Responses to Questions from the Honorable Carolyn B. Maloney on Census Undercount Adjustment

Kenneth Darga, Senior Demographer
Office of the State Demographer
Michigan Information Center
Michigan Department of Management and Budget

June 19, 1998

STATE OF MICHIGAN

JOHN ENGLER, Governor
DEPARTMENT OF MANAGEMENT & BUDGET
P.O. BOX 30026, LANSING, MICHIGAN 48909
MARY A. LANNOYE, State Budget Director

Office of the State Demographer
Michigan Information Center

June 19, 1998

Honorable Carolyn B. Maloney
Ranking Minority Member
Subcommittee on the Census
U.S. House of Representatives

Dear Representative Maloney:

Thank you for the twenty-five questions which you included in your memo to me on May 13, 1998. These questions raise a wide range of important issues, and I hope that the following responses will lead to a dialog which advances the debate on the techniques proposed for use in the next census.

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 or redesign it?**

   Very early in my career, I had an experience with this dilemma which I believe can shed a great deal of light on the process of computing adjustments for census undercount.

   In the first government agency I worked for, I was once asked to do a quick analysis to show the cost of excess hospital capacity in Michigan. I had a pretty good idea what to expect based on the published literature on the subject, but my first calculations showed just the opposite of what I expected.

   Naturally, the question I asked myself was *"What did I do wrong?"* When I reviewed my computer program with this question in mind, I found a simple computational error that explained a large part of the problem. The figures still didn't point in the expected direction, but at least they didn't point so strongly in the "wrong" direction.

I couldn't find any more mistakes in my program, so the next question I asked myself was *"How can I improve the analysis?"* Since I had been taking a very simple approach to a very complex question, it didn't take long to find that I had left out some important factors which biased the results in the "wrong" direction. When I repeated the calculations with allowances for those factors, I got the results that I expected.

Unquestionably, the changes which I made were improvements. I had produced an analysis that was consistent with my expectations about what was true and with the published literature on the subject. But that experience left me with two important questions:

(1.) *What would have happened if my initial results had been consistent with my expectations?* Would I even have found my computational error if I hadn't had to ask myself "What did I do wrong?"

(2.) *What would have happened if my expectations and the initial results had been the opposite of what actually happened?* What if I had expected excess hospital capacity to *decrease* hospital expenditures instead of increasing them, and what if my first calculations had shown an increase? Would I have been able to find some legitimate factors that were left out of the initial analysis which biased the results in this new direction? Would I have "improved" the analysis in the opposite direction if I had the opposite expectations?

I had encountered a dilemma which faces all researchers, whether they are aware of it or not:

> On the one hand, it is probably impossible to produce good research on a complicated problem without finding and correcting mistakes and modifying methods based on new insights that are gained in the course of the analysis. And a principal way to find those mistakes and gain those new insights is by finding things that are contrary to expectations and figuring out either what went wrong or how the data and the analysis can be improved.
>
> On the other hand, when the corrections and refinements are driven by expectations of what the results should be, the research will tend to conform to those expectations regardless of whether those expectations are correct and regardless of whether the data and methodology are sound.

[See also the answer to Question 8 below.]

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?**

The central flaws of the proposed method of adjusting for undercount, which are explained in the papers that I submitted to the Subcommittee on 5/5/98,[1,2] are (a.) that it misses many of the same people who are missed by the census, and (b.) that many—in fact, most—of the people that it identifies as missed by the census were not missed at all. Thus, any differences it suggests between states are not so much differences in the amount of undercount as they are differences in the amount of error that the Census Bureau makes in trying to measure undercount.

Several of the sources of bias noted in my testimony are of particular relevance at the state level. For example:

- The exclusion of homeless people from the post-enumeration survey results in a bias against states whose homeless people are more likely to be staying with households during the April census than during the subsequent post-enumeration survey.

- Differences in weather and climate can affect the level of fabrication in the post-enumeration survey, which in turn can have a very serious impact on the apparent undercount rate.

- Because differences in weather and climate influence the likelihood that people will be at home when an enumerator visits, they can affect the proportion of successful PES interviews in different states. A high rate of unsuccessful interviews or proxy interviews in the PES can seriously increase the level of error in measuring undercount.

