more optimistic than Lyberg and Lundstrom that further investment in methodology will pay dividends. One example of methodology that we cited repeatedly in our paper is Zaslavsky (1993a); this work not only advances the state of the art for census undercount estimation, but it serves as a useful case study that could be adapted to other statistical arenas as well. Overall, however, we appreciate their endorsement of our general perspective on the adjustment controversy.

Ericksen, Fienberg and Kadane make few comments directly about our paper. We would simply point out that some of their recent references (Kadane, Meyer and Tukey, 1992; Darroch, Fienberg, Glonek and Junker, 1993) also serve to illustrate that progress is still being made on undercount-related issues, yielding both new theory and new methods.

We understand that in their rejoinder, FW cite a personal communication from us. We offer the following comment in the spirit of "setting the record straight."

In the initial version of his paper, Breiman made a stronger claim about the increasing proportion of unresolved cases in the Evaluation Followup Survey (EFU) when one reads across his Table 12, which we saw as the kind of nitpicking criticism that deserves to be pushed to the margin. Originally, after we pointed to Breiman's curious claim that one might do just as well in imputing for unresolved cases by flipping a coin with probability 15% of heads, we had written, "The higher proportion of remaining unresolved cases in the higher imputed probability categories is explained in large part by the fact that names were not recorded for many PES individuals." However, it turned out that our explanation was inaccurate; although cases without names constituted approximately 70% of the P-sample cases receiving probabilities of having been enumerated of 75–100%, these cases without names were largely excluded from the EFU and so were not reflected in Breiman's Table 12.

We acknowledge that there were a substantial number of unresolved cases in the EFU and that there is remaining uncertainty about the accuracy of the imputation methods. Our essential point is that there is not much to criticize based on available data, which agree with predicted values extremely well (Belin et al., 1993). To attribute our earlier statement to us as if it is our current view is a misrepresentation.

Overall, although we anticipate that our Berkeley colleagues will continue to support one another, we are pleased at the signs of consensus in this exchange.

Rejoinder

Leo Breiman

I thank the discussants. The descriptions of the methods used in Australia, Great Britain and Sweden were interesting and form a compact introduction to the diversity of methods for estimating population counts. They also underline the difficulty of the census undertaking in the United States. The discussion by Ericksen, Fienberg and Kadane and the Belin–Rolph article contain most of the direct comments about my paper.

BACKGROUND

The effort to adjust the census counts was a complex process. After the initial error evaluation, additional errors were discovered, some of which are discussed in my article. Because the original error analysis has not been updated to take these additional errors into account, the widespread impression remains that the adjustment process was proven to produce more accurate counts than the census.

The validity of any such proof is currently in serious doubt. For one thing, errors of various types are now acknowledged to account for the major part of the original national undercount estimate of 2.1%. The initial loss function analysis used earlier estimates of the bias that, on the national level, were too small by at least a factor of 2. The analysis was also flawed by a significant underestimation of sampling variances (Fay and Thompson, 1993; Freedman, Wachter, Cutler and Klein, 1994). There are also questions about the additional local bias due to heterogeneity (Freedman and Wachter, 1994), the errors resulting from smoothing the adjustment factors (Freedman et al., 1993) and many of the assumptions going into the loss function analysis (Freedman, Wachter, Cutler and Klein, 1994).

This careful scrutiny was possible, in part, due to the availability of three sets of numbers: the census counts, the adjustments and the extensive evaluation data. We view the controversy over the adjustment as having some healthy outcomes. Methodology and implementation were openly discussed and debated. More statisticians than ever became aware of the problems of carrying out an accurate census. We hope that the planned single-number census for the year 2000 will allow similar careful scrutiny and discussion.

SUMMARY OF MY REJOINDER

In my rejoinder, I first comment on some statements of the overseas discussants, then on the Ericksen, Fienberg and Kadane discussion, and last, on the Belin–Rolph article. The Belin–Rolph remarks are surprisingly angry and include some serious and totally unjustified personal attacks.

The proposed 1991 adjustments will be referred to as the DSE adjustments or the DSE estimates. My article pointed out that major components of the DSE estimate adjustments are data errors stemming from different sources. None of the discussants disagree with my main conclusion concerning the presence of a large amount of error in the DSE estimates, at least at the national level.

