin et al. (1993) comparing predictions from the P-sample imputation model to results from evaluation follow-up (EFU) interviewing, an operation that in part investigated individuals whose enumeration status in the census was unresolved. We later describe the EFU operation in more detail, and we also examine the accuracy of imputation further in the context of Breiman's paper. For present purposes, it is relevant that fewer than half of the individuals whose census enumeration status was unresolved in the PES were able to be resolved through the additional review. However, among those EFU cases that were resolved, 31.6% were determined to have been enumerated in the census, and the imputation model had predicted that 32.2% of these cases had been enumerated with a predictive standard error of 2.2%. Belin (1993) asked Freedman:

> I am curious, since you seem to think that the EFU data constitute evidence against the approach we used, can you offer a hypothetical scenario in which the EFU data would come out in such a way that you would not criticize our imputation approach? I really do not mean to be facetious here; I simply do not know what might silence some of the critics on this point.

Freedman (1993a) wrote the following:

> [W]hat would have to happen in EFU to persuade me? If all or nearly all of the imputed cases in production had been in the sampling universe for EFU, and a very substantial fraction of the sampled cases had been resolved in EFU, and the match status as determined in EFU had on average matched the imputed probabilities (conditional on some covariates of interest), and EFU fieldwork had been done more or less properly, then I would be quite persuaded, indeed, this would be a very serious argument. Since none of these conditions are met, I find the assertion that EFU validates the model to be somewhat frivolous.

Freedman's position here is very extreme. With the designation of "unresolved" status coming much later than the original interview, roughly six months later in the 1990 PES, one could never design a PES–EFU program to guarantee the high response rate that Freedman demands as a precondition to "validation." There is uncertainty inherent in any adjustment operation due to the inevitable presence of unresolved cases and the inevitable difficulty of following them up. We think it makes sense to weigh this uncertainty against the foregone conclusion that the census headcount will never be perfectly calibrated in all parts of the country for all population subgroups.

It strikes us as an extreme position to be so intolerant of uncertainty from a source such as unresolved enumeration status that one would reject adjustment out of hand without considering the degree of bias evidence in the census headcount. Meanwhile, the assertion that 31.6% does not constitute close agreement with $32.2 \pm 2.2\%$ raises questions about what standard Freedman is using in his evaluation of the data.

Ultimately, the Freedman–Navidi position, "We think the census is imperfect, but good," begs the question, because the determination that the census is "good" is in the eye of the beholder. Is there no room in statistical science for those who say, "We think the census could be better"?

2.4 Producing Census Counts as a Decision Problem

The 1990 attempted headcount resulted in completely implausible census counts. So did the proposed 1990 adjustment. By "completely implausible" we mean that there is evidence in available data that suggests the presence of systematic errors in the allocation of population counts to areas. The attempted headcount was completely implausible because at aggregated levels there was a clear bias in the coverage of minority populations. The proposed adjusted counts were completely implausible because there is demonstrable heterogeneity in undercount rates within poststrata assumed to have constant within-stratum undercount rates (FW; Hengartner and Speed, 1993). That a choice has to be made between imperfect alternatives is obvious. Population counts have to be produced, so we have to choose from among "implausible sets" of counts.

Zaslavsky (1993a) has proposed what appear to be less implausible models than either of these other two, a development that we applaud. Surely Zaslavsky's models are not perfect, but violations of the assumptions of Zaslavsky's models appear to be harder to find than violations of the assumptions underlying either an unadjusted census or an adjustment based purely on dual-system estimation. In our opinion, the census would be better if it were harder than it currently is to find systematic errors in it.

2.5 Guidelines for a Decision on Adjustment

Important constraints attend this decision problem, notably time, available methodology, and the need to have the solution be broadly acceptable to a skeptical public and its elected officials. In such a decision setting, there are many valid nonstatistical considerations that touch on issues of fairness, precedent for future censuses, and so on. However, there is a need for clarification of standards or rules to guide the decision-making process.

The 1990 adjustment decision differed from that in 1980 in that the Commerce Department was required as part of settlement of a lawsuit to produce "guidelines" for its decision. These guidelines, which themselves were the target of a subsequent legal challenge, put forward some general presumptions and highlighted several issues as relevant for decision-making purposes. It bears repeating that there is an important distinction to be drawn between statistical and nonstatistical considerations. "Guidelines" as we conceive of them are primarily useful as a way of prioritizing nonstatistical issues so that statistical results can usefully inform the decision-making process. In his paper, Breiman comments as follows:

> The most important [guideline], from a statistical point of view, was Guideline 1: "The Census shall be considered the most accurate count of the population of the United States, at the national, State, and local level, unless an adjusted count is shown to be more accurate."

One might also say that Guideline 1, which we do not necessarily endorse, was the most important guideline from a political point of view, in the sense of that term used by Wachter (1991).

Freedman and Navidi (1992) advance the notion that there is a "burden of proof" that any proposed adjustment must surmount in order to be accepted, apparently viewing such a burden as something of a basic canon of science since adjustment is based on statistical models. Meanwhile, in distinguishing themselves from "modelers," Freedman and Navidi appear not to acknowledge the assumption that the attempted headcount leads to constant undercount rates across poststrata. However, this assumption is part of the implicit model that underlines the unadjusted counts. In arguing that the burden of proof is a statistical issue, Freedman and Navidi shed more light on their own biases than they do on the broader adjustment debate:

> As the exchange with Fienberg indicates, modelers are reluctant to accept the burden of proof. Once they make an assumption, it is taken as truth unless it can be disproved. Even then, they may view the assumption as useful until it can be replaced by some other assumption.
>
> Language is used in a specialized way. An assumption is "reasonable" if the modelers think it is reasonable. If questioned, they introspect again. The introspection

> confirms the original conclusion; after all, the assumptions are by now familiar parts of the technical literature. The modelers become indignant at those who do not share the faith.

As for the psychological profile of the "modeler," we think it applies equally well to those whose "faith" has it that the attempted headcount leads to constant undercount rates across poststrata. In fact, the assumption that the undercount rate is constant everywhere is stronger than the assumption that undercount rates are constant within poststrata, so perhaps the criticisms apply to a greater extent to those whose faith favors unadjusted counts. Meanwhile, it is naive to suggest that "modelers" have a monopoly on using language in a specialized way for rhetorical purposes, as our discussion below of Breiman's paper amply demonstrates. However, Freedman and Navidi are certainly correct that rhetoric has created a gulf across which it is hard to move toward consensus.

To summarize, "personal and political" opinions deserve to be considered in the adjustment debate, and developing guidelines may be a useful way to summarize nonstatistical issues, but there is not a sound basis for rejecting adjustment out of hand on scientific grounds.

2.6 What Is the Right Loss Function?

Crucial to any decision about statistical estimation in the census are the "measures of performance" (Isaki, Schultz, Diffendal and Huang, 1988) chosen to assess the accuracy of census numbers. Such measures have been variously referred to as measures of improvement (Wolter, 1987b), measures of adjustment success (Schirm and Preston, 1987) and measures of misproportionality (Tukey, 1983), but commonly are referred to using the generic term "loss functions" (e.g., Spencer, 1980a,b; Citro and Cohen, 1985; Woltman, 1991; Woltman et al., 1991; Freedman and Navidi, 1992; Mulry and Spencer, 1993; Zaslavsky, 1993a). We see a prospect for genuine consensus in this area.

We begin by recounting some statements made in the report of the Panel on Decennial Census Methodology in 1985 that we believe apply equally well today as they did back then:

1. "[I]t is impossible to determine a single loss function that is appropriate for evaluating every effect of an error in census numbers: each use of the census numbers has a different effect resulting in different components of loss" (Citro and Cohen, 1985, page 279).
2. "It must be accepted that no adjustment procedure can be expected to simultaneously reduce the error of all census information for every location in the United States" (Citro and Cohen, 1985, page 281).
3. "[A]djustment should be undertaken when there is reasonable certainty that appreciable reduction in the general differential coverage error will be achieved. A relatively trivial reduction would not be worthwhile, since adjustment will surely cost time and resources to implement, and doubt about whether the adjustment did or did not reduce the differential coverage error would impair public confidence in census figures. Furthermore, knowledge of a subsequent adjustment might reduce public cooperation, thus lowering the completeness of the census count" (Citro and Cohen, 1985, page 281).
4. "The panel believes that it is substantially more important to reduce the general error per person than the general error per place. Hence, the panel does not recommend the use of loss functions for measuring the total error that weight each political jurisdiction equally, e.g., that determine the proportion of the 39,000 revenue sharing jurisdictions that gained or loss through adjustment, regardless of the number of people in each jurisdiction. Rather, the panel believes that the contribution to total loss attributable to an area should reflect the size of its population" (Citro and Cohen, 1985, page 282).

There is widespread agreement that distributional accuracy, or focusing on population shares rather than population totals, is the appropriate goal (Schirm and Preston, 1987; Undercount Steering Committee, 1991; Mosbacher, 1991b; Mulry and Spencer, 1993; Zaslavsky, 1993a; Rolph, 1993). However, the fourth point, that it is important to consider the size of jurisdictions in assessments of accuracy as opposed to counting the number of domains made more accurate, is underappreciated. Much of politics is place-oriented, so political interest is often in error per place, not error per person. This is an issue where getting statisticians together with students of politics might be helpful in sharpening the debate. The rationale for the panel's recommendation that the focus be on reducing error per person was based on the idea that each citizen's interest should be weighted equally for fairness reasons.

This point evidently did not sink in among key decisionmakers on adjustment of the 1990 census. Then-Secretary of Commerce Robert Mosbacher, the day after he announced his decision not to adjust the 1990 census, commented in congressional testimony (Mosbacher, 1991b):

> Based on the measurements so far completed, the Census Bureau estimated that the proportional share of about 29 States would be made more accurate and about 21 States would be made less accurate by adjustment.... When the Census Bureau made allowances for plausible estimates of factors not yet measured, these comparisons shifted toward favoring the accuracy of the census enumeration. Using this test, 28 or 29 States were estimated to be made less accurate if the adjustment were to be used.

This testimony was given only a few weeks after the Census Bureau's Undercount Steering Committee, in its report, took great pains to point out that it is a fallacy to suppose that the break-even point between adjustment and nonadjustment occurs where the shares are more accurate in exactly half of the jurisdictions (Undercount Steering Committee, 1991):

> Suppose [conditions hold that imply] the census distribution [of proportionate shares for states] is a close approximation to the truth but differs randomly by about 1 percent, and the PES is unbiased but has a sampling variance of the same magnitude. Thus, the sets of two proportions are equally accurate.... Under assumptions of normality, the negative sign [for the difference between the loss for a given state under the unadjusted census versus under the adjusted census, implying an advantage to the unadjusted census] should appear in an expected 68 percent of the states, or about 34 out of 50. Intuition that the break-even point is when half of the states have negative losses and half have positive is *not* correct. [emphasis in original] Further, when the ratio of [the variance of the biases in unadjusted state population shares] to [the variance of PES estimates of state population shares] is 2 to 1, strongly favoring an adjustment, then a similar calculation gives that 59 percent would have negative signs, or about 29.5 states. The expected number of negative signs is about 21 at a ratio of [the variance of biases in unadjusted state shares to the variance of PES estimates of state shares] of 5 to 1. Under this simple model, observing 21 negative signs is consistent with a strong positive effect of adjustment on the measurement of the true population proportions.

The secretary's reliance on a summary of the results that probably had the effect of misleading the assembled legislators is troubling. It may reflect the place-oriented outlook of a political appointee who was unaware of the apparent paradox or a conscious choice to present results in a way that favored a particular interpretation. Either way, this portion of Mosbacher's rationale for not adjusting the census is weak, sounding like an argument that the presidency of the United States should be awarded to the person who wins the most number of states.

The National Academy panel in the mid-1980's (Citro and Cohen, 1985) cites earlier work by Kadane (1984) that provides a rationale for the choice of a loss function for congressional seat allocation and by Spencer (1980b) for choosing a loss function appropriate for revenue-sharing purposes. Tukey (1983) and Wolter (1987b), along with the National Academy panel, provide strong arguments about the importance of taking into account the size of units in assessing their contribution to aggregated loss. Zaslavsky (1993a) adds to these arguments a seemingly reasonable condition that reduces the field of contenders of candidate loss functions, namely, that the aggregated loss should be insensitive to division of domains of consideration if undercount is uniform within that domain. Zaslavsky accordingly restricts attention to absolute error $\sum|X_k - T_k|$, where X_k is the estimated population share for domain k and T_k is the true population share for domain k, and size-weighted squared relative error $\sum(X_k - T_k)^2/T_k$. We endorse the loss function choices in Zaslavsky (1993a) for the reasons outlined there and by earlier authors (e.g., Hartigan, 1992).

Finally, we note that the methods of evaluation in Zaslavsky (1993a) represent a substantial advance over the loss function analysis strategy implemented in recent Census Bureau decisions on adjustment. Rather than assuming that adjusted counts corrected by estimated biases can be taken as true population counts, Zaslavsky reflects the fact that the estimated biases have variances of their own that need to be included in the analysis for a proper accounting of uncertainty. This is a technical advance that could remove a major source of controversy over adjustment evaluation.

2.7 What Is the Right Level of Analysis?

Beyond the issue of choosing among loss functions, another concern is the level of geographical or other aggregation at which to apply loss functions. Freedman (1991) discusses small-sample issues that attend the problem of making adjustments at disagregated levels. Freedman uses as an example the city of Stockton, California, which has a population of about 200,000, roughly a quarter of whom are Hispanic. Stockton cannot have very

many individuals included in a PES sample of 5,000 blocks around the country, and no stable estimates of a Hispanic undercount rate would be available from data in Stockton alone. The Census Bureau's 1990 stratification pooled Stockton's Hispanics with other Hispanics in Southern California and pooled its non-Hispanics with other non-Hispanics throughout the Western states in order to estimate undercount rates. Freedman argues that "Stockton is the rule, not the exception," since there are far more local jurisdictions around the country than there were blocks in the 1990 PES; implicitly, the argument seems to be that adjustment is untenable on its face.

On the other hand, as noted by Zaslavsky (1993b), direct sample-based estimates of undercount rates for demographic subgroups at the most aggregated level, that is, for the country as a whole, are accurate enough to detect significant differences between groups. For example, no one disputes that blacks and whites are differentially undercounted. At a less aggregated level, for example, at the level of 50 states, sophisticated "composite" estimates, or weighted averages of unadjusted and adjusted counts with weights derived from estimates of the precision of the unadjusted and adjusted counts, may be shown to be more accurate than unadjusted counts (Zaslavsky, 1993a). However, there is some level of aggregation, which Zaslavsky (1993b) terms "the gray zone," where composite estimates improve accuracy, but the case is difficult to prove. Also, there is what Zaslavsky (1993b) terms "the red zone," a level of aggregation where adjustment may make estimates less accurate, although again it may be hard to prove that these small-area adjustments give less precise estimates of population shares.

Once again, "personal and political" considerations can also enter the picture, as loss functions may not reflect fairness considerations that many deem important. Cain (1992) reports a systematic effect at the level of congressional districts which raises fairness questions that would be difficult to capture in an explicit loss function analysis. In the trial on the 1990 adjustment, Cain presented a regression of estimated percent undercount on percent Democratic registration among registered voters in the 52 newly apportioned California congressional districts, which were created based on unadjusted data. He found that a difference of 10% in Democratic registration was associated with a difference of 0.8% in estimated undercount, with worse undercounts in districts that were more Democratic.

If the choice of a level of aggregation at which to evaluate census numbers is not decided in advance, it will be extremely difficult not only to build consensus around a particular set of census numbers but even to design an undercount evaluation program like the 1990 PES that would be broadly acceptable.

3. A SUMMARY OF THE ATTEMPTED HEADCOUNT AND ADJUSTMENT-RELATED OPERATIONS

Although the descriptions of the census-taking process in the papers by Freedman and Wachter and by Breiman are not incorrect, they are incomplete because they do not mention some key features of the applied setting that result in a decision-making framework where the alternatives are imperfect and there are crucial constraints, such as time. We give our own synopsis of the census-taking process to provide a context for our later comments. See Hogan (1992a, 1993), Ericksen, Estrada, Tukey and Wolter (1991) and Freedman et al. (1993) for more detailed accounts. The process that generated the official 1990 population counts we term an "attempted headcount" for reasons that the descriptions below will make clear.

3.1 Address Lists and Initial Enumeration Attempt

In 1990, in most areas the census assembled address lists from a variety of public and private sources and sent a census form to each household on the address list. In other areas, notably rural areas where addresses such as "Rural Route 1" might be common to a number of households, census enumerators attempted to visit each housing unit in person. Four different enumeration strategies were used by the Census Bureau, with all areas of the country being classified into one of these four "types of enumeration areas." Most households received one census form, but some received none and some received more than one. The official Census Day was April 1, 1990, with respondents to the census asked to list residents of the housing unit as of that day. In addition to trying to include all housing units in its address list, the Bureau attempted to enumerate homeless individuals with teams of enumerators visiting places thought "likely" to have homeless people spending the night. The 1990 PES was focused solely on the population in housing units and was not designed to measure undercount of the homeless population; in the Census Bureau jargon that Breiman invokes, the homeless population was "out-of-scope" for the PES.

3.2 Follow-up

When a household did not respond to the first attempt of the Census Bureau to count individuals living there, census enumerators sought an interview with a household member during May or June of

1990 in "nonresponse follow-up." When no household member was available, a "last-resort" interview was sought with a neighbor, landlord, or some other person who claimed knowledge about the residents of a household. When a census enumerator suspected that a housing unit was occupied but did not have direct evidence, individuals were added to households through "imputation" or "substitution," whereby the number of individuals from a nearby household is taken as the number for the household in question.

In addition, there were several "coverage improvement" programs. These include the following: counting people who might be vacationing or traveling; double-checking households that were originally classified as vacant; running newspaper advertisements asking "Were You Counted?" that people could clip and send to the Bureau; and reviewing administrative lists in an attempt to count parolees and probationers. One of these operations, the "Parolee/Probationer Follow-up Program," was not "prespecified," or planned in advance of its mid-1990 inception, but was undertaken when an originally planned "Parolee/Probationer Check" resulted in much lower than expected response. About 2.9 million people were added to the population between August and December of 1990 (Ericksen and DeFonso, 1993); these later enumerations were considerably more likely to be erroneous (duplicate, fictitious, etc.) than enumerations obtained closer to the official census day (Ericksen, Estrada, Tukey and Wolter, 1991; Belin and Diffendal, 1991; Griffin and Wajer, 1993; Ericksen and DeFonso, 1993).

3.3 Post Enumeration Survey (PES)

Before headcounting efforts began, PES workers (also employed by the Census Bureau but not involved in the attempted headcount) canvassed a sample of blocks, recording addresses and sketching maps, that were to be included in the PES based on a stratified sampling scheme considering region of the country, "place type" (i.e., central city or one of a few types of non–central city locations), 1980 percentage of minority population in the area, 1980 percentage of renters in the area, and "type of enumeration area" (i.e., census-taking methodology used in the area, reflecting the source of the household address in the master address list and the strategy for contacting the household, such as by mail or by personal interview) (Woltman, Alberti and Moriarity, 1988). The PES interviewing was done in person, mostly in July and August, or three to four months after census day. Some areas with high levels of nonresponse in the initial stages of PES interviewing were recanvassed by experienced interviewers who work regularly for the Census Bureau on its monthly household surveys.

The PES design was a stratified cluster sample, with all housing units in sampled blocks targeted for inclusion in the PES sample, except in some large blocks where subsampling of the block took place. The stratification in the design was intended to improve the precision of estimates of differential undercount between certain population subgroups, with some groups anticipated to have high undercounts sampled at a greater than proportional rate. Sampling weights based on inverse probabilities of selection of the block were spread among the individuals in the block, with an adjustment made to individual sampling weights when subsampling was done and when interviews for particular households were not obtained.

3.4 Matching the E-Sample and P-Sample

Breiman uses the Census Bureau jargon "E-sample" and "P-sample," the former consisting of the individuals living in PES blocks on April 1, 1990, and the latter consisting of the individuals living in PES blocks at the time of the PES. These samples overlap, but one reason they can differ is that people move between census day and the time of the PES.

Names of individuals from census forms, which are not routinely entered into census databases, were keypunched into an E-sample data file to supplement demographic information from the census form, and names and associated characteristics from the PES were entered into a separate P-sample data file. Name, address, and personal characteristics provide the basis for matching records from these two data files; when names are not available, matching is not possible, so undercount estimation needs to accommodate such cases. A computer matching operation identified obvious matches, after which information was printed out onto forms that were reviewed by clerical matching teams, one form for each housing unit with a line for each individual in that housing unit. Sometimes, for reasons that will become clear, this process is called before-follow-up matching; Breiman refers to these matching efforts by their Census Bureau acronyms, SMG1 and SMG2. Separate matching operations were undertaken for individuals who moved between the time of the census and the PES; for these individuals, queries in the PES interview about the address where the individual resided on April 1 were used to search census data files to see whether the person had been enumerated.

3.5 PES Follow-up

The "PES follow-up," which took place in October and November of 1990, attempted to resolve discrepancies between the census and PES rosters by send-

ing interviewers to the households in question. Both census and PES enumerators can fabricate the existence of people, and both sets of interviewers miss residents. Thus, for both E-sample cases that did not match a PES case and P-sample cases that did not match a census case, interviewers attempted to find out whether the individual lived at the sampled housing unit on census day. The rule of thumb is that two "sightings" in the field establish the existence of individuals. So, for example, individuals observed in both the census and the PES need not be followed up; in addition, a second "sighting" in PES follow-up is taken as establishing the existence of individuals who had been included in the PES but not the census.

Information from PES follow-up interviewing was returned to processing offices for "after-follow-up matching" by teams of matching clerks, who attempted to resolve individuals as having been enumerated or not having been enumerated based on the additional information from the field. Some cases inevitably remain unresolved in this process, perhaps because a follow-up interview was not obtained. Since these cases were selected for follow-up interviewing on account of a discrepancy between the census and the PES, they are clearly not a random subsample of the PES sample. Thus, ignoring unresolved cases in undercount estimation would lead to predictable biases.

3.6 Estimation for Unresolved Cases

For these remaining unresolved cases, rather than attempting further follow-up interviewing, logistic regression models based on characteristics observed on each individual were used to impute a probability of having been correctly enumerated. These probabilities were used to estimate undercount rates and to characterize uncertainty about those estimated rates (Schafer and Schenker, 1991; Belin et al., 1993). The logistic regressions used not only personal characteristics such as age, race and sex, but also processing characteristics such as "before-follow-up match status," which reflects information such as that there is a person in the census who appears to be a possible match for the given unresolved P-sample individual or that the unresolved person was part of a whole household of individuals who did not match anyone in the census even though the address where they resided was included in the census address list. As shown in Belin et al. (1993), such distinctions clearly have great predictive value.

3.7 Adjustment Poststrata

Estimation of undercount rates was done for various poststrata defined by the following: geography, place type, race or ethnic origin, owner or renter status, age, and sex. For the 1991 proposed adjustment discussed by Breiman, there were 1,392 poststrata; for the 1992 proposed adjustment of the base for intercensal estimates discussed by FW, there was a coarser stratification into 357 poststrata. Sampling variances were estimated by a jackknife method, where the units being jackknifed were the primary sampling units, namely, the sampled blocks.

3.8 Smoothing Models

For the 1991 proposed adjustment, a hierarchical "smoothing" regression model was used that related poststratum-level undercount rates to covariates. Because large estimated undercounts are typically associated with large variance estimates or, more specifically, because the error in estimating the mean for a Poisson-like undercount rate is apt to be correlated with the estimated variance, "raw" sample variance estimates were modeled to produce "presmoothed" variance estimates for input into the smoothing of undercount estimates (Hogan, 1993).

There were a few poststrata where the combination of presmoothing and smoothing led to undercount estimates that did not appear to pass a test of face-validity. For example, Asian–Pacific Islander males aged 20–29 in New York City had a direct estimate of a 34% undercount, and presmoothing approximately halved its estimated standard deviation from 14.4% to 6.2%. That poststratum became an influential point in the regression; for 12 poststrata involving Asians and Pacific Islanders in New York City, the mean $\pm$ SD of their estimated undercount rates went from $4.3 \pm 20.4\%$ before smoothing to $13.4 \pm 2.0\%$ after smoothing. The considerable shrinkage of the undercount estimates for these poststrata is a reflection of limited sample sizes, and this degree of sensitivity was unusual. To lessen the impact of influential and outlying poststrata, which were identified through analysis of regression residuals (Bell, 1991; Hogan, 1993), the variances of the undercount estimates for these poststrata were left "un-presmoothed." This yielded final undercount estimates for the 12 Asian–Pacific Islander poststrata in New York City of $9.8 \pm 1.9\%$, these being regarded as superior to the estimates with all variances presmoothed as a matter of face-validity.

Much controversy attended the use of such models in the *New York v. Commerce Department* lawsuit (Freedman, 1992; Freedman et al., 1993); the issue of smoothing was perhaps the most prominent technical issue argued there. In that context, the second author expressed strong support for such models being used as they were by the Census Bureau in their 1991 proposed adjustment (Rolph, 1992, 1993).

While far from a perfect fix to a challenging technical problem, the decision to presmooth was based on the technical issue of avoiding bias in estimating Poisson-like means. We mention this to refute the recent suggestion by a group of authors (Freedman et al., 1993) that a "circular" argument was the basis for presmoothing, or, as those authors put it, that presmoothing is "defended because it delivers answers that agree with a priori expectations [about minority undercounts]." These comments ignore the rationale for the decision four years earlier at the Census Bureau to carry out presmoothing for bias-reduction reasons.

3.9 Evaluation Studies

To inform officials better in the decision-making process, the Census Bureau attempted to synthesize components of error in undercount estimates into a "total error model" (Mulry and Spencer, 1991, 1993). Such errors were estimated by a series of projects such as a rematch study, in which additional matching clerks reviewed a sample of cases, with results later compared to those of the original PES matching clerks, and "evaluation follow-up" (EFU) interviewing, where a final attempt was made to reach individuals in the PES sample to assess undetected PES fabrication, the actual enumeration rate of individuals who were unresolved in PES processing and so on. Breiman (1994) focuses on the accuracy of these studies.

In addition, a "loss function analysis" was used to inform the adjustment decision by estimating the costs of errors associated with the attempted headcount and the adjusted counts. Here, simple functions are used to approximate the costs of errors, or losses. Comparisons of the unadjusted and adjusted counts took place at the level of a small number (13 in the 1991 loss function analysis and 10 in the 1992 loss function analysis) of groupings of PES poststrata into "evaluation poststrata" that reflect broad geographical and racial–ethnic characteristics, and comparisons also took place at more disaggregated levels such as the state level. In both the 1991 and 1992 analyses, the latter of which corrected some of the errors cited by Breiman, notably the computer processing error, the Census Bureau calculations favored the adjusted counts (Mulry and Spencer, 1993; Fay and Thompson, 1993). Not all observers have found these loss function analyses to be decisive, and some commentators have criticized these approaches as being based on arguable approximations in the calculations of total error and for relying on undercount estimates that have been corrected for estimated bias (e.g., Wachter and Speed, 1991).

In this broad context, Freedman and Wachter have focused on one issue that was ignored in the Census Bureau's loss function analysis, namely, heterogeneity of undercount rates within poststrata, and Breiman's article targets many quantities that figure into these loss calculations and some other quantities that were omitted. We now turn to the interesting and informative article by Freedman and Wachter.

4. FREEDMAN AND WACHTER'S STUDY OF HETEROGENEITY

Heterogeneity of undercount rates within strata or poststrata is but one of many sources of error underlying sample-based population estimation. Presumably, the amount of heterogeneity bias in estimates depends heavily on the level of aggregation at which analysis takes place. Hengartner and Speed (1993) and Schafer (1993) engage in a lively debate over heterogeneity in undercount rates in the context of the proposed 1990 adjustment, shedding further light on the issue of an appropriate level of aggregation for analysis; Wolter and Causey (1991) discuss heterogeneity in the context of the proposed 1980 adjustment. Freedman and Wachter attempt to add to our understanding the heterogeneity bias, and their efforts are most welcome.

We commend FW for adopting the approach of analyzing proxies to gain insight into the plausible effects of adjustment. This methodology was used in Isaki, Schultz, Diffendal and Huang (1988), which has been invoked by some as a rationale for adjustment. By lending legitimacy to the general approach, FW make it easier to discuss the merits of the technical issues. Being able to agree upon methods of analysis is an essential step toward broader consensus on these issues.

A major point of agreement we have with FW is their conclusion about standard errors for small areas. Without focusing on details, we agree with their assessment at the outset of Section 5: "Heterogeneity is appreciable, adding substantial uncertainty to estimated undercounts for states and smaller areas." Omitting heterogeneity from small-area standard errors can clearly be a substantial error. Perhaps this work by FW could be used to estimate analogs to survey design effects to inflate standard errors by a fixed factor or by a set of factors that might depend on characteristics of the area. Would FW be willing to embrace the idea of taking this next step, or do they have an alternative solution in mind to the problem of estimating small-area standard errors?

Freedman and Wachter also amply demonstrate that the synthetic assumption for carrying adjustments down to small areas is false. Schafer's (1993) comment on the similar conclusion of Hengartner

and Speed (1993) was, "The Census Bureau is not surprised," and neither are we. This is an empirical conclusion about which there is broad consensus.

Our main criticism of FW is their interpretation of results from their loss function analysis. Their Table 9 considers seven proxy variables, four of which show the estimated risk difference to be lower than the true risk difference and three of which show the estimated risk difference to be higher. This does not surprise us. Their own algebra shows that the bias in the estimated risk difference for domain j is given by $2 \times (\text{ash}_j - \text{csh}_j) \times (\text{tsh}_j - \text{ash}_j)$, where tsh_j is the true population share, csh_j is the census population share and ash_j is the adjusted (synthetically estimated) population share. Contributions to bias are positive if the true undercounts are higher than the synthetic undercount estimates in the same domains where the adjusted population shares are estimated by the synthetic method to be higher than the census population shares, that is, in those domains where the PES identified large undercounts. As noted by Fay and Thompson (1993), it does not seem obvious why such a correlation would occur. Without such a correlation, one would expect the risk difference to be fairly symmetrically distributed around zero, as it is in FW's Table 9. A reviewer pointed out the following to us (emphasis in the original):

> A priori I would expect that at the local level, areas with high proportions of hard-to-count groups (highly segregated minority areas, areas in which everybody is a renter) would tend to have worse undercount FOR A GIVEN POSTSTRATUM than those with more of a mix (e.g., comparing black renters in a segregated renter-only area to black renters in an integrated area with many homeowners); this makes the covariance positive, favoring adjustment. It is less obvious to me how this would work out at the state level.

We think FW do a commendable job pointing up the limitations of proxy variable analysis while also illustrating its utility. So when they conclude, "In short, Table 9 shows that almost anything can happen," we think another reasonable voice might have phrased it, "These results support the theoretical argument that heterogeneity bias should, on average, favor neither adjusted nor unadjusted figures." It appears to us that FW are arguing a point as if they were in an adversarial proceeding. We favor the more balanced interpretation in Fay and Thompson (1993).

The possible presence of heterogeneity bias might not be irrelevant to a decision on adjustment, however, if there is some presumption in advance favoring either adjustment or nonadjustment. Fay (1993) pointed to the risk of heterogeneity bias as one of several factors that led him to change his mind about the 1991 adjustment decision. Fay had originally supported adjustment, but a few months after the decision he concluded from a subsequent analysis that the sampling variances may have been underestimated by a substantial factor and reconsidered his original support for adjustment. When it was pointed out to him, in a meeting where he was presenting his findings, that the amount by which he claimed the sampling variances were underestimated still would not have overturned the loss function advantage of the adjusted counts, Fay (1992b) replied, "Yes, but our margin of comfort is gone." When asked what he had in mind when he made that comment, Fay (1993) explained:

> The loss function analysis was incomplete. We didn't know whether we had all the biases in it fully measured. We can make quite a list including that we didn't know what to do at the time about heterogeneity bias.... I was saying if there are pieces left out of the puzzle, then...we have to allow for more uncertainty.

5. BREIMAN'S PAPER

Breiman's paper poses a substantial challenge. He focuses on a number of details of the adjustment procedure about which neither of us had more than general knowledge prior to writing this article, along with a few topics about which we have specialized knowledge. A primary reason that it is hard to respond to Breiman is that some of his depictions are very misleading. We wish he had not included material in his article that we believe deserves the harshest kind of criticism, but he did, and we feel obliged to point out what we view as unprofessional practices.

5.1 Breiman's Planned Testimony and the Current Paper

Although Breiman states that he assisted the defendants in the *New York v. Commerce Department* lawsuit, he did not note that this paper has the same title and numerous similarities to a manuscript (Breiman, 1992b) that he prepared to offer in support of his testimony (which it turned out he did not give). Indeed, many aspects of the current text are identical to the version prepared for the trial. This earlier version is relevant to material we discuss later.

5.2 The Computer Coding Error and a Contradiction in Breiman's Analysis

Breiman is correct to consider the potential effect of factors not included in the Total Error Model anal-

ysis on decisions about census adjustment, but inconsistencies in the standards he applies make it appear that his primary aim is to discredit the whole idea of adjustment. One of Breiman's main points is that if every area were undercounted by the same amount, then adjustment would be superfluous and would only add error: "The most important thing about estimating the undercount is not its total magnitude, but its differential effect on the 50 states and on thousands of counties and cities." Why then, in an apparently contradictory approach, does he summarize evidence on data quality, and on the computer coding error in particular, in terms of effects on the net national undercount rate?

That the computer coding error substantially affected the overall undercount estimate and looms as a warning signal for any future effort to undertake a complicated adjustment is widely acknowledged. The error was that E-sample cases that match at the PES sample address with a PES mover should be counted as erroneous enumerations rather than correct enumerations since the PES person reported that they should have been counted elsewhere on census day. This would affect differential undercounts between jurisdictions to the extent that people move to some types of areas more than they move to other types of areas, but that type of analysis is absent from Breiman's paper. The issue was addressed by Mulry and Spencer (1993), who wrote:

> The accuracy of the adjustment is affected by many kinds of errors. Some of the errors cancel, others do not.... To decide whether adjustment improves accuracy, it is necessary to consider net error.... There is no guarantee that all kinds of error can be identified, or even that the identified errors' moments are accurately estimated. For example, the original total error analysis failed to detect a large processing error that occurred when the wrong computer program was used to edit clerical match codes. Underestimating errors in the adjustment does not necessarily imply overestimation of the accuracy of the adjustment relative to the census, however.

Breiman's attempt to stretch the left endpoint of an interval estimate for the national undercount rate is at odds with his stated view that the issue of differential undercount is paramount.

When Breiman does focus on differential undercount in his Table 16, his commentary reflects the widespread view that the issue of differential undercount of minorities is important. Whey, then, does he not label evaluation poststrata as minority or nonminority? To allow readers to see the differential minority undercount, we reproduce Breiman's Table 16 with information on minority status added (see Table 1). Other authors (e.g., Mulry and Spencer, 1993; Zaslavsky, 1993a) add interval estimates in addition to labeling minority status.

TABLE 1

Minority evaluation strata	Total corrected undercount	Nonminority evaluation strata	Total corrected undercount
1	3.7	2	−2.5
3	2.6	4	−1.6
5	1.5	6	0.3
8	3.0	7	0.6
11	3.9	9	−0.5
		10	−0.7
		12	3.1*
		13	−0.2

*Might be 1.1.
Source: Breiman (1994, Table 16).

Even after accounting for the computer error, the Census Bureau's loss function calculations favor the adjusted over the unadjusted counts at the state level (Mulry, 1992a,b; Fay and Thompson, 1993) and at other less aggregated levels as well (Mulry, 1992a,b; Mulry and Spencer, 1993). In fact, the reanalysis of corrected data in Zaslavsky (1993a, b) is generally more favorable to adjustment than the analysis with uncorrected data. Zaslavsky considers several strategies for combining census counts and dual system estimates (DSE) of poststratum population totals (i.e., adjusted counts based on both census and PES data) to minimize the estimated risk of the estimator for a given loss function; he evaluates the performance of these strategies under various loss functions, such as the absolute-error and size-weighted squared relative error (SWSRE) criteria mentioned in our Section 2.6. Using the data set developed for the proposed 1991 adjustment, Zaslavsky finds that the linear combination between census and DSE minimizing the SWSRE estimated risk attaches 12.3% weight to the census and 87.7% weight to the DSE; using a revised data set that corrected the computer coding error and included other modifications (Mulry, 1992b), a similar analysis suggests that the SWSRE estimated risk is minimized when negative weight is attached to the census.

5.3 Matching Error and Another Contradiction

Breiman states that net undercount should be measured roughly as a difference between omissions and erroneous enumerations, allowing some errors in the census to be canceled by other errors in the census. However, in the PES, particularly in his

discussion of matching error, Breiman does not take into account that errors can cancel. Why, for example, does Breiman report matching error results primarily in terms of "disagreement rates"? If that is the standard approach, then the disagreement rate between the census and the PES is not his 2.1% net undercount figure, but more like 5.6% (General Accounting Office, 1991).

False matches and false nonmatches have biasing effects in opposite directions, and the loss function analyses carried out by the Census Bureau attempt to account for such effects. Curiously, after criticizing PES matching for its high disagreement rates, Breiman then uses the Census Bureau's calculations of the effects of matching error in his Table 15.

5.4 Fabrication and Obfuscation

On the issue of fabrication, Breiman states:

> The report concludes [P6, page 15] that "Overall, between .9 and 6.5% of the interviewers were found to have high nonmatch rates. This compares favorably with the expectation that between 2 to 5% of interviewers are dishonest in their data collection." If these latter figures are anywhere near truth, then the DSE numbers overestimate the undercount by millions of persons.

If Breiman's "millions of persons" had any basis in reality, then PES follow-up results would have been replete with fictitious PES enumerations. In fact, PES follow-up revealed 86 individuals who were fictitious enumerations in households where nonfictitious enumerations were recorded and 138 whole households that were fabricated out of 377,381 individuals in the P-sample in 171,390 housing units. Needless to say, fictitious enumerations detected in PES follow-up are not included in dual system estimates of the population. They do not inflate counts; they prevent the counts from being inflated.

In addition, quality-control checks suggesting 2–5% interviewer fabrication does not translate to 2–5% of cases entered in the P-sample database being fabrications. Indeed, the evidence suggests that undetected fictitious enumerations in the PES were a small fraction of detected fictitious enumerations (West, 1991), and rough calculations assuming average weights suggest that detected fictitious enumerations prevented an inflation of the population total by some tens of thousands of individuals. Thus, using Breiman's calculations, the quality control operations in the district offices helped prevent the dual system estimates from being inflated by "millions of people." However, despite his rating of the PES in his conclusion as "excellent judged only as a sample survey and matching operation," Breiman discounts the possibility that quality control at the original interviewing stage, PES follow-up interviewing and evaluation follow-up interviewing would have been able to catch fictitious enumerations.

5.5 Correlation Bias and the Weight to Attach to Imprecise Estimates of Nonsampling Errors

Although Breiman criticizes the [P16] total error analysis for ignoring sources of nonsampling error, he then states:

> Report [P16] gives a second set of undercount estimates that includes an additional error source called model bias, more commonly known as correlation bias. This is the bias ascribed to the existence of persons unreachable by any survey. However, because these bias estimates are (and must be) based on highly speculative assumptions and have only a tenuous connection with any data, they are not included in our discussion.

Correlation-bias estimates have more than a "tenuous connection with any data"; on the contrary, estimates are based on combining PES data with evidence from vital records about the ratio of the size of the male population to that of the female population. These data are used to estimate a parameter that characterize dependence in omission rates between the census and PES (Bell, 1993).

Bell (1993) does point out that assumptions underlying these estimates are speculative, so that Breiman's treatment of attaching no weight to the results, or, equivalently from a subjectivist point of view, attaching an infinite standard error to correlation bias estimates, may not be completely unreasonable. We would caution, however, that Bell (1993) and other empirical studies of correlation bias such as Zaslavsky and Wolfgang (1993) suggest that correlation bias is appreciable; in Bell's words, "assuming independence (no correlation bias) is even more restrictive [than alternative estimators of correlation bias] and appears to be refuted for adult males by the data (subject to the limitations of data quality, including those discussed in this section)." We also read Bell's Table 3 to imply that the primary evidence of differential correlation bias appears to be in the 20–29-year-old age group, with the differential bias going in the direction that suggests even higher black undercount.

In sum, less accurate bias estimates should be given less weight in loss calculations, whether formal or informal. Using this standard, Zaslavsky (1993a)

provides a more reasoned approach to the reflection of evidence about biases than does Breiman's paper, which appears to treat some bias estimates as having infinite standard errors and other bias estimates as having zero standard errors.

5.6 Breiman's Table 13 as "Bad Data"

Breiman's Table 13, which originally appeared in a Census Bureau report, contains a flaw that makes the table's column headings not directly comparable with the entries in the interior of the table. This flaw was pointed out in Census Bureau memoranda written by Belin to senior Census Bureau officials (Belin, 1991a,b). Prior to receiving these memoranda, there had been plans to raise in the Undercount Steering Committee report the same concern Breiman has raised with his Table 13; however, based on the information in Belin's memoranda, a decision was made not to mention the issue. Because the details are cumbersome, we postpone a full discussion to the Appendix.

Breiman may not have known about the relevant internal Census Bureau documents when he was writing earlier drafts of the article, but Belin was contacted by one of the referees of Breiman's paper on its initial review and gave the reviewer copies of these memoranda. Our understanding is that the reviewer mentioned the flaw in Table 13 in a referee's report to Breiman. Breiman appears to have made no effort to address this criticism, as the table itself is unchanged from the earlier version of his paper, and the text, which is quoted in the Appendix, is identical in substance.

5.7 The Prism through which Breiman Views Tables 11 and 12

A *Statistical Science* article is not a place to play games with the base of percentages. An informative comparison in Breiman's Tables 11 and 12 is between the imputed probabilities and the fraction of cases resolved in EFU that were found to have been included in the census. This comparison yields an average predicted probability of 32.2% for PES cases that were originally unresolved but later resolved in the EFU as compared to a proportion of 31.6% for the enumeration rate among resolved EFU cases (Belin et al., 1993). Some of the decimal places of accuracy are lost in Breiman's rounding, but it is plain enough from Table 11 that $12/(12 + 27) = 30.8\%$, so rounding error does not appear to be the problem here. A similar perspective emerges from our Table 2.

Belin et al. (1993) provide a more detailed analysis of the same EFU data that includes appropriate predictive standard errors, demonstrating good agreement between imputed probabilities and enumeration rates from the EFU not only in the aggregate but also across important subgroups. Wachter (1993b) criticized the interpretation of Belin et al. (1993) because the EFU-resolved cases do not amount to a random subsample of the EFU sample, which generated discussion of the value of the EFU data in the presence of nonresponse, but Wachter did not deny that the agreement between what was observed and what was predicted was extremely good. Breiman calls the interpretation of Belin et al. (1993) "optimistic."

TABLE 2
Data from Breiman's Table 12 for resolved evaluation follow-up (EFU) cases

Imputed probability of having been in census	0–25%	25–50%	50–75%	75–100%
Fraction of resolved EFU cases determined to have been in census	12%	33%	53%	68%

Regarding Table 12, the 1992 version of Breiman's paper reads: "Note that the percent in the match category is virtually constant from an imputed probability of 25% to 100%. One would do just as well by assigning each person to match or other than match by tossing a coin with probability 15% of heads." The current version reads: "As the match probabilities increase, the proportion of resolved cases that result in matches increases. Another, more curious, trend also seems to be present—the higher the model match probabilities, the higher the proportion of unresolved persons the matching team finds." The latter is an improvement, but it seems to us that Breiman is grasping for something to criticize.

For us, the important question raised by this discussion is not about imputation methods. The real question is whether the statistics profession will require Census Bureau statisticians to respond to every criticism raised by a professional outside the Census Bureau, no matter how questionable, before sanctioning some kind of adjustment methodology.

5.8 Breiman's "Unresolved Issues" with the Census Bureau

Breiman attempts to use the fact that he circulated a draft of his paper at the Census Bureau to corroborate his conclusions: "The results of this paper were circulated to the Census Bureau in March 1992. In the reviews there was agreement with the numbers given in Table 15 except, perhaps, in two areas...[that] are discussed in the Appendix." In the Appendix, he discusses "census day address error," which refers to cases where the PES incorrectly recorded an individual's residence as of April 1, 1990, implying that the PES matching operation may have been searching in the wrong location for a census

enumeration. Such PES cases may then be incorrectly designated as nonmatches to the census, which would inflate undercount estimates. In his interaction with Bureau staff, Breiman questioned the formula for computing census day address error, and he asserts that the matter was never resolved:

> In December 1991, I was informed that the formula was incorrect, that the tables in [P4] were wrong and that the census day address error computation in [P4] and [P16] included all of the errors found in P-sample reinterviews regardless of whether they were census day address error or not. This was surprising, since this error is consistently referred to both in [P4] and [P16] as census day address error. Furthermore, some of the tables in [P4] are reproduced in the published paper titled "Address reporting error in the 1990 post-enumeration study" (West, Mulry, Parmer and Petrick, 1991).

Breiman then states that he was "unable to obtain from the Bureau any more specific information regarding their method for computing census day address error."

Our conversations with Census Bureau officials lead us to conclude that some of Breiman's assertions are misleading. First, Breiman's inability to obtain the information he desired should not be interpreted as suggesting that the Census Bureau was uncooperative. John Thompson (1994) of the Census Bureau explained to us:

> Leading up to the lawsuit, we transmitted a lot of information to Leo and others. To my knowledge, we had provided the answers to all of his requests, and I was unaware that he was waiting for anything. I thought we had agreed with Leo that there was a mistake in the earlier documentation that had been corrected in the documentation of the 1992 adjustment decision.

Mary Mulry (1994) of the Census Bureau echoed this sentiment, pointing out that the original calculation was correct and that it was just the documentation that was wrong. The effect of census day address error on undercount estimates goes in one direction, but other PES data collection errors could have effects in either direction. As Breiman notes, the Bureau had included all sources of data collection error in its original calculations but had summarized them under the label "census day address error." Responding to input from Breiman, Bureau staff changed their documentation by substituting "error in address reporting and other P-sample data collection errors" (Mulry, 1992b) for what had previously been labeled "census day address error." Thus, Breiman's inability to obtain "more specific information regarding their method for computing census day address error" appears to have resulted from Bureau officials deciding that separate calculations for each type of error were inessential so long as the accounting of cases was complete and the documentation of methods was accurate. Breiman cites the report by Mulry (1992b) in his paper, but he ignores the change in the documentation.

Clearly, having the ability to reproduce results is crucial, and questions arise as to appropriate standards for documentation. Is it enough that a Census Bureau employee familiar with detailed codes and jargon can reproduce results, or must the documentation be understandable to outside professional statisticians? What is the significance of a document that is in error if the actual calculation that was carried out is correct?

Many aspects of the 1990 adjustment procedure were rushed, so for the future we need to be realistic in assessing how long various operations take. Meanwhile, Breiman's criticism notwithstanding, our view is that the 1990 adjustment procedure was acceptably documented. We also think that it would have been better for Breiman to resolve his dispute somewhere other than in these pages.

Also, the Census Bureau's corroboration of Breiman's numbers should not be interpreted as Census Bureau endorsement of his conclusions. For example, according to the July 20, 1992, minutes of the Committee on Adjustment of Postcensal Estimates (CAPE) at the Census Bureau, that meeting was devoted to discussing a 13-page critique of an earlier version of Breiman's work by Hogan (1992b). Thompson (1994) also said: "His numbers appear to be correct, but the inferences are Leo's interpretation of the data. The main point I'd like to make is that if journals publish criticisms of the Census Bureau, then they should offer the Census Bureau an opportunity to comment." The Census Bureau is a professional agency. Ultimately, we do not believe that Breiman does justice to the Census Bureau in his paper.

6. YEAR 2000: WILL CONSENSUS EMERGE AROUND A "ONE-NUMBER CENSUS"?

The Census Bureau is planning for the year 2000 census and beyond (Holmes, 1994). Tortora, Miskura and Dillman (1993) summarize the design options from the perspective of senior Bureau officials. Current plans call for selecting the design of the next census by December 1995. A primary goal is to de-

sign a "one-number census," which Tortora, Miskura and Dillman (1993) define in the following way (emphasis in original):

> The term "One-Number Census" names the concept that *the decennial census is designed to produce the best possible single set of results by legal deadlines, and that those results are based on an appropriate combination of counting, assignment, and statistical estimation techniques.*

The evolution of the one-number census flows in many ways from the adjustment experience in 1990. We briefly describe some of the connections.

Apart from the technical issues we have discussed, the political reality is that there cannot be two population numbers, no matter how good either one of them is, without politicizing the process. In 1991, we saw the Secretary of Commerce being required to choose one set of numbers or the other after the consequences of his decision were apparent, a scenario that would open any decision to the charge of being politically motivated. As Barbara Everitt Bryant, the director of the Census Bureau during the adjustment controversy over the 1990 census, recently wrote (Bryant, 1993):

> It becomes clear from the experience of 1990 that a two-number census did not work. With a 200-year history of never adjusting the census, the production of two numbers made one number the adversary of the other, with the consequent demand for statistically impossible proof that one number is better than the other at all levels at which data are produced.

The statistical community may think of adjustment as a technical issue, and policymakers may endorse this view so long as the technical decisions are made in advance of knowing their effects. The concept of a one-number census recognizes this imperative.

Although Tortora, Miskura and Dillman (1993) do not require that "statistical estimation techniques" be used in order for a census to qualify as a one-number census, current plans for the 1995 test census suggest that statistical methods are high on the Census Bureau's research agenda. The debate over adjustment is thus imbedded in the consideration of a one-number census, with one distinction being that "adjustment" in a one-number census is built into the census design and is viewed as an integral part of the census. One reflection of how the thinking, or at least the nomenclature, has evolved at the Bureau is that the paper by Tortora, Miskura and Dillman does not contain the word "adjustment."

Tortora, Miskura and Dillman (1993) are aware that there are some major hurdles to clear before a one-number census could become reality, which these Bureau officials pose as an opportunity and a challenge to statisticians outside the Bureau:

> We have proposed that the 1995 Census Test be a one-number census.... Determining how to do this combines some intriguing technical, perceptual, and operational issues. One set of issues involves the estimation-based coverage improvement methods to be used.... Another set of issues has to do with determining if the new one-number approach, which will rely less on counting techniques and more on estimation than the 1990 census, yields the desired improvement.... What measures are needed to make a convincing case? Is it sufficient to show that we have improved the relative coverage of population groups; or is it important to also show that the "proportionate shares" of the total population across all areas are improved? In either case, what is the geographic level at which such improvement must be shown?... We are currently discussing these issues within the Census Bureau and invite the input of the technical community on them.

Along with the political imperative mentioned above, the push for a one-number census appears also to be based on other nonstatistical considerations, such as operational feasibility. Time constraints imposed by legal deadlines would seem to eliminate the possibility of carrying out a 1990-style PES and EFU program and then allowing enough time to deliberate over a comparison of an "unadjusted" and an "adjusted" census. Bob Fay, the Senior Mathematical Statistician at the Bureau, described recently in the following exchange with Tom Belin (Fay, 1993):

> Fay: The state of science will not be such in the year 2000 that we can achieve consensus by looking at a lot of evaluation data over a four-week period and deciding how to adjust the census. It now seems that if we're going to make this change we'll have to just decide ahead of time to live with something, and then, go back and look and see what we've done over a two-year period. But in fact, even the notion of guidelines or the two-track approach may not be the way the society decides to adjust its data.

> Belin: In other words, the type of guidelines, if you will, that you're imagining would say, "Just do it" and then, review it.
>
> Fay: Yes. And in some ways I think Barbara Bailar and Kirk Wolter were ahead of me on this by ten years.... The only subtle difference is to design the census around the idea that it has to be ready within the usual scheduled timetables for the census rather than delay the census eight months or longer.

Another compelling consideration, mentioned by Bryant (1993), is the desire to avoid tying up the top statisticians at the Census Bureau in litigation.

A number of important voices have registered support for the idea of a one-number census. The report of the Panel to Evaluate Alternative Census Methods (Committee on National Statistics, 1993b, page 2) states:

> One key message is that the dual objectives of reducing the differential undercount and controlling costs will require expanded use of sampling and statistical estimation.... We endorse the Census Bureau's stated goal of achieving a one-number census in 2000 that incorporates the results from coverage measurement programs, including programs involving sampling and statistical estimation, into the official census population totals. We recommend that research on alternative methodologies continue in pursuit of this goal.

Bryant (1993) also endorses the idea of a one-number census in forceful terms:

> The 1990 experience also demonstrated that differential undercount cannot be reduced by more intensive enumeration efforts. It is not possible to count 100% of the population, although the public and many elected officials have an unrealistic expectation that this can be done. Combining statistical estimation of those missed with the best reasonable effort to enumerate all in a one-number census should improve overall accuracy.... The census has changed evolutionarily with the times. Now is the time to make a larger than evolutionary leap.

To facilitate discourse, terms like "overall accuracy" should be defined to distinguish, for example, between "state-level" and "block-level" accuracy, as we discuss in Section 2.7. However, there appears to be a growing consensus around the idea of integrating statistical estimation into the census-taking process instead of generating two sets of counts and deciding between them. Hopefully, this confluence of views will create momentum for statisticians with widely varying perspectives to reach the kind of consensus envisioned by the National Academy panel in the mid-1980's.

7. CONCLUSION

It is true that there are a great many sources of error in adjustment and that the details matter greatly. Not all sources of uncertainty in the proposed adjustment of the 1990 census were quantified, and there was a major processing error discovered after senior Census Bureau officials had already recommended adjustment. Freedman and Wachter zero in on one of the fundamental issues about which there is genuine philosophical disagreement, and despite their own leanings against adjustment they have produced analyses that help clarify a technical issue of great interest. This is constructive. On the other hand, Breiman is not careful with his facts, and one of the most harmful aspects of Breiman's paper, we believe, is that Breiman has cultivated a sense of distrust of government statisticians that we regard as unjustified and irresponsible.

Ultimately, we would press our colleagues to embrace the idea of a "negotiated settlement" to the adjustment controversy. Crucial design decisions need to be made in the next few years. As a profession, we should press policymakers in Congress to exercise their prerogative and responsibility to, as the Constitution puts it, "direct" the census. By this we mean that Congress should state what it needs from the Census Bureau and should support the best professional judgment of the Census Bureau on how to get there. If present trends in designing the next census continue and get widespread support from the statistics community, we may be able to avoid full-blown court cases after the year 2000 census that would waste time, money and the energies of outstanding members of the profession. To statisticians who pride themselves on improving society by avoiding wasteful public and private spending, this looms as a calling.

APPENDIX: BREIMAN'S TABLE 13 AS "BAD DATA"

Breiman's Table 13 on the accuracy of E-sample imputation originated in the preliminary report for project P3 at the end of May 1991 (Gbur, 1991b). The preliminary report also contains a table summarizing evidence about P-sample imputation, which Breiman has reproduced as his Table 12. Belin, who

played a prominent role in producing the imputed probabilities, was still working at the Census Bureau at the time the preliminary P3 report was circulated. At the time, the P-sample results were generally received by senior Census Bureau officials as being supportive of the imputation methodology, contrary to Breiman's interpretation of the results. However, the E-sample results were a source of concern among Bureau officials, much as they are in Breiman's account, largely due to the much greater than expected fraction of individuals classified as correctly enumerated from the EFU who received imputed probabilities of 50% or less of having been correctly enumerated. Belin was told that the Undercount Steering Committee, which consisted of senior Bureau officials, was planning to include a comment about the discrepancy in the E-sample results in their final report. This is important because the absence of any mention of this discrepancy in Undercount Steering Committee (1991) confirms that senior Bureau officials agreed with Belin's analysis of the situation at the time.

Belin and his colleagues quickly identified a flaw in the construction of the E-sample table. PES matching rules for E-sample cases that were unmatched in before-follow-up matching declare that the E-sample case is a correct enumeration only if three conditions hold: a follow-up interview must establish the person to exist, the E-sample address has to be correctly assigned to the sample block or to an adjacent block, and the individual cannot be duplicated in a search area defined as a ring of three "surrounding blocks." The latter condition is required because otherwise it would remain unclear whether the correctly enumerated individual was the one in the sample block or the one in the surrounding block. To reflect this uncertainty in the case of a duplicate in a surrounding block, the standard Census Bureau procedure, as described in Belin et al. (1991), was to assign an imputed probability of correct enumeration of 50% when an E-sample case had one surrounding block duplicate, 33.3% for an E-sample case that matched two other census records in surrounding blocks, and so on.

The preliminary P3 report, according to its author, placed all E-sample cases having surrounding block duplicates in the 0–50% category in the table. However, EFU operations were only concerned with the first of the three conditions mentioned above; there had been no review of whether the E-sample case in the sample block was "the" correct enumeration as opposed to one of the surrounding block duplicates being "the" correct enumeration. Another contribution to the imputed probability of correct enumeration, namely, the possibility of incorrect assignment of residences to geographical locations, known as geocoding error, was also not part of the EFU review. Almost 90% of the cases that appear in the 0–50% category of what is now Breiman's Table 13 were attached to surrounding block duplicates.

Belin (1991a, b) wrote two memoranda to Census Bureau superiors on this matter. Using the same data set as was used to generate the preliminary P3 report, he found that the average imputed probability for individuals in the 0–50% column, excluding the contributions of surrounding block duplicates and possible geocoding error, was 88.7%. The correct enumeration rate among resolved cases from the EFU was $67/(67 + 10) = 87.0\%$.

As noted earlier, there is no mention of the discrepancy in the E-sample table in Undercount Steering Committee (1991). However, the author of the P3 preliminary report, despite receiving copies of both of Belin's memoranda, unfortunately left the table unedited in the final P3 report (Gbur, 1991c). It is understandable both that Breiman would have taken the table to support the position he was set to argue in court and that he would not have been aware of Belin's memoranda, which, although included in the administrative record of the adjustment decision, amounted to only five pages out of thousands.

During the review of Breiman's article for *Statistical Science*, an anonymous referee who knew that Belin had worked on imputation in the PES contacted Belin to discuss Table 13. Belin provided copies of the memoranda, and we understand that the referee conveyed an explanation of this issue to Breiman.

Between Breiman (1992) and Breiman (1994), Table 13 did not change, and the discussion surrounding it changed from, "There is no sign of a correlation between the probabilities computed by the imputation model and the enumeration status as determined by the EFU reinterview information" to "There is no evidence here of an association between the probabilities computed by the imputation model and the enumeration status as determined by the EFU reinterview information." Belin wrote the Executive Editor of *Statistical Science* to make him aware of the situation and to request an opportunity to respond, which grew into our current collaboration.

ACKNOWLEDGMENTS

We are grateful to Bob Fay, Mary Mulry and John Thompson for helpful discussion and for responding to requests for information. We also thank Norman Bradburn, Louis Gordon, Rob Kass, Don Ylvisaker and two referees, whose comments greatly improved the text.

REFERENCES

ALHO, J. M., MULRY, M. H., WURDEMAN, K. and KIM, J. (1993). Estimating heterogeneity in the probabilities of enumeration for dual-system estimation. *J. Amer. Statist. Assoc.* **88** 1130–1136.

ANDERSON, M. J. (1988). *The American Census: A Social History*. Yale Univ. Press.

ANOLIK, I. (1988). The rural post-enumeration survey in east central Mississippi, Statistical Research Division Report, Series RR 88/10, Bureau of the Census, Washington, D.C.

ANOLIK, I. (1989). The 1987 Post-Enumeration Survey. In *Proceedings of the Survey Research Methods Section* 710–715. Amer. Statist. Assoc., Alexandria, VA.

ANOLIK, I. (1990). Matching movers in a post-enumeration survey: results from 1988 and plans for 1990. In *Proceedings of the Survey Research Methods Section* 536–540, Amer. Statist. Assoc., Alexandria, VA.

ANOLIK, I. and HOGAN, H. (1991). Non-response conversion: the experience of the 1990 Post-Enumeration Survey. In *Proceedings of the Survey Research Methods Section* 658–663 (P3). Amer. Statist. Assoc., Alexandria, VA.

BATEMAN, D., CLARK, J., MULRY, M. and THOMPSON, J. (1991). 1990 Post-Enumeration Survey evaluation results. In *Proceedings of the Social Statistics Section* 21–30. Amer. Statist. Assoc., Alexandria, VA.

BELIN, T. R. (1991a). Interpretation of results from evaluation project P3. Statistical Support Division, Bureau of the Census, Washington, D.C. (June 5, 1991.)

BELIN, T. R. (1991b). Comparison of E-sample results from P3 evaluation of production imputation results. Statistical Support Division, Bureau of the Census, Washington, D.C. (June 26, 1991.)

BELIN, T. R. (1993). E-mail correspondence to D. Freedman, August 31, 1993.

BELIN, T. R. and DIFFENDAL, G. J. (1991). Results from the handling of unresolved enumeration status, missing characteristic data, and noninterviews in the 1990 Post-Enumeration Survey. STSD Decennial Census Memorandum Series V-112, Bureau of the Census, Washington, D.C.

BELIN, T. R., DIFFENDAL, G. J., MACK, S., RUBIN, D. B., SCHAFER, J. L. and ZASLAVSKY, A. M. (1993). Hierarchical logistic–regression models for imputation of unresolved enumeration status in undercount estimation (with discussion). *J. Amer. Statist. Assoc.* **88** 1149–1166.

BELIN, T. R., DIFFENDAL, G. J., RUBIN, D. B., SCHAFER, J. L. and ZASLAVSKY, A. M. (1991). Documentation of handling of unresolved enumeration status in 1990 census/Post-Enumeration Survey. STSD Decennial Census Memorandum Series V-98, Bureau of the Census, Washington, D.C.

BELL, W. R. (1991). Standardized residuals from regression models for variances of adjustment factors. Statistical Research Division, Bureau of the Census, Washington, D.C. (May 16, 1991.)

BELL, W. R. (1993), Using information from demographic analysis in Post-Enumeration Survey estimation. *J. Amer. Statist. Assoc.* **88** 1106–1118.

BIEMER, P. and STOKES, S. (1989). The optimal design of quality control samples to detect interviewer cheating. *Journal of Official Statistics* **5** 23–39.

BREIMAN, L. (1992). The 1991 census adjustment: Undercount or bad data? Preliminary version, technical report, Dept. Statistics, Univ. California, Berkeley.

BREIMAN, L. (1994). The 1991 census adjustment: Undercount or bad data? *Statist. Sci.* **9** 458–475.

BRYANT, B. E. (1991). Recommendation to Secretary of Commerce Robert A. Mosbacher on whether or not to adjust the 1990 census. Bureau of the Census, Washington, D.C. (June 1991.)

BRYANT, B.E. (1992). 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. *Federal Register* Docket 920895–2347, December 29, 1992.

BRYANT, B. E. (1993). Guest commentary. Census-taking for a litigious, data-driven society. *Chance* **6** (Summer) 44–49.

BUREAU OF THE CENSUS (1992a). Request for comments on the proposed options for incorporating information from the Post-Enumeration Survey into the intercensal population estimates produced by the Bureau of the Census. *Federal Register* **57** 35,562–35,564 (August 10).

BUREAU OF THE CENSUS (1992b). Assessment of accuracy of adjusted versus unadjusted 1990 census base for use in intercensal estimates. Report of the Committee on Adjustment of Post Censal Estimates, (August 7) Suitland, MD.

BUREAU OF THE CENSUS (1992c). Decision on whether to incorporate information from the Post-Enumeration Survey into the base for intercensal population estimates. *Federal Register* **57** 42,939–42,940 (September 17).

BUREAU OF THE CENSUS (1992d). Additional research on accuracy of adjusted versus unadjusted 1990 census base for use in intercensal estimates. Addendum to the Report of the Committee on Adjustment of Post Censal Estimates, Suitland, MD (November 28).

BUREAU OF THE CENSUS (1993). 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. *Federal Register* **58** 69–77. (January 4).

CAIN, B. (1992). Exhibit for *New York et al. v. Commerce Department et al.* U.S. District Court, Eastern District of New York.

CHILDERS, D., DIFFENDAL, G., HOGAN, H., SCHENKER, N. and WOLTER, K. (1987). The technical feasibility of correcting the 1990 census In *Proceedings of the Social Statistics Section* 36–45. Amer. Statist. Assoc., Alexandria, VA.

CHILDERS, D. R. and HOGAN, H. (1989a). The 1988 Post Enumeration Survey, methods and preliminary results. Statistical Research Division, Bureau of the Census, Washington, D.C. (October 11, 1989.)

CHILDERS, D. R. and HOGAN, H. (1989b). The 1988 Post Enumeration Survey, methods and preliminary results. In *Proceedings of the Survey Research Methods Section* 685–689. Amer. Statist. Assoc., Alexandria, VA.

CHILDERS, D. R. and HOGAN, H. (1990). Results of the 1988 dress rehearsal post enumeration survey. In *Proceedings of the Survey Research Methods Section* 547–552. Amer. Statist. Assoc., Alexandria, VA.

CHOLDIN, H. M. (1994). *Looking for the Last Percent: The Controversy over Census Undercounts.* Rutgers Univ. Press.

CITRO, C. F. and COHEN, M. L., eds. (1985). *The Bicentennial Census: New Directions for Methodology in 1990.* National Academy Press, Washington, D.C.

COHEN, M. L. (1990). Adjustment and reapportionment—analyzing the 1980 decision. *Journal of Official Statistics* **6** 241–250.

COHEN, M. L. (1991). Choice and conflict about census data: adjusting the American census count. *Journal of Public Policy* **11** 357–398.

COMMITTEE ON NATIONAL STATISTICS (1993a). Planning the decennial census: interim report. Report of the Panel on Census Requirements in the Year 2000 and Beyond, National Research Council, Washington, D.C.

COMMITTEE ON NATIONAL STATISTICS (1993b). A census that mirrors America: interim report. Report of the panel to Evaluate Alternative Census Methods, National Research Council, Wahington, D.C.

CRESSIE, N. (1992). REML estimation in empirical Bayes smoothing of census undercount, *Survey Methodology* **18** 75–94.

DARROCH, J. N., FIENBERG, S. E., GLONEK, G. F. V. and JUNKER, B. W. (1993). A three-sample multiple-recapture approach to census population estimation with heterogeneous catchability. *J. Amer. Statist. Assoc.* **88** 1137–1148.

DATTA, G. S., GHOSH M., HUANG, E. T., ISAKI, C. T., SCHULTZ, L. K. and TSAY, J. H. (1992). Hierarchical and empirical Bayes methods for adjustment of census undercount: the 1988 Missouri dress rehersal data. *Survey Methodology* **18** 95–108.

DAVIS, M. C., MULRY, M., PARMER, R. and BIEMER, P. (1991). The matching error study for the 1990 post enumeration study. In *Proceedings of the Survey Research Methods Section* 248–253. Amer. Statist. Assoc., Alexandria, VA.

DEPARTMENT OF COMMERCE (1989). *Federal Register* **54**(238) 51,002–51,004.

DEPARTMENT OF COMMERCE (1991a). *Department of Commerce News* June 13.

DEPARTMENT OF COMMERCE (1991b). Decision on whether or not a statistical adjustment of the 1990 decennial census of population should be made for coverage deficiencies resulting in an overcount or undercount of the population. Docket No. 91281–1181, July 15, 1991. Office of the Secretary of Commerce, Washington, D.C. *Federal Register* **56** 33,582–33,642 (July 22nd).

DIFFENDAL, G. (1988). The 1988 test of adjustment related operations in Central Los Angeles County. *Survey Methodology* **14** 71–86.

DIFFENDAL, G. and BELIN, T. (1991). Results of procedures for handling noninterviews, missing characteristic data, and unresolved enumeration status in 1990 census/Post-Enumeration Survey. STSD Decennial Census Memorandum Series #V-112, Bureau of the Census, Washington, D.C.

ERICKSEN, E. P. and DEFONSO, T. K. (1993). Beyond the net undercount: how to measure census error, *Chance* **6** (Fall) 38–43.

ERICKSEN, E., ESTRADA, L., TUKEY, J. and WOLTER, K. (members, Special Advisory Panel) (1991). Report on the 1990 census and the Post Enumeration Survey. Report submitted to the secretary of the Department of Commerce, June 22, 1991.

ERICKSEN, E. P. and KADANE, J. B. (1985). Estimating the population in a census year (with discussion). *J. Amer. Statist. Assoc.* **80** 98–131.

ERICKSEN, E. P., KADANE, J. B. and TUKEY, J. W. (1989). Adjusting the 1980 census of population and housing. *J. Amer. Statist. Assoc.* **84** 927–943.

FAY, R. E. (1992a). Inferences for small domain estimates from the 1990 Post-Enumeration Survey. Technical report, Bureau of the Census, Washington, D.C.

FAY, R. E. (1992b). Personal communication (presentation in a meeting during Census Bureau Annual Research Conference).

FAY, R. E. (1993). Personal communication (interview with T. Belin).

FAY, R. E., PASSEL, J. S., ROBINSON, J. G. and COWAN, C. D. (1988). *The Coverage of the Population in the 1980 Census*. U.S. Department of Commerce, Washington, D.C.

FAY, R. E. and THOMPSON, J. H. (1993). The 1990 Post Enumeration Survey: statistical lessons in hindsight. In *Proceedings of the 1993 Annual Research Conference* 71–91. Bureau of the Census, Washington, D.C.

FIENBERG, S. E. (1989–1992). An adjusted census in 1990? Assorted articles. *Chance* **2–5**.

FIENBERG, S. E. (1992a). Comment on "Should we have adjusted the U.S. census of 1980?" by D. A. Freedman and W. C. Navidi. *Survey Methodology* **18** 25–28.

FIENBERG, S. E. (1992b). Bibliography on capture–recapture modelling with application to census undercount adjustment. *Survey Methodology* **18** 143–154.

FIENBERG, S. E. (1993). The New York City adjustment trial: witness for the Plaintiffs. *Jurimetrics* **34** 65–84.

FIENBERG, S. E. (1994a). Ethical and modelling considerations in correcting the results of the 1990 decennial census. In *Ethics Modelling* (W. A. Wallace, ed.) 103–114. Pergamon Press, New York.

FREEDMAN, D. (1991). Adjusting the 1990 census. *Science* **252** 1233–1236.

FREEDMAN, D. A. (1992). Testimony in *New York et al. v. Commerce Department et al.* U.S. District Court, Eastern District of New York, 89 CV 3474.

FREEDMAN, D. A. (1993a). E-mail correspondence with T. Belin, September 5, 1993.

FREEDMAN, D. A. (1993b). Adjusting the census of 1990. *Jurimetrics* **34** 99–106.

FREEDMAN, D. and NAVIDI, W. (1986). Regression models for adjusting the 1980 census (with discussion). *Statist. Sci.* **1** 1–39.

FREEDMAN, D. and NAVIDI, W. (1992). Should we have adjusted the U.S. census of 1980? (with discussion). *Survey Methodology* **18** 3–74.

FREEDMAN, D. and WACHTER, K. (1994). Heterogeneity and census adjustment for the intercensal base. *Statist. Sci.* **9** 476–485.

FREEDMAN, D., WACHTER, K., COSTER, D., CUTLER, R. and KLEIN, S. (1993). Adjusting the census of 1990: the smoothing model. *Evaluation Review* **17** 371–443.

FREEDMAN, D., WACHTER, K., CUTLER, R. and KLEIN, S. (1994). Adjusting the census of 1990: loss functions. *Evaluation Review* **18** 243–280.

GBUR, P. M. (1991a). Missing data in the 1990 Post Enumeration Survey. In *Proceedings of the Survey Research Methods Section* 646–651. Amer. Statist. Assoc., Alexandria, VA.

GBUR, P. M. (1991b). PES Evaluation Project P3: evaluation of imputation methodology for unresolved match status cases. Preliminary version, Statistical Support Division, Bureau of the Census, Washington, D.C. (May 31, 1991.)

GBUR, P. M. (1991c). PES Evaluation Project P3: evaluation of imputation methodology for unresolved match status cases. Final version, 1990 Coverage Studies and Evaluation Memorandum Series #C-2, Statistical Support Division, Bureau of the Census, Washington, D.C. (July 12, 1991.)

GENERAL ACCOUNTING OFFICE (1991). The 1990 census: final preparations for a possible adjustment. Report GAO/T-GGD-91-26, June 19, 1991.

GRIFFIN, D. H. and WAJER, S. C. (1993). Results from the 1990 search match operation. In *Proceedings of the Survey Research Methods Section.* Amer. Statist. Assoc., Alexandria, VA. To appear.

HARTIGAN, J. A. (1992). Comment on "Should we have adjusted the U.S. census of 1980?" by D. A. Freedman and W. C. Navidi. *Survey Methodology* **18** 44–50.

HENGARTNER, N. and SPEED, T. P. (1993). Assessing between-block heterogeneity within the post-strata of the 1990 Post-Enumeration Survey (with discussion). *J. Amer. Statist. Assoc.* **88** 1119–1129.

HOGAN, H. (1989). Nine years of coverage evaluation research: What have we learned? In *Proceedings of the Survey Research Methods Section* 663–668. Amer. Statist. Assoc., Alexandria, VA.

HOGAN, H. (1990). The 1990 Post-Enumeration Survey: an overview. In *Proceedings of the Survey Research Methods Section* 518–523. Amer. Statist. Assoc., Alexandria, VA.

HOGAN, H. (1991). Downweighting outlier small blocks. STSD 1990 Decennial Census Memorandum Series #V-109, addressed to John Thompson, Chief, Statistical Support Division, Bureau of the Census, Washington, D.C. June 18, 1991.

HOGAN, H. (1992a). The 1990 Post-Enumeration Survey: an overview. *Amer. Statist.* **46** 261–269.

HOGAN, H. (1992b). Discussion of "The 1991 census adjustment: undercount or bad data" (June 1992 version) by L. Breiman. Minutes of the Committee on Adjustment of Postcensal Estimates, Bureau of the Census, Washington, D.C. (July 20, 1992.)

HOGAN, H. (1993). The 1990 Post-Enumeration Survey: operations and results. *J. Amer. Statist. Assoc.* **88** 1047–1060.

HOGAN, H. and WOLTER, K. (1988). Measuring accuracy in a post-enumeration survey. *Survey Methodology* **14** 99–116.

HOLMES, S. A. (1994). Census officials plan big changes in gathering data. *The New York Times*, May 16, 1994, page A1.

ISAKI, C. T., SCHULTZ, L. K., DIFFENDAL, G. J. and HUANG, E. T. (1988). On estimating census undercount in small areas. *Journal of Official Statistics* **4** 95–112.

JARO, M. (1989). Advances in record-linkage methodology as applied to matching the 1985 census of Tampa, Florida. *J. Amer. Statist. Assoc.* **84** 414–420.

KADANE, J. B. (1984). Allocating congressional seats among the states when state populations are uncertain. Technical Report 309, Dept. Statistics, Carnegie Mellon Univ.

KIM, J. J., BLODGETT, R. and ZASLAVSKY, A. (1993). Evaluation of the synthetic assumption in the 1990 Post-Enumeration Survey. Technical report, Bureau of the Census, Washington, D.C.

KRUSKAL, W. (1991). Recommendation to Secretary of Commerce Robert A. Mosbacher on adjustment of the 1990 census, June 1991.

MARTIN, E., BROWNRIGG, L. A. and FAY, R. E. (1990). Results of ethnographic studies of 1988 dress rehearsal census coverage. In *Proceedings of the Survey Research Section*. 530–535. Amer. Statist. Assoc., Alexandria, VA.

MEYER, M. M. and KADANE, J. B. (1992). Reconstructing the adjusted census for Florida: A case study in data examination. *Journal of Computational and Graphical Statistics* **1** 287–300.

MITROFF, I., MASON, R. and BARABBA, V. (1983). *The 1980 Census: Policymaking Amid Turbulence*. Heath, Lexington, MA.

MOSBACHER, R. A. (1991a). The decision of the secretary of commerce on whether a statistical adjustment of the 1990 Census of Population and Housing should be made for coverage deficiencies resulting in an overcount or undercount of the population. *Federal Register* **56** 140, 33,582–33,688, (July 22, 1991.)

MOSBACHER, R. A. (1991b). Testimony before the House Subcommittee on Census and Population, Committee on Post Office and Civil Service, Serial No. 102–21, July 16, 1991. U.S. Government Printing Office, Washington, D.C.

MULRY, M. H. (1992a). Loss function analysis for the Post Census Review (PCR) estimates. Decennial Statistical Studies Division, Bureau of the Census, Washington, D.C.

MULRY, M. H. (1992b). Total error of Post Census Review estimates of population. Decennial Statistical Studies Division, Bureau of the Census, Washington, D.C. (June 1992.)

MULRY, M. H. (1994). Personal communication, January 25, 1994.

MULRY, M. and DAJANI, A. (1989). The matching error study for the 1988 Dress Rehearsal. In *Proceedings of the Social Statistics Section* 704–709. Amer. Statist. Assoc., Alexandria, VA.

MULRY, M. and SPENCER, B. (1991). Total error in PES estimates of population (with discussion). *J. Amer. Statist. Assoc.* **86** 839–863.

MULRY, M. H. and SPENCER, B. D. (1993). Accuracy of the 1990 census and undercount adjustments. *J. Amer. Statist. Assoc.* **88** 1080–1091.

PLEWES, T. (1994). The redesigned current population survey. *Chance* **7** 35–41.

ROBINSON, J. G., AHMED, B., DAS GUPTA, P. and WOODROW, K. A. (1993). Estimation of population coverage in the 1990 United States census based on demographic analysis (with discussion). *J. Amer. Statist. Assoc.* **88** 1061–1079.

ROLPH, J. E. (1992). Testimony in *New York et al. v. Commerce Department et al.* U.S. District Court, Eastern District of New York, 89 CV 3474.

ROLPH, J. E. (1994). The census adjustment trial: reflections of a witness for the plaintiffs. *Jurimetrics* **34** 85–98.

ROYCE, D. (1992). A comparison of some estimators of a set of population totals. *Survey Methodology* **18** 109–126.

SCHAFER, J. L. (1993). Comment on "Assessing between-block heterogeneity within the post-strata of the 1990 Post-Enumeration Survey" by N. Hengartner and T. P. Speed. *J. Amer. Statist. Assoc.* **88** 1125–1127.

SCHAFER, J. L. and SCHENKER, N. (1991). Variance estimation with imputed means. In *Proceedings of the Survey Research Methods Section* 696–701. Amer. Statist. Assoc., Alexandria, VA.

SCHENKER, N. (1988). Handling missing data in coverage estimation, with application to the 1986 test of adjustment related operations. *Survey Methodology* **14** 87–97.

SCHENKER, N., ed. (1993). Special section on the 1990 undercount. *J. Amer. Statist. Assoc.* **88** 1044–1166.

SCHIRM, A. L. (1991). The effects of census undercount adjustment on congressional apportionment. *J. Amer. Statist. Assoc.* **86** 526–541.

SCHIRM, A. L. and PRESTON, J. (1987). Census undercount adjustment and the quality of geographic population distributions (with discussion). *J. Amer. Statist. Assoc.* **82** 965–990.

SINGH, M. P., ed. (1992). Special section on census undercount measurement methods and issues. *Survey Methodology* **18** 1–154.

SPENCER, B. D. (1980a). *Benefit–Cost Analysis of Data Used to Allocate Funds. Lecture Notes in Statist.* **3** Springer, New York.

SPENCER B. D. (1980b). Implications of equity and accuracy for undercount adjustment: a decision-theoretic approach. In *Proceedings of the Conference on Census Undercount* 204–216. Bureau of the Census, Washington, D.C.

STEFFEY, D. L., ed. (1993). *A Census That Mirrors America.* National Research Council, Washington, D.C.

STEFFEY, D. L. and BRADBURN, N. M., eds. (1994). *Counting People in the Information Age.* National Research Council, Washington, D.C.

STOKES, L. and JONES, P. (1989). Evaluation of the interviewer quality control procedure for the Post-Enumeration Survey. In *Proceedings of the Survey Research Methods Section* 696–698. Amer. Statist. Assoc., Alexandria, VA.

THOMPSON, J. H. (1994). Personal communication, February 1, 1994.

TORTORA, R. D., MISKURA, S. M. and DILLMAN, D. A. (1993). Onward towards a 2000 census design: research results. In *Proceedings of the Survey Research Methods Section* 120–128. Amer. Statist. Assoc., Alexandria, VA.

TREMBLAY, A., STOKES, S. L. and GREENBERG, B. S. (1991). Estimation of PES fabrications from quality control data. In *Proceedings of the Survey Research Methods Section* 242–247. Amer. Statist. Assoc., Alexandria, VA.

TUKEY, J. W. (1983). Affidavit in *Cuomo v. Baldridge*, Civil Action 80-4550, U.S. District Court, Southern District of New York.

TUKEY, J. W. (1991). Testimony before the House Subcommittee on Census and Population, Committee on Post Office and Civil Service, Serial No. 102–19, June 27, 1991. U.S. Government Printing Office, Washington, D.C.

UNDERCOUNT STEERING COMMITTEE (1991). Technical assessment of the accuracy of unadjusted versus adjusted 1990 census counts, June 21, 1991. Bureau of the Census, Washington, D.C.

U.S. DISTRICT COURT (1991). Eastern District of New York (Judge Joseph M. McLaughlin), Memorandum and Order 88 CV 3474, 92 CV 1566, 92 CV 2037, April 13, 1993.

U.S. DISTRICT COURT (1992). *New York et al. v. Commerce Department et al.* U.S. District Court, Eastern District of New York (Judge Joseph M. McLaughlin), 89 CV 3474, 92 CV 1566, 92 CV 2037.

WACHTER, K. W. (member, Special Advisory Panel) (1991). Recommendations on 1990 census adjustment. Report submitted to the Secretary of Commerce, June 17, 1991.

WACHTER, K. W. (1993a). The census adjustment trial: an exchange. *Jurimetrics* **34** 107–115.

WACHTER, K. W. (1993b). Ignoring nonignorable effects. Comment on "Hierarchical logistic regression models for imputation of unresolved enumeration status in undercount estimation" by T. R. Belin, G. J. Diffendal, S. Mack, D. B. Rubin, J. L. Schafer and A. M. Zaslavsky. *J. Amer. Statist. Assoc.* **88** 1161–1163.

WACHTER, K. and FREEDMAN, D. (1994). Local heterogeneity in 1990 census coverage. Technical report, Dept. Statistics, Univ. California, Berkeley.

WACHTER, K. W. and SPEED, T. P. (1991). Comment on "Total error in PES estimates of population: the dress rehearsal of 1988" by M. H. Mulry and B. D. Spencer. *J. Amer. Statist. Assoc.* **86** 858–861.

WEST, K. K. (1991). PES Evaluation Project #P5a: analysis of P-sample fabrication from evaluation followup data, 1990 Coverage Studies and Evaluation Series Memorandum #G-2, Statistical Support Division, Bureau of the Census, Washington, D.C. (July 12, 1991.)

WEST, K., CORBY, C. and VAN NEST, G. (1989). Quality assurance for the clerical matching operation in the 1990 Post-Enumeration Survey. In *Proceedings of the Survey Research Methods Section* 699–703. Amer. Statist. Assoc., Alexandria, VA.

WEST, K. K., MULRY, M. H., PARMER, R. and PETRIK, J. (1991). Address reporting error in the 1990 Post-Enumeration Survey. In *Proceedings on Survey Research Methods Section* 236–241. Amer. Statist. Assoc., Alexandria, VA.

WOLTER, K. (1986a). Comment on "Regression models for adjusting the 1980 census," by D. A. Freedman and W. C. Navidi. *Statist. Sci.* **1** 24–28.

WOLTER, K. (1986b). Some coverage error models for census data. *J. Amer. Statist. Assoc.* **81** 338–346.

WOLTER, K. (1987a). Technical feasibility of adjustment. Memorandum to the Undercount Steering Committee, Bureau of the Census, Washington, D.C.

WOLTER, K. M. (1987b). Comment on "Census undercount adjustment and the quality of geographic population distributions" by A. L. Schirm and S. H. Preston. *J. Amer. Statist. Assoc.* **82** 978–980.

WOLTER, K. M. (1991). Accounting for America's uncounted and miscounted. *Science* **253** 12–15.

WOLTER, K. M. and CAUSEY, B. (1991). Evaluation of procedures for improving population estimates for small areas. *J. Amer. Statist. Assoc.* **86** 278–284.

WOLTMAN, H. F. (1991). Final report for loss function analysis for places. Statistical Support Division, Bureau of the Census, Washington, D.C.

WOLTMAN, H., ALBERTI, N. and MORIARITY, C. (1988). Sample design for the 1990 census Post-Enumeration Survey. In *Proceedings of the Survey Research Methods Section* 529–534. Amer. Statist. Assoc., Alexandria, VA.

WOLTMAN, H. F., GRIFFIN, R. A., NAVARRO, A., BATES, L. and MULRY, M. H. (1991). Post-Enumeration Survey Evaluation Project P16: total error model—loss function evaluation. Statistical Support Division, Bureau of the Census. Washington, D.C.

YLVISAKER, D. (1991). A look back at TARO. Technical report, Dept. Mathematics, UCLA.

ZASLAVSKY, A. M. (1993a). Combining census, dual-system, and evaluation study data to estimate population shares. *J. Amer. Statist. Assoc.* **88** 1092–1105.

ZASLAVSKY, A. M. (1993b). Composite estimators for multiple domains and information sources via hierarchical Bayes linear models and Gibbs sampling. Presentation at UCLA Workshop on Model Selection, Bayes Factors, and Sensitivity Analysis. (February 19, 1993.)

ZASLAVSKY, A. M. and WOLFGANG, G. S. (1993). Triple system modeling of census, Post-Enumeration Survey, and administrative list data. *Journal of Business and Economic Statistics* **11** 279–288.

Comment

Ian Diamond and Chris Skinner

From this side of the Atlantic, the amount of attention devoted to the adjustment of the U.S. census can sometimes seem bewildering. Censuses are conducted for many purposes and raise many methodological problems. For perhaps most purposes, the effects of coverage errors in the census seem likely to be minor compared, for example, to the effects of nonresponse in most sample surveys. Nevertheless, there are important uses of census data in Great Britain, as in the United States, where coverage errors do matter, most notably in the preparation of local area population estimates, which are widely used by central government for the allocation of resources. Furthermore, the adjustment issue has particular significance for the 1991 censuses of England and Wales and of Scotland since, unlike in 1981 and before, the estimates of the national under-

Ian Diamond and Chris Skinner are Professors, Department of Social Statistics, University of Southampton, Southampton S017 1BJ, United Kingdom.