- When people migrate from one state to another on a seasonal basis, the post-enumeration survey can assign them to a different state from the one they reported as their "usual" state of residence when they filled out their census form.

[See also the answer to questions 4 and 20 below.]

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?**

In the "Notice of Final Decision" regarding adjustment of the 1990 Census, Secretary Mosbacher wrote:

> 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... While we know that some will fare better and some will fare worse under an adjustment, we don't really know how much better or how much worse. If the scientists cannot agree on these issues, how can we expect the losing cities and states as well as the American public to accept this change? [*Federal Register,* 7/22/91, page 33583]

This statement by Secretary Mosbacher is not at odds with my testimony. The figures cited, which involve comparing the adjusted counts to calculations based on assumptions about actual undercount in each state, are consistent with everything I have said about high levels of error in the Post-Enumeration Survey. An adjustment methodology that seemed to be less accurate than the census for 21 or 28 or 29 states in 1990 can hardly be considered a sound basis for fine-tuning the results of the next census.

[See also the answers to Question 3 above and Question 20 below.]

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?**

In addressing this question, it is important to remember that the 1980 and 1990 censuses were the most successful in history with respect to minimizing undercount. Based on the Census Bureau's "demographic analysis" method, the 1.8% estimated undercount in 1990 compares favorably to the estimated undercounts for 1940 (5.4%) through 1970 (2.7%). Likewise, the estimated undercount for blacks in 1990 (5.7%) compares favorably to the estimated undercounts for blacks for 1940 (8.4%) through 1970 (6.5%). The estimated 1990 undercounts for blacks, for other races, and for the population as a whole are the second best ever recorded; the only census with better results was the 1980 Census. (See Figure 1 of my first paper.[1])

My assessment of these figures is that the Census Bureau has made a lot of progress through the so-called "traditional methods." Since a number of promising improvements have been incorporated in the plans for Census 2000 and further improvements remain to be explored, it appears that the "traditional methods" hold promise for further progress. [See also the answer to Question 2 above.]

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?**

As you note, there are substantial discrepancies between the undercounts suggested by the post-enumeration survey and those suggested by demographic analysis. These discrepancies can be seen in Figure 2 of my first paper[1]: Relative to the results of demographic analysis, the undercount adjustments that were proposed for the 1990 census were 36% too low for black males but 43% too high for black females at the national level. The adjustments for other males were about right at the national level, but the adjustments for other females were 133% too high. Subsequent to correction of several errors, the adjustments proposed in September 1992 were 42% too low for black males and 33% too high for black females at the national level. The adjustments for other males were 25% too low, and the adjustments for other females were 50% too high. The situation was even worse at the regional level, where the proposed adjustments presented an inconsistent mosaic of high and low adjustments for different age, race, and sex categories.

The birth data and other data used in demographic analysis provide a very solid basis for estimating the relative number of males and females that were missed by the census. The discrepancies between the PES and demographic analysis therefore demonstrate quite clearly that the undercount adjustments derived from the PES are implausible and unreliable. However, one obviously cannot go beyond that to characterize them as consistently overstating or understating the undercount of minorities.

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 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 various techniques for conducting a more accurate enumeration—including those listed in my response to Question 2 above, those discussed in reports by the National Academy of Sciences, those proposed by the Census Bureau, and others as well—can be expected to promote a modest improvement in undercount rates. As explained in my response to Question 2 above, I believe that the remaining undercount is best addressed through population estimates rather than through census adjustments.

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?**

Although the Census Bureau tries very hard to count everybody and makes special efforts to count minorities and persons in poor communities, there are still some people who are missed. Regardless of whether they are missed because their living arrangements make them hard to count or because they intentionally avoid the census, it is desirable to know how many people each community really has and what their characteristics are.

However, the methodology that has been proposed for adjusting the census is not acceptable: it reflects survey matching error more than it reflects undercount, it would greatly reduce the value of sub-national census data, it

bias affect the accuracy count of those traditionally undercounted, Blacks, Hispanics, Native Americans, renters?