Ericksen, Fienberg and Kadane note that there was uncertainty in some of the evaluation data, with the implication that aspects of the evaluation could have been better planned and timed. I agree. Both Ericksen, Fienberg and Kadane and Belin–Rolph maintain that my results do not have any major implications for the distribution of DSE estimates of population at local levels. They also criticize me for "assuming zero correlation bias." These points are responded to in my rejoinder to Ericksen, Fienberg and Kadane.

OVERSEAS DISCUSSANTS

The overseas discussants disclaim familiarity with the DSE adjustment process. However, their comments contain some misunderstandings. For instance, Lyberg and Lundström believe that because the PES was more accurate than the census, the DSE adjustments will be more accurate than the census counts. This does not follow. As shown in Section 4 of my article, small errors in matching can cause the adjusted counts to be less accurate than the census counts.

Steel likes the 1.7% undercount figure obtained by correcting for the coding error because it is consistent with the 1.85% demographic estimate of the undercount and "consistency offers some reassurance." However, the 1.7% number is defensible only if one ignores all of the evaluations carried out by the Census Bureau. If there is anything consistent going on, it is that almost every source of error investigated further lowers the DSE undercount estimate.

Diamond and Skinner like interval estimates and note that the 95% interval (1.00%, 2.25%) given in Mulry and Spencer (1993) agrees with the demographic estimate of 1.85%. However, the Mulry–Spencer article does not correct for the large coding error and other errors found since 1991. Thus, the interval cited is not accurate in terms of current information.

ERICKSEN, FIENBERG AND KADANE

Ericksen, Fienberg and Kadane (EFK) open by making the point that the census had many errors. The first 10 pages of the report of Special Advisory Committee members favoring adjustment also makes this point (Ericksen, Estrada, Tukey and Wolter, 1991). So does a 95-page appendix to this report. About one-third of Fienberg's article on adjustment (Fienberg, 1993) is spent describing errors in the census. Referring to Freedman, Wachter and myself, EFK say: "That the census is replete with errors and that the errors have differential impact on minorities never seems to be addressed or acknowledged by these authors."

The first lines in Wachter (1993a) are: "The 1990 census had flaws. It missed, net, between one and three percent of the population. It missed more men than women. It missed more blacks than whites. These facts are not in dispute." Wachter's article appears in the same journal issue as Fienberg (1993). This issue had four short articles on adjustment. It is hard to imagine circumstances under which EFK could have overlooked Wachter's statement.

Ericksen, Fienberg and Kadane keep raising the issue that the census had flaws, but nobody disagrees with them. No knowledgeable statistician I know denies that the census had many errors and that there was a differential minority undercount. Neither do I. Now that we are all agreed, let us get on to the fundamental question: would the DSE adjustments produce more accurate counts than the census?

The point of my article is that the DSE adjustments are mainly a reflection of bad data. Since an error-filled adjustment will only superimpose more noise on the census counts, EFK have to face the central issue: is it or is it not true that the DSE estimates are mainly a reflection of bad data?

Here are the points that EFK make regarding this central issue:

1. My final estimate is not believable.
2. My analysis is exaggerated and extreme.
3. I used questionable data.
4. I assumed zero correlation bias.

5. The Bureau's evaluation is better.
6. I worked on the wrong problem.
7. My results do not make much difference.

I will take these up, one at a time.

1. My final estimate is not believable.
Ericksen, Fienberg and Kadane begin with a statement from my article, which reads: "The largest part of the original undercount estimate is due to bad data and processing error—80% on the national level." They claim this has the following implication: "In Breiman's terms, he believes that the correct estimate of undercount may be as low as 1 million." Then they go on to show that the true undercount could not be as low as 1 million.

Their logic is wrong and they have overlooked the statement in my conclusion section which reads: "The results of this study should not be taken to mean that I believe that the true 1990 undercount is as low as 0.4% [1 million] or even 0.9%. My focus was on whether the 1990–1991 DSE process produced reasonable estimates of the true undercount. To that, my answer is no, there were simply too many sources of error."

2. My analysis is exaggerated and extreme.
To quote EFK, "Many of Breiman's judgments appear to be exaggerated and his evaluations extreme." They give a single example to support this statement. The example is followed by the sentence "Breiman's arguments here seem strained at best." Now strained is quite a demotion from exaggerated and extreme. Still, since this is EFK's only shot, let us see how strained my argument is.

The example is brief and occurs in my Section 5.5 titled "Reliability of interview data." As part of the Census Bureau's evaluation, a small fraction of the PES households were reinterviewed. This was called the Evaluation Followup Survey (EFU). The example essentially starts with the sentence in my article reading "Match status could be changed from the PES production match status only if new, relevant and reliable information regarding a case was present in the EFU interview." This is followed by a table showing what percentages changed match status as a result of the EFU data.

