# Philip B. Stark, Ph.D.

1157 Cragmont Ave
Berkeley, CA 94708-1641

Phone: 510-540-0703
FAX: 510-486-1157
email: stark@stat.berkeley.edu

26 June 1998

The Honorable Carolyn B. Maloney
Ranking Minority Member
Subcommittee on the Census
Committee on Government Reform and Oversight
2157 Rayburn House Office Building
Washington, DC 20515-6143

Thank you for your questions of 13 May 1998. I shall answer them by number.

1) *Can you tell us about a statistical or scientific activity that you've worked on that either worked perfectly the first time you tried it, or that didn't work as well as you had hoped the first time so you abandoned the idea altogether without making an effort to improve it?*

It has happened on several occasions that I had a conjecture I hoped was true, tried to prove it, found a counterexample, and immediately abandoned it. It has also happened several times that the first approach to a problem I tried worked perfectly. Sometimes a technique "almost" works, and I try to improve it. The sampling-based (DSE) approach to adjusting the census did not "almost" work in 1990. The problems with the DSE are not minor details that can be repaired by increasing the sample size or other incremental refinements: the experience from 1990 suggests that the approach is unworkable, because its biases are so large. The biases come from failures of the assumptions on which the method is based, and from insurmountable practical problems in implementing the approach on such a large scale. The situation is analogous to finding a counterexample to a conjecture. Science progresses by finding counterexamples and publishing them, so that others can pursue more promising approaches. The experience in 1990 seems to be a counterexample to the hypothesis that DSE can be used to improve the accuracy of the census.

2) *Despite the fact that the Census Bureau made improving the count among minorities a major goal of the 1990 Census, the 4.4 percent differential in the 1990 undercount between Blacks and non-Blacks was the highest ever recorded. Experts have repeatedly said that spending more money on traditional methods will not reduce this differential. If not through statistics, how do you propose to reduce this differential?*

First of all, the 4.4 percent figure you quote is not a fact---it is an estimate, and I am unsure of its source. I believe it to be based on demographic analysis, which has uncertainty of its own. The true undercount differential is unknown. Regardless, every

set of data has some limit on its accuracy. The 1990 sampling-based adjustments really seem to make the accuracy worse, not better. The primary problem with the census is non response. The single best thing that could be done to improve census accuracy and decrease its cost is to motivate the public, especially undercounted groups, to fill out and return their census forms in a timely way. This is an area in which elected public leaders can make a big contribution.

If the question were "we can afford to spend x dollars on the census---how can we get the highest accuracy at that cost?," the answer might involve sampling, at least sampling for non-response follow-up. However, the results would probably be less accurate than a full head count.

3) *You have mentioned your concerns about block level accuracy. Can you discuss [your] thoughts on the accuracy of census numbers at the state level if Dual System Estimation is used in 2000? Do you have any evidence that suggests that the census counts will be more accurate at the state level in 2000 if DSE is not used?*

My testimony concerned state-level accuracy, not block-level accuracy. The evidence that adjusting the 1990 census using DSE would have made the accuracy of state shares worse is quite strong---see the "Technical Notes" section of my 5 May 1998 written testimony. Based on that evidence, and my review of the details available for the proposed 2000 ICM, I believe the 2000 census counts would be more accurate at the state level if DSE is *not* used. Many serious problems with the 1990 DSE are present in the 2000 ICM, so the failure of the 1990 DSE is evidence that the proposed 2000 ICM would be less accurate than a simple census.

4) *Secretary Mosbacher, in testimony before both the House and the Senate, said that the Post Enumeration Survey would make the majority of the states more accurate. Is that statement correct? If so, why is his testimony so at odds with your testimony?*

I do not have a copy of Secretary Mossbacher's testimony. I would be happy to read it and reply in detail if you wish. I believe that using the 1990 Post-Enumeration Survey and Dual System Estimate would have made state shares less accurate.

5) *The 1990 census cost 20 percent more per household in real dollars than the 1980 census. The 1980 census cost twice as much per household in real dollars as the 1970 census. That is an increase in real dollar cost per household of 250 percent with no improvement in the differential undercount. Does that suggest to you that spending more on traditional methods will reduce the differential undercount?*

I think that there must be ways to motivate more of the population to respond to the census by mail. That would improve accuracy, and cut follow-up costs. Whether or not it

would decrease the differential undercount is an empirical question that I cannot answer *a priori.*

6) *Demographic analysis showed higher undercounts of African Americans than the undercounts demonstrated by the Post Enumeration Survey. That suggests that the Post Enumeration Survey understates, not overstates, the undercount, especially for minorities. In other words, isn't it likely that the 1990 census missed more African-Americans [than] would have been added back into the census by the Post Enumeration Survey?*