Your understanding of correlation bias is correct. Correlation bias should lead to a very substantial underestimate of the undercount for those groups which tend to be missed by both surveys.

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

That is only one of the risks. Another problem is that some communities might have more correlation bias than others. This is one of several factors that can cause the adjusted counts to be less indicative of a community's share of the nation's population than the original counts.

Another problem with correlation bias is that analysts who dismiss it as innocuous sometimes seem to forget that it is there. Correlation bias should result in adjustments for undercount that are much too low. However, the undercount adjustments derived from the 1990 Post-Enumeration Survey were *not* much too low: they were much to high for some segments of the population, much to low for others, and about on target for the national population as a whole. Analysts who forget about correlation bias and focus only on the seemingly plausible picture of undercount for the national population as a whole can make the mistake of thinking that the PES provides reasonably accurate information about undercount. However, for analysts who do *not* forget about correlation bias, the fact that the adjustments derived from the PES are not consistently too low is a clear sign that there is something seriously wrong with them.

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.**

While I do not disagree with this statement, I would add that statistical sampling methods can be used in ways that are effective and ways that are ineffective, in ways that are objective and ways that are biased, and in ways that assist and ways that detract from the census process. Like any tools,

statistical sampling methods work better for some purposes than for others, and they can be used in both appropriate and inappropriate ways.

12. **Dr. Lawrence Brown also testified before Senator Thompson that the Sampling for Non-Response 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.**

My testimony and analysis have focused exclusively on the issue of undercount adjustment, and I have not comprehensively reviewed the methodology proposed for handling non-response. Nevertheless, the following observations should be helpful for understanding some of its shortcomings.

An underlying premise of sampling for non-response is that each census statistic will be based mostly on actual responses, and that it will therefore not be seriously affected by minor errors in estimating the characteristics of the remaining 10% or so of the population from a sample.

One critical statistic for which this premise does not hold is the vacancy rate. Obviously, most vacant households will not respond to the census. It is my understanding that most of them are to be excluded from follow-up based on reports by letter carriers that they are vacant. (The plan calls for a sample of these housing units to be followed-up, however, in order to adjust for inaccuracies in the letter carriers' vacancy reports.) Any vacant units that the letter carriers do not report as vacant are to be followed up on a sample basis along with other non-responding households. Unfortunately, neither the letter carrier reports nor the proposed samples will produce reliable vacancy data. The letter carrier reports tend to be inaccurate, their errors cannot be corrected very well through the proposed sample, and the routine sampling of non-responding housing units will be subject to error as well.

In its preliminary testing, the Census Bureau found that 42% of the housing units that letter carriers identified as vacant were actually occupied, and that half of the units pre-identified as vacant were not identified as such by the letter carriers. If this result is at all indicative of the level of error to be expected in the letter carrier reports, they provide a very poor basis for determining vacancy status.

These deficiencies of the letter carrier reports cannot be corrected adequately even through the 30% sample recently proposed. Variations in the accuracy of letter carrier reports from neighborhood to neighborhood and from carrier to

carrier will present a serious dilemma: If the correction factors are derived from broad geographic areas, they will not be applicable to neighborhoods where vacancy status is particularly easy or particularly hard to determine, nor to neighborhoods where the letter carrier has particularly high or particularly low levels of skill and conscientiousness in determining vacancy status. But if they are derived from small geographic areas, they will tend to be dominated by sampling error. Whichever way the Census Bureau chooses to resolve this dilemma, the correction factors will be unreliable for small units of geography. The poor overall quality of the letter carrier reports, in turn, will cause those unreliable correction factors to have a very large impact on the vacancy rates.

A similar dilemma arises in connection with vacant units in the "regular" sample of non-responding households. The number of vacant units missed by the letter carriers can be expected to vary widely from neighborhood to neighborhood: Data derived from broad geographic areas will therefore not be indicative of local conditions, but data derived from small geographic areas will tend to be dominated by sampling error. Finding even one vacant housing unit in the sample can cause several housing units to be considered vacant, which can substantially change the vacancy picture for a census block or a small community. Any error—whether sampling error or non-sampling error—will therefore tend to have a serious impact. And since we are talking about measuring a (usually) small proportion of households through a small sample drawn from a small population, relatively high levels of error can be expected.