Ericksen, Fienberg and Kadane go through some computations to show that, although individual cases may change match status, the marginal totals in each match category remain about the same. Having established this, they note, with some triumph, that "the impact of these changes is minimal." This totally misses the point. A person changes match status only if the PES interview information for that person is judged unreliable compared to the EFU information. Thus, the percentages of people changing match status gives a measure of the reliability of the production PES information, and that is what Section 5.5 is about; EFK's example is a complete misreading of a minor issue.

3. I used questionable data.
EFK write that my analysis is based largely on the three sources of evaluation data: the rematching study, quality control and the field reinterviews (EFU). This is only true if "largely" is given a generous interpretation. The one-million-person coding error is not connected with the evaluation data. The later rematching of 104 selected blocks gave another error correction of 250,000 persons.

They note that the EFU data was gathered 5–6 months after the initial PES and that this delay would cause uncertainty in the data. I agree, and I also wish that the EFU had occurred earlier, but there was good reason for the timing. The PES follow-up took place about 3–4 months after the initial PES. The EFU could not be carried out until the PES follow-up was over.

Ericksen, Fienberg and Kadane imply that the effect of the uncertainty would be to inflate the error estimates. There is no evidence supporting this, and some to the contrary. Since each EFU interviewer carried around the previously completed PES interview forms for the households being reinterviewed, disagreements between the EFU and the PES would, if anything, be biased low (see Biemer and Forsman, 1992). At any rate, the EFU provides the only reinterview field data available and was used both by the bureau in their evaluation (which EFK approve) and myself.

4. I assume zero correlation bias.
If some persons avoid official surveys, they will be difficult to count both by the census and by the PES. The capture–recapture assumptions will not apply, and the DSE estimates will be biased low. This is called correlation bias. Now, the argument is that, although data errors tend to cause an overestimate of the undercount, such errors are largely canceled out by the effect of correlation bias (see Ericksen, Estrada, Tukey and Wolter, 1991).

Both EFK and BR are concerned that "zero correlation bias is assumed." This is not correct. Zero correlation bias is not assumed. I willingly admit that in places correlation bias is likely. However, correlation bias is irrelevant to my study. What I look at is how much of the DSE undercount estimates is attributable to bad data. This has no relation to what is assumed about correlation bias.

5. The Bureau's evaluation is better.
Early on in their comments, EFK say, referring to the Bureau's evaluation, "This evaluation was sum-

marized by Mulry and Spencer (1993)." Later, they add: "We prefer to rely upon the evaluation made by the Census Bureau, partly because they have studied the questions more thoroughly, but also because we believe them to be more objective."

The Mulry–Spencer article is not a summary, but a selective extract from the bureau's 1991 evaluation report [P16]. For instance, the estimated bias due to bad data is not reported separately as it is in [P16]. Instead, the only bias figures given are the lower ones obtained by adding a problematic estimate of correlation bias to the data error bias. Thus, the reader is not given the [P16] information that there is a 0.7% bias at the national level due to bad data nor are they given the corresponding [P16] information for the evaluation strata.

The article, although published in late 1993, does not include corrections for the one-million-person coding error discovered by the Bureau two years earlier. It does not correct for other errors detailed in my article that the Bureau later incorporated into its intercensal error analysis (Mulry, 1992b). These include the late late census data and the rematching of 104 blocks.

The bureau has also admitted that they overlooked the new-out-of-scopes problem and to making errors in computing the census day address error (see the Appendix to my article). These latter two, along with many other error sources, were discussed in my article but not in the Bureau evaluation nor in Mulry and Spencer (1993). EFK's view regarding thoroughness and objectivity is not well founded. They are putting their reliance on an analysis known to both the Census Bureau and other statisticians as outdated and erroneous.

6. I worked on the wrong problem.

"... Breiman's paper seems misdirected." "... Breiman focused his time and energy on the wrong problem." "Because Breiman studied the wrong problem ... we do not feel that his conclusions matter greatly."

It is the wrong problem because "Breiman studied ... the national net undercount as opposed to the distribution of this undercount ... ." Not so—my article explicitly gives the undercount estimates and the effects of the errors on them for the 13 evaluation strata (see Table 16). As EFK know, the only evaluation data available was at the level of these evaluation strata. These data were used by the bureau and myself to see how much of the DSE estimates at the national and stratum level could be attributed to bad data.