I think the primary issue is shares, not totals. Shares can be worse if people are put in the wrong place than if no adjustment were made. For example, suppose there are only two states, A and B; only two ethnicities, pink and green; and no gender. Suppose the census finds:

| State | pink | green | total |
|---|---|---|---|
| A | 100 | 10 | 110 (55.6%) |
| B | 80 | 8 | 88 (44.4%) |
| total | 180 (90.9%) | 18 (9.1%) | 198 |

Suppose we know (from some perfect demographic analysis, perhaps) that nationwide, 3 pink people (1.7 percent) and 1 green person (5.6 percent) are missing. Then the true population fraction of pink people is 90.6 percent, the true population fraction of green people is 9.4 percent, and the differential undercount rate is about 3.9 percent. The DSE says 2 pink people and 1 green person are missing, all from state A. It would appear that adjusting the counts is a good idea, because it makes the totals closer to the Demographic Analysis. The adjusted counts would be:

| State | pink | green | total |
|---|---|---|---|
| A | 102 | 11 | 113 (56.2%) |
| B | 80 | 8 | 88 (43.8%) |
| total | 182 (90.5%) | 19 (9.5%) | 201 |

The percentages of pink and green people in the overall population in the adjusted census are closer to those in the demographic analysis. Suppose the DSE adjustment is mostly bias in the DSE. In fact, the 3 missing pink people are missing from state B, and the 1 missing green person is missing from state A. Then the truth is:

| State | pink | green | total |
|---|---|---|---|
| A | 100 | 11 | 111 (55.0%) |
| B | 83 | 8 | 91 (45.0%) |
| total | 183 (90.6%) | 19 (9.4%) | 202 |

Adjustment made state shares less accurate (they are off by 1.2 percent, while the census was off by only 0.6 percent), even though it made the totals more accurate.

The situation is the same for the 1990 DSE: most of the adjustment is bias, and it is implausible that the adjustment put the missing people more or less where they belonged. As a result, the adjusted state shares are probably less accurate than the census state shares. Even if the DSE added the right number of people nationally, it probably put them in the wrong places. The result is less accurate state shares.

7) *You have talked a lot about bias in the Post Enumeration Survey but have not talked much about the bias in the census. The differential undercount measured by demographic analysis shows that the bias in the census is quite real. If there is no Integrated Coverage Measurement, is it not the case that this bias in the census will continue?*

The census does seem to be biased at the level of national totals, and is probably biased at the level of state shares. The ICM is unlikely to fix the bias in the census. It just adds different biases.

8) *Do you believe that it is acceptable for the census to consistently miss certain segments of the population -- [African] Americans, Latinos, Asian Americans, poor people in rural and urban communities -- at greater rates than the White population? If that is not acceptable, what do you propose be done to reduce the differential undercount? Can you offer any evidence that [your] proposal(s) will reduce the differential undercount?*

It is a regrettable fact that the census makes mistakes. It is a regrettable fact that DSE does not fix those mistakes---it just makes different mistakes. I wish the differential undercount could be eliminated, or at least reduced. The best way to decrease the differential undercount is to motivate undercounted groups to respond to the mail-out census questionnaires.

9) *It has been stated that one of the faults of the 1990 PES was correlation bias. Can you explain correlation bias? I understand that it is the likelihood that the people missed in the census may be the same people missed in the PES. Said another way, both the census and the survey miss the same people, for example, young Black makes. How does correlation bias affect the accuracy count of those traditionally undercounted, Blacks, Hispanics, Asians, Native Americans, renters?*

"Correlation bias" is a label for two kinds of failure of the hypotheses on which the DSE is based: (i) being "caught" by the census can influence the chance of being "caught" by the PES, and (ii) different individuals within a post-stratum have different chances of being caught either by the census or by the PES. The existence of people who are unreachable by both the census and the PES is a failure of the second kind. Correlation

bias does not affect the accuracy of the census; it is a source of error in DSE adjustments. Some demographers say that such unreachable people are especially likely to be in dense inner cities, which often have large minority populations. Because such people are "caught" neither by the census nor by the PES, DSE adjustment does not take them into account.

10) *Wouldn't the only risk of correlation bias be minimization of the undercount rather than an overestimation of the undercount?*

No. If correlation bias is different in different places, that can reduce the accuracy of state shares estimated by the DSE.

11) *In testimony before the Senate Committee on Governmental Affairs approximately one year ago, Dr. Lawrence Brown, Professor of Statistics at the University of Pennsylvania, stated that, "Statistical sampling methods can be used in an effective and objective way to assist the census process." Do you agree with Dr. Brown's statement? If you disagree, please explain why.*

I agree. For example, I understand that sampling methods are used successfully by the Census Bureau for quality control of interviews.