These problems would be much less serious if 90% of the data on vacancy were based on actual enumerations and only 10% of the data were subject to substantial error. However, that will not be the case due to the fact that most vacant housing units do not respond to the census. Unlike most other census statistics, the numerator of the vacancy rate is to be *almost entirely* based on very imprecise data.

A problem with faulty vacancy rates is far more critical than it may seem at first glance. In addition to being an important statistic in its own right, the vacancy rate plays a crucial role in determining the census count itself. If the estimated vacancy rate for a unit of government is 2 percentage points too low, then people will be imputed as living in vacant housing units and we can expect the population count to be a little more than 2 percentage points too high. If the estimated vacancy rate is 2 percentage points too high, then housing units that are occupied will be assumed to be vacant and we can expect the population count to be a little more than 2 percentage points too low.

Errors of this magnitude and greater would be quite likely for many units of government, particularly where there is a substantial amount of seasonal or vacation housing. For example, 49% of the units of government in Michigan had vacancy rates of 10% or more in 1990, 31% had vacancy rates of 25% or more, and 14% had vacancy rates of 50% or more. The proportion of housing units in these areas whose vacancy status would be determined by very imprecise methods would therefore be quite substantial, and the resulting census "counts" could easily be off by *several percentage points*.

As a demographer involved in the production of intercensal population estimates, I am very much aware of the weaknesses and limitations of those estimates and of the need for periodically benchmarking them to new census counts. I am therefore alarmed by the prospect that the proposed methodology might produce census "counts" for many units of government that are less reliable than their intercensal population estimates based on the 1990 Census, and that future population estimates for these areas might have no accurate basis at all.

Another potential problem with sampling for non-response is the possibility of distortions in local population data caused by replicating cases encountered in the sample. For example, if the methodology turns one household with a grandmother caring for grandchildren into several local households with grandmothers caring for grandchildren, or one household with twelve children into several local households with twelve children apiece, then the local census data will be seriously distorted. Thus, it would not be appropriate to replicate the findings from the sample within a small geographic area. (It may be appropriate, however, to use large-area samples as a basis for assigning weights to local census responses in order to influence the composition of the "deck" used for imputing the characteristics of non-responding households. See item 1(g.) under Question 2 above.) [Problematic aspects of sampling for non-response are discussed further in Question 24 (c.) below.]

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 now believe 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 strongly disagree with this statement. The two papers which I submitted as testimony to the Subcommittee on 5/5/98 are entirely directed toward explaining my position on this question.[12]

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.**

Subjective decisions can bias the results in ways that are not necessarily even intentional, conscious, or politically motivated. The most frequent and most likely way for this to happen is for personnel at various levels of the ICM effort—particularly interviewers, matchers, and the managers and statisticians responsible for implementing the methodology—to be influenced in their subjective decisions by their expectations about undercount. For example, when the match status of a particular record is not clear, it is possible for the classification to be influenced by whether the matcher expects people in that demographic category to have a high level of undercount. When a PES interviewer fabricates data on a hot or rainy day for people who never seem to be at home, the characteristics assigned to those people will naturally reflect the expectations of that interviewer. When a decision is made about whether to send a group of records back for re-matching or to downweight a group of records as outliers, that decision can influenced by whether the initial findings for those records were consistent with expectations about undercount and by whether the overall level of apparent undercount is higher or lower than expected.

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 statement? If you disagree, please explain why.**

I agree with that statement. A substantial portion of this problem is already handled through the Census Bureau's traditional "imputation" or "substitution" process for non-respondents and partial respondents. The importance of this element of the census process is frequently overlooked and, as explained in the answers to questions 2 and 12, this process can be improved. The remainder of

the problem, as explained in the answers to questions 2 and 8, can be better solved through an *estimate* of undercount rather than an *adjustment* for undercount.

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.**

One of the interesting things about measuring undercount through a post-enumeration survey is that the process has several fatal flaws, any one of which is sufficient by itself to explain why it produces such unacceptable results. One such flaw is sampling error due to a sample size that was insufficient to support the detailed stratification which the undercount adjustments require. This was such a big problem that there is no implausible aspect of the 1990 adjustments for which it is not a sufficient explanation.