If I studied the wrong problem, so did the Bureau, but EFK never claim, either in this discussion or in any of their past articles and reports, that the Bureau studied the wrong problem. They do not explain why they believe that the Bureau studied the right problem, but I studied the wrong one. Perhaps EFK think that if they say I am working on the wrong problem often enough, then somehow my conclusions will go away.

7. My results do not make much difference.

Now EFK go to their final line of defense: even if I studied the right problem, my results do not make much difference. Referring to the undercount estimates in the evaluation strata, they say "his numerical results are not greatly different from those of the bureau [DSE adjustments], especially when we consider between-area differentials."

They then try to show that the differences between the DSE estimates and the error-corrected estimates do not matter much; in particular, that the minority–nonminority undercount differentials remain about the same. To do this, they use a somewhat convoluted procedure leading to the conclusion that the differences matter in only one half of the country. One half of the country is a pretty big slice.

Since EFK consider the minority–nonminority differential to be an important indicator, let us see what the effect is using a simple computation. We take as our measure the estimated population proportion of the five minority strata. Using the DSE estimates, the increase in this proportion over the census is 51% larger than the change computed using the error-corrected estimates. A difference of 51% can hardly be called minor.

To see what effect the corrections have on the stratum-level distribution, define the population share of a stratum to be its fraction of the total population estimate. A standard measure of the difference in two population distributions is the sum over the strata of the squares of the differences of the shares. Using this measure, the difference between the DSE adjustments and the census is 74% larger than the difference between the corrected adjustments and the census. By any definition, 74% is substantial.

The right problem was studied and the differences matter. The errors listed in my article significantly affect minority–nonminority differentials and population shares in the evaluation strata. They are likely to have even more effect at lower levels of aggregation. Furthermore, the strata most affected by the errors were the minority strata (see my Table 16). The implication is that the DSE adjustments are likely to be the worst just where we would want them to be the best.

Conclusions about EFK

The above summarizes EFK's comments on my article and gives my rejoinder. The Census Bureau, in its original [P16] analysis, conceded that about 30% of the national undercount estimate was due to bad data. Ericksen, Fienberg and Kadane do not seem to be aware that in 1992 the Bureau's estimate of undercount, corrected for bias due to data errors, dropped to 0.9%, a tacit admission that almost 60% of its original undercount estimate was due to bad data (Mulry, 1992b). My estimate of 80% is more realistic. Regardless of how error-filled the census is, you cannot solve the problem by combining it with another error-filled set of numbers.

BELIN AND ROLPH (BR)

The Belin–Rolph comments on my article are sharply critical. I will go through their remarks section by section before responding to their accusations of professional misconduct. Here is a list of subjects covered in their sections:

Section 5.1, the source of my article;
Section 5.2, a contradiction, national versus differential;
Section 5.3, a contradiction, two ways of looking at matching error;
Section 5.4, obfuscation in analyzing fabrications;
Section 5.5, the assumption of zero correlation bias;
Sections 5.6, 5.7 and the Appendix, two tables of imputation data;
Section 5.8, my dealings with the Census Bureau.

Section 5.1 notes that my present article has its roots in a manuscript prepared for the adjustment lawsuit. Quite true, but so?

Section 5.2 refers to a supposed contradiction in my analysis. The contradiction is that, even conceding that the estimates aggregated to the national level are largely due to errors, it has not been shown that the estimated population distributions at the local level are affected by the errors. This point was raised by EFK and responded to above.

This section has a number of questionable remarks. For instance, in the first paragraph BR think they have discovered one of my big ideas: "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." Neither this "main point" nor anything resembling it occurs anywhere in my article.

Further on BR ask "Why, then, does he not label evaluation poststrata as minority or nonminority?" Table 4 in my article, which defines the evaluation strata, clearly labels each stratum as minority or nonminority. Belin and Rolph continue: "To allow readers to see the differential minority undercount, we reproduce Breiman's Table 16 with information on minority status added...." Not so! My Table 16 shows the initial 1991 adjustments and the substantial reductions in differential minority undercount when the 1991 adjustments are corrected for errors. This information is eliminated (without comment) in the BR "reproduction."

Next BR say that "after accounting for the computer error, the Census Bureau's loss function calculations favor the adjusted over the unadjusted counts at the state level" and give some references in support. None of their references do what BR claim. Mulry (1992b) deals with the intercensal adjustments. These differ considerably from the original DSE adjustments. Mulry and Spencer (1993) do not account for the computer error; Fay and Thompson (1993) give a review of the Mulry (1992b) work; and Zaslavsky (1993a) is based on the Mulry (1992b) adjustments.