12) *Dr. Lawrence Brown also testified before Senator Thompson that the Sampling for Nonresponse Follow-up plan "is an objective procedure all the way around [and] has a very good chance of working as desired." Do you agree with that statement? If you disagree, please explain why.*

The plan appears to be objective (although it involves many *ad hoc* choices), but it seems unlikely to reduce the biases in the census. I believe that sampling for non-response follow-up will decrease data quality, and introduce a new source of error into DSE adjustments. However, I am more troubled by the sampling-based DSE adjustments than by sampling for non-response follow-up.

13) *In addition, Dr. Brown testified that the Census Bureau's 2000 census plan had been "drastically simplified and improved ... [these changes] make it possible to believe that that the Integrated Coverage Measurement might work as well as desired to correct the undercount." Do you agree with that statement? If you disagree, please explain why.*

I agree that the current proposal for the 2000 ICM is simpler than some past proposals, and that the data analysis is simpler in some respects than the 1990 DSE. The statement you cite is hardly an endorsement of the planned 2000 ICM: it is possible to believe that the proposed ICM might reduce the undercount, but I am convinced that will make state shares less accurate. For the ICM to improve state shares would require an implausible

cancellation of large errors. Moreover, there will never be a way to tell whether such a cancellation occurrs. Therefore, it cannot be shown that the ICM improves the census.

14) *With regard to concerns that the Integrated Coverage Measurement process could be manipulated to achieve a particular outcome in terms of the population counts, Dr. Brown testified that, "if all of this planning is done in advance, it is very, very hard for me to see how one could direct these subjective decisions towards any desired goal." Do you agree with Dr. Brown that if the procedures and protocols for the Integrated Coverage Measurement are set forth in advance and subject to expert and public scrutiny, that it is very unlikely that the sampling and statistical estimation process will be subject to manipulation, possibly for political advantage? If you disagree, please explain why.*

I have no opinion about this.

15) *Dr. Brown also testified that even after the non-response follow-up phase of the census is complete, there "would still [be] the undercount problem of those people who just refuse to be counted or are very difficult to count." Do you agree with that? If you disagree, please explain why.*

I agree.

16) *With regard to the post-enumeration survey in the 1990 census, Dr. Brown testified that many of the difficulties with the procedure "can be traced to the fact that the PES sample was much too small to support the kind of objective, reliable analyses that are desired." Do you agree with that? If you disagree, please explain why.*

The sample size was inadequate, but there were many other serious problems with the analysis, such as the biases discussed in my 5 May 1998 testimony. Increasing the sample size would not decrease those biases. It would probably exacerbate them.

17) *The size of the sample in the Integrated Coverage Measurement (ICM) is 750,000 households. Is that a proper size for such an endeavor?*

There is no proper sample size for the ICM, because the main problem is bias, not sampling error.

18) *The results of the PES in 1990 showed that census was less accurate than its predecessor. That result was confirmed by demographic analysis, which has been performed on every census since 1940. We certainly know that the 1990 census was much more expensive than the 1980 census. Do you agree with the conclusion that 1990 was also less accurate than 1980?*

Because demographic analysis does not estimate state shares, it is not possible to tell from demographic analysis whether the 1990 census was less accurate than the 1980 census at the level of states, or for state shares. Because of the uncertainties in demographic analysis, it is not clear whether the 1990 census was less accurate than the 1980 census at the national level, but the evidence suggests that at the national level the 1990 census was the second most accurate census, if not the most accurate census, in U.S. history.

19) *Please explain the difference between net over- or undercount in the 1990 census count and actual over- and undercounts (mistakes) made [in the] 1990 count. I know that a net undercount of 1.6% sounds relatively small but for census purposes, aren't those 26 million mistakes a concern?*

Net undercount is the number of people counted erroneously, minus the number of people who were not counted. Both of these terms are computed at the block level, not at the national level. That is, the same person, who really lives somewhere in the US, can contribute both an erroneous enumeration and a gross omission, if his or her address is incorrect in the census (the person will be a gross omission where the person really lives, and an erroneous enumeration at the incorrect address). The importance of the two errors depends on the geographic level one cares about: at the block level, both errors are important, but for such a person, the errors cancel at the national level. Overall, the gross omissions and erroneous enumerations in the census cancel to some degree, although not perfectly, when aggregated to states or the nation. The figure of 1.6% you cite appears to reflect some of the revisions in the PES since it was first published; I believe the figure of 26 million mistakes may not reflect those revisions. The large size of the revisions should make such estimates suspect.