It would be a fallacy, however, to conclude that sampling error is therefore the only explanation or even the chief explanation for the many implausible aspects of the 1990 adjustment factors. There are several other documented problems which are also sufficient by themselves to explain them. For example, the documented level of uncertainty and error in matching is sufficient to explain any of these implausible results. The level of fabrication in typical surveys, which was generally confirmed by the various studies of fabrication in the PES, is comparable in size to undercount and sufficient to explain any of these implausible results. Likewise, any of the implausible results can be explained by the fact that such an attempt to measure a small component of the population is extremely sensitive to tiny errors in the insurmountable task of classifying the remainder of the population. (See pages 6 through 9 of my first paper.[1]) It would be foolish to presume that solving only one of these problems would be sufficient to "fix" the proposed process for measuring undercount. There would be more than enough problems remaining to invalidate the results.

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

It is more than sufficient for the post-enumeration survey's traditional role of evaluating census questions and procedures. However, no increase in sample

size would be sufficient to produce valid adjustments for undercount through a post-enumeration survey, since sample size is not the only problem or even the chief problem. As explained in the answer to Question 16 above and in the papers which I submitted to the Subcommittee as testimony on 5/5/98,[1,2] the attempt to measure undercount through a post-enumeration survey has several fatal flaws that are not caused by insufficient sample size. These flaws account for much of the estimated undercount and, since they involve non-sampling error, they obviously will not be reduced by enlarging the sample. In fact, an increased sample size, coupled with a very tight time schedule and questionable staffing levels, is likely to increase the problems of fabrication, proxy interviews, and matching error which plagued the 1990 PES.

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?

I have not studied this issue in detail. However, as explained in the answer to questions 2 and 5 above, it is appropriate to say that the Census Bureau's "demographic analysis" method indicated that the 1980 Census was the most accurate in history and that the 1990 Census was only the second most accurate in history with respect to undercount.

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?

There are three sets of terms that need to be explained: (a.) *actual* gross overcount and undercount, (b.) gross *measured* overcount and undercount, and (c.) net measured overcount and undercount.

(a.) "Actual gross overcount" is the number of people *actually* counted twice by the census or counted in error. For example, people who were born after April 1 or who died before April 1 are sometimes counted by the census even though they should not be. College students who are counted at their parents' home instead of at the school where they lived are considered part of the "overcount" of their parents' community and part of the "undercount" of their college community. Overcount is usually

referred to as "erroneous enumeration." Similarly, "actual gross undercount" is the number of people *actually* missed by the census.

(b.) "Gross measured overcount" and "gross measured undercount" are appropriate terms for the number of people identified as erroneous enumerations by the Post-Enumeration Survey and the number of people identified as undercounted by the Post-Enumeration Survey. The "26 million mistakes" to which the question refers represent gross measured overcount and gross measured undercount. These numbers are much higher than *actual* gross overcount and *actual* gross undercount for several reasons:

- Much of the measured undercount and overcount is due to measurement errors in the post-enumeration survey rather than actual undercount and overcount in the census. This is the central point developed in my papers. [See pages 6 through 9 of my first paper[1] and pages 3 through 13 of my second paper.[2]]

- All of the people who are added to the census count through the substitution process and all of the people whose census responses are too incomplete to be used for matching are considered to be erroneous enumerations. The corresponding people who are found in those housing units by the Post-Enumeration Survey are considered to be part of the gross undercount. While this is appropriate in the context of the PES analysis, it does tend to make the gross measured overcount and gross measured undercount misleadingly high.

- People who seem to be counted in the wrong location by the census are counted as part of the undercount in one place and part of the overcount in another. This is appropriate in the context of the PES analysis, but it tends to make the total number of errors appear misleadingly high.

- Matching errors in the PES analysis typically involve a census record which should be matched with a PES record but which fails to match for any one of a number of reasons. In most such cases, the census record becomes part of the gross measured overcount and the PES record becomes part of the gross measured undercount. Again, this is appropriate in the context of the overall PES analysis, but it does tend to make the gross measured overcount and gross measured undercount misleadingly high.