Section 5.3 refers in its title to another contradiction in my work. They state: "... particularly in his discussion of matching error, Breiman does not take into account that errors can cancel." Of course I take this into account and explicitly discuss it in Section 6.1. Then they say: "Curiously ... Breiman then uses the Census Bureau's calculations of the effects of matching errors in his Table 15." As far as I can make out, the contradiction referred to is that I cite both disagreement rates over cases and marginal disagreement rates. The reason for this is carefully explained in the first three paragraphs of my Section 5.1 on matching; BR have either overlooked the technical reasoning or do not understand it.

In brief, the differences in the marginal totals of the match–rematch data for the evaluation strata are direct estimates of the matching error and were used by the Bureau and myself. However, I show that marginal disagreement rates at highly aggregated levels tend to average out and are not good indicators of marginal disagreement rates at less aggregated levels. The disagreement rates over cases are better indicators and are cited for this reason.

Section 5.4 on fabricated interviews refers in its title to obfuscation in my work. The section begins with a quote, supposedly from my article. It is not in my article nor in the original court document. The sentence in my article following the extract from [P6] does not have the faintest resemblance to the quote that BR attribute to me. Then BR imply that I claim the undercount is overestimated by millions due to undetected fabrications in the PES. This is not what

I say, and if readers are in doubt, I would request that they review my Section 5.2.

Further BR comments indicate lack of careful reading. For instance, they state that "quality-control checks suggesting 2–5% interviewer fabrication does not translate to 2–5% of cases entered in the P-sample database being fabrications." Of course not, but they have the 2–5% numbers wrong. The 2–5% numbers come from report [P6] and have nothing to do with quality control. This confusion leads them on a merry chase.

Their concluding statement is: "Breiman discounts the possibility that quality control... would have been able to catch fictitious enumerations." No such discounting occurs. Nobody, least of all the Census Bureau, claims to have a foolproof way of detecting all fabrications. The question is how many got through the PES quality control procedures. What I present is the evidence given in the P-studies concerning this question.

Section 5.5 raises the issue of assuming zero correlation bias. See my response to EFK on this subject.

Sections 5.6 and 5.7 and the Appendix, a third of BR's comments, are about two tables. These two, Tables 12 and 13, appear in my Section 5.7 on imputation and form a minor part of the study. Table 12 contains a summary of the P-sample (PES) imputation evaluation data, and Table 13 a summary of the E-sample (census) imputation evaluation data. They are taken directly from Tables 3.1 and 3.2 of Bureau report [P3].

Section 5.6 is titled "Breiman's Table 13 as 'Bad Data'." Belin and Rolph and I agree that Table 13 does not look good for imputation. They believe that this table is erroneous and have a long story about why I am to be blamed for reproducing the [P3] version. First are two memos Belin wrote to his superiors in the Bureau, but which he admits might have been difficult for me to know about. Next, I should have corrected the error because Belin told a referee about the corrections and the referee was supposed to tell me. There were no such referee comments.

It is remarkable that BR spend so much energy criticizing me for not knowing about and not correcting a supposed error in a table taken from an official Census Bureau report. They are hassling the wrong person. Their argument is with the Bureau and not with me. If the table is wrong, then report [P3] needs correction, and that should be done by the Bureau.

Section 5.7 concerns Table 12 and its collapsed version, Table 11. The section title refers to the "prism" through which I view these tables. They begin with a misreading of my paper. In their first paragraph, BR note that in my Table 11 the percentage of matches among the resolved is 12/39 = 30.8%, while the percentage given in Belin et al. (1993) is 31.6%. Belin and Rolph comment, in a quite annoyed way, that this difference could not be due to rounding. They are right. As stated in my article and in report [P3], the numbers in Table 11 are weighted to the nation. In the Belin article the results are unweighted.

Next BR exhibit a truncated version of Table 12. The evaluation data on P-sample imputations contains almost 60% unresolved data, analogous to nonresponses in a survey. The BR game is to deal with the unresolved data by pretending that it is not there, ergo, the truncated table. Then things look better for imputation. The Bureau does not see it the BR way. Report [P3], in summarizing Table 12, states: "Thus, for P-Sample persons, the imputation process is consistent with EFU results. However, the high percentage of unresolved persons in the EFU (58.55 percent) may limit the utility of this result."