20) *I understand that improvement in the average does not necessarily mean that there will be improvement in every case. In 1990, there was criticism about the strata being broken down by region. If statistical methods are used in 2000, with strata broken down by state in 2000, can we expect more states with improved accuracy than there were in 1990?*

No. First of all, bias is probably more important than the sampling error. The bias in 1990 was so large that, in my opinion, the 1990 DSE was not trustworthy. I have not seen anything in the 2000 plan that would reduce the level of bias to the point that adjustment reasonably would be expected to improve census accuracy. Furthermore, even though the proposed sample size is larger, the number of post strata is also larger, so there is a tradeoff that might increase the sampling error too.

21) *Representative Sawyer pointed out that the longer the Census Bureau is in the field, the higher the error rate in the information collected. I believe that information came from one of the many GAO studies he and his Republican colleagues commissioned. You have stated*

*your concern about the Census Bureau not [being] in the field for enough days in the 2000 plan.  Can you explain the difference in opinion?*

The quality of data will suffer if the Census Bureau tries to work so quickly that it uses poorly trained or less competent field workers, or allows too little time for it to be possible to do their work well.  Data quality will also suffer if too much time goes by, because people move and memories fade.  Therefore, I see no contradiction.

22)  *In order to address the problem of declining public response, the GAO suggested exploring a radically streamlined questionnaire in future censuses.  Would you give us your thoughts on how effective this approach might be in increasing response, and also its effect on perhaps diminishing the usefulness of census data?*

Everyday experience suggests that it is easier to get 5 minutes of someone's time than 2 hours.  Data from a shorter questionnaire could be less useful.

23)  *In its 1992 capping report on the 1990 census, the GAO concluded that "the results and experience of the 1990 census demonstrate that the American public has grown too diverse and dynamic to be accurately counted solely by the [traditional] 'headcount' approach and that fundamental changes must be implemented for a successful census in 2000."  Do you agree with that conclusion?  If you disagree, please explain why.*

I believe that a headcount is the most accurate method available.  Perhaps someday someone will devise a better approach, but the 1990 experience indicates that the DSE is less accurate than a headcount.

24)  *After the 1990 census, GAO concluded that "the amount of error in the census increases precipitously as time and effort are extended to count the last few percentages of the population ... This increase in the rate of error shows that extended reliance on field follow-up activities represents a losing trade-off between augmenting the count and adding more errors."  In the last months of the follow-up efforts in 1990, the GAO estimated that the error rates approached 30 percent, and that this problem was probably exacerbated by the use of close-out procedures.  This appears to be a problem inherent to the methodology of the 1990 census.  Do you agree?*

*Do you have any information on the error rates for information gathered using close-out procedures?*

*Even if sampling is not perfect, isn't its error rate well below the levels for the last percentages of the population using more traditional follow-up procedures?*

*If this is the case, then doesn't that logically lead to GAO's and the Commerce Department's Inspector General's conclusion that sampling at least a portion of the nonresponding households would increase the accuracy and decrease the cost of conducting the census?*

The problem in reaching the last few percent does not go away with sampling---one still needs to reach the last few percent of the sample, or the same kinds of errors occur and are magnified. The likely cost savings from a 90 percent sample with complete follow-up in the sample, versus the 1990 approach to head counting, seems rather small, and accuracy would probably suffer. If follow-up within the *sample* is incomplete, the resulting errors are just magnified by the sampling ratio. Only if follow-up within the sample is truncated could there be significant cost savings, but that would substantially reduce the accuracy for the hardest households to count, which are already the biggest problem. For both the census and the PES, the data quality is worst for the cases that are hardest to follow up, and a disproportionate part of the expense is in following up the hardest cases. Furthermore, sampling for non-response follow-up will make the DSE even more difficult, and even less accurate. Data quality problems in the PES follow-up are magnified enormously by the DSE. Thus the problem is worse for the DSE than for a headcount.

25) *GAO also concluded after the 1990 census that a high level of public cooperation is key to obtaining an accurate census at reasonable cost. Unfortunately the mail response rate has fallen with every census since 1970, and was only approximately 65 percent in 1990. The reasons for this decline are in many instances outside of the Census Bureau control, for example the increase in commercial mail and telephone solicitations and in nontraditional household arrangements. For these reasons, the Bureau is planning a public education campaign for the 2000 census, surpassing any previous attempts. Given the response in 1990, do you believe this is money well spent?*

*Do you believe that this public education campaign can succeed in arresting the decline in response rate?*

*Even if it does, wouldn't some use of sampling be warranted to solve the problems associated with reaching the last few [percent] of nonresponding households?*

I am not expert at motivating the public, but I think that such a campaign could be very helpful. The details of the campaign would be crucial to its success. See my answer to the previous question (24) in response to the last part of this one.