[It should be noted that matching error does not always result in offsetting errors in gross overcount and gross undercount. For example, if the person described by the unmatched census record really does exist, it might be difficult to prove that they don't exist and they therefore might not become part of the measured overcount. This is one of the ways that matching error introduces bias into the undercount adjustments.]

- Looking at the PES in a broader sense, it can be expected that the number of people erroneously identified as overcounted or undercounted will naturally tend to exceed the number of people erroneously identified as counted correctly. This is because only a very small proportion of the population is actually overcounted or undercounted: in other words, there are very few people at risk of being *erroneously* identified as counted correctly. However, the vast majority of people are counted correctly by the census, and they are therefore at risk of being erroneously identified as overcounted or undercounted. This results in a large upward bias in the gross measured overcount and the gross measured undercount. [This issue is discussed in more detail on pages 6 through 9 of my first paper.[1] On page 9 of that paper, there is a list of eighteen problems which make it very difficult to match people correctly between two surveys so that they can be classified accurately as overcounted, undercounted, or correctly counted.]

(c.) "Net measured undercount" can be simply computed by subtracting gross measured overcount from gross measured undercount. (If an area has more measured overcount than measured undercount, its "net measured overcount" can be calculated by subtracting its net measured undercount from its net measured overcount.)

Thus, the frequently cited figure of "26 million mistakes" is greatly inflated, and it does not reflect the actual level of accuracy in the 1990 Census.

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?

Since the undercount adjustments reflect error in measuring undercount more than they reflect undercount itself, any prediction of how the numbers will fall out in any particular census is very uncertain. With that caveat, my expectations are as follows:

(a.) Estimating the adjustments for each state individually will negate most of the advantage otherwise gained from a larger sample size in terms of sampling error.

(b.) The factors which introduced geographic bias into the 1990 undercount adjustments will tend to affect individual states in the same way that they affected regions in 1990. [See answer to Question 3 above.]

(c.) Since state boundaries are as artificial as regional boundaries in terms of having a logical relationship with undercount rates, I see no reason at this time to expect an increase in accuracy resulting from this change in stratification.

[See also the answers to questions 3 and 4 above.]

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.

There is no contradiction between the findings which you cite and the concern about trying to process more interviews with inadequate staff in a shorter period of time. In fact, the findings reinforce the concern.

The higher error rates during the final weeks of follow-up do not result simply from "being in the field too long." The first weeks in the field result in more accurate data because they involve actual interviews with people who are willing to be counted. The final weeks in the field result in less accurate data because they involve more interviews with people who have resisted repeated attempts to count them, more proxy interviews to "close out" cases for which a direct interview cannot be obtained, and more fabrication of interviews in response to pressure to close out as many cases as possible before the deadline.

Shortening the amount of time in the field does not eliminate those final weeks of interviewing in which high error rates can be expected. The final weeks of interviewing will still be there, with all of their pressure to close out the difficult cases. Instead of eliminating the *final* weeks of interviewing, the current plan would, in effect, eliminate the *initial* weeks of interviewing in which lower error rates can be expected. By calling for more PES interviews in a shorter period of time with inadequate staff, the current plan creates a danger that the initial weeks of interviewing will be as error-prone as the final weeks of interviewing were in 1990.

It should be noted that the accuracy problems in the final weeks of interviewing and the concerns about truncated time frames apply both to the census itself and to the post-enumeration survey. Proxy interviews, fabrication by interviewers, and unreliable reports by respondents are problems for the PES as well as for the census—in fact, they are even more serious when they occur in the PES. The timetable for Census 2000 involves very tight time frames for both the census and the PES.

[See also the response to Questions 24(a.) and 24(c.) below.]

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?

I have not studied this question in detail. I understand that the Census Bureau has concluded from its research that shortening the form would not have a large impact on response rates. I do know, based on the involvement of my office in the Census Bureau's survey of data users and from its work in disseminating census data and in using census data to address needs of data users, that the information on both the long form and the short form is very widely used in both the public and private sectors. A radically shortened questionnaire would greatly diminish the value of the census. However, *if* we have a successful census in 2000, and *if* the Continuous Measurement program is adequately funded and successfully implemented, it should be possible to eliminate the long form in 2010.

23. In its 1992 capping report on the 1990 census, the GAO concluded that "the results and experiences 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.

It is not entirely fair to criticize a statement removed from its context within a larger report, so the following comments should not be interpreted as a criticism of the GAO or its 1992 report.

(a.) First, it is important to realize that our diverse and dynamic population is not a new development. Our history has included settlement of the frontier, Indian wars, emancipation of slaves, massive foreign immigration, industrialization, urbanization, the Great Depression, suburbanization, inter-state redistribution of population, and many other events and changes that have always made our population diverse, dynamic, and challenging to count. As difficult as it is to develop a precise Master Address File for Detroit in 1998, it would have been far more difficult in 1898.

(b.) I agree with the notion that there is considerable room for improvement in the census and that census methods should adapt to changes in the population. However, I am not sure exactly what is meant by "fundamental" changes. The concept of finding out how many people there are by counting them is sound, and I would characterize the required improvements as "incremental" rather than "fundamental."

(c.) The deficiencies of the census require not simply "change" but rather "change for the better." It should be clear from my testimony and the testimony of the other members of the 5/5/98 panel that the particular uses of sampling that have been proposed for Census 2000 would be very serious changes for the worse.

(d.) The 1990 Census approached our "diverse and dynamic" society, in which it is often difficult to find people at home, through a mail-back census form with instructions available in 34 different languages. It is somewhat ironic that the innovation proposed for dealing with these problems is a post-enumeration survey that relies exclusively on *personal interviews* by enumerators, most of whom speak fewer than 34 languages. The proposed innovation is more poorly adapted to our diverse and mobile society than the census itself.

24. **(a.) 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, 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. Don't you agree?**

It is inherent not just to the census, but to any survey which must obtain information about people who are difficult to reach or resistant to being counted. These problems apply even more to Sampling for Non-Response and to the post-enumeration survey required for Integrated Coverage Measurement than they do the census itself. These efforts not only involve exhaustive follow-up of difficult cases, but any errors will be multiplied when the sample results are inflated to represent the sampled universe. In fact, given the proposed constraints of time and resources discussed under Question 21 above, the proposed plans for Census 2000 can be expected to make these problems even worse. Again, it must be stressed that we need not just "change," but "change for the better." The proposed changes are even more susceptible to this problem than the old procedure was.

[See also the response to Question 21 above and Question 24 (c.) below.]

(b.) Do you have any information on the error rates for information gathered using close-out procedures?

The Census Bureau would be the most authoritative source for such information.

(c.) 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?

The premise underlying this question appears to be that sampling is somehow an *alternative* to traditional follow-up procedures. However, traditional follow-up procedures are just as much a part of the proposed uses of sampling as they are of the conventional census: follow-up is a critical part of Integrated Coverage Measurement, and follow-up is what Sampling for Non-Response is all about. Both of these efforts involve exhaustive efforts to obtain information about that last percentage of the population, and the associated errors will be

compounded when the sample findings are inflated to represent the sampled universe. The pertinent comparisons would therefore be between the overall error of the traditional census and the overall error of the modified census, or else between the error resulting from close-out procedures for the samples and the error resulting from close-out procedures for a traditional census. It should be obvious from the discussion above that these comparisons would not be favorable to the proposed sampling methodology.

That having been said, we are still left with a question about the overall error rate for sampling. With regard to sampling for undercount, a Census Bureau report estimated that identified errors accounted for about 33% of the net undercount suggested by the 1990 PES. A subsequent analysis by the same author raised this estimate to about 57%, and a further analysis by Dr. Leo Breiman raised the estimate to about 70%. (These reports are cited on pages 11-13 of my second paper.[2]) Similarly, the Census Bureau's Report of the Committee on Adjustment of Postcensal Estimates (the "CAPE Report," released on 8/7/92) stated that "about 45% of the revised estimated undercount is actually measured bias and not measured undercount. In 7 of the 10 evaluation strata, 50% or more of the estimated undercount is bias." These error rates compare unfavorably with error rates for virtually any aspect of the census process, regardless of