After all the time BR spend on these two tables, this is what we are left with: Table 12 is okay—no questions about its accuracy. Belin and Rolph need to work with the Bureau to resolve their claims about Table 13. They end their discussion of these two tables with the revealing statement that for them the real issue is not the data and substance underlying imputation, but personalities and politics.

Section 5.8 contains a lengthy lecture based on the sentence in my Appendix reading "I have been unable to obtain from the Bureau any more specific information regarding their method for computing census day address error." The section consists mainly of moral disapproval of my efforts to understand how the census day address error was computed and little that is relevant to the substantive issues. I stand by the accuracy of my statements.

In this section, John Thompson, a senior Bureau statistician, is quoted as saying that "[Breiman's] numbers appear to be correct..." but that the Bureau should have the chance to comment and give their interpretation. I agree. It would be instructive to have a Bureau viewpoint represented.

BELIN AND ROLPH'S ACCUSATIONS

Belin and Rolph begin their review of my work by accusing me of "unprofessional practices" and in the conclusion they state "...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."

This is an quite an indictment and should be accompanied by convincing proof of unethical behavior and distortion of facts. Belin and Rolph have not come up with a single fact that I have been careless with, nor any instance of unethical behavior. On the other hand, they have gotten their facts wrong, have been careless in reading, have found contradictions and obfuscations where there are none and have spent most of their time on irrelevant side issues and morality mongering.

Rejoinder

D. Freedman and K. Wachter

1. INTRODUCTION

Census adjustment is not an easy topic. We are grateful to the discussants for their efforts at clarifying the issues. One other idea will not be controversial: Rob Kass and Ram Gnanadesikan deserve thanks for putting this exchange together and bringing it to a successful conclusion.

The commentaries fall naturally into two groups, those from outside the United States and those from inside. It is valuable to have perspectives gained from experience in other countries. We marvel, naturally, at errors measured in hundreds, which Lyberg and Lundstrom attribute to Sweden's PIN-keyed registers. Australia as described by Steel makes an interesting contrast to Britain (Diamond and Skinner), in terms of the trust accorded to results from demographic analysis in Britain and the distrust in Australia—even though Australia has effective monitoring of international migration, which removes one of the chief components of uncertainty in demographic analysis for the United States.

Belin and Rolph (BR) have a free-spirited and wide-ranging commentary which reviews many previous exchanges on the census. Much as we like the authors, we differ with them on readings of the technical and historical record and on matters of scientific principle. With respect to the census, Ericksen, Fienberg and Kadane (EFK) are among the oldest and most familiar of our opponents; but on this occasion, as we shall explain, their critique is off the mark entirely.

According to the rules of engagement, we do not comment on BR's rejoinder and they do not comment on ours, so we get the last word in this exchange—on these pages of this journal. Silence cannot be interpreted as consent: we are sure that BR and EFK will continue the argument in some other forum.

The Census Bureau's latest thinking on the 1991–1992 adjustments is described in Fay and Thompson (1993). Our discussants frequently refer to this paper for arbitration, and we shall too. We hope Bob Fay and John Thompson will not mind such close textual analysis.

Despite the scope of BR's remarks, our paper was not really about the bottom-line question: whether adjustment would have made errors in state population shares better or worse. It was, rather, about a "wild card" in the Census Bureau's assessments of state and local coverage error: heterogeneity. Heterogeneity was omitted from the bureau's loss function analysis. Statisticians on all sides have been arguing ever since what kind of difference that could have made.

In our paper, we measured the difference that heterogeneity does make, in a context that allows an exact answer. We used the same loss function that the Bureau did, with proxy variables instead of undercounts. In our context, the adjustment factors are known with perfect accuracy, so that errors of adjustment are due purely to heterogeneity, and loss itself can be calculated. We found the following:

1. The omission of heterogeneity does bias the estimated risks.
2. Depending on the proxy, the bias can be small or it can be large.
3. The bias can go either way, for or against adjustment.

In particular, we established that loss function analysis can be strongly biased in favour of adjustment; EFK and BR react quite critically to this finding. Before we answer them, let us recall the larger picture in which such arguments take their place.

2. BACKGROUND

Would the proposed adjustment of the 1990 census, or of the intercensal estimates, have improved the accuracy of population shares held by the various states? "Loss function analysis" attempts to balance errors in the census against errors in adjustment, and it seems to be the principal statistical argument that adjustment would improve on the census (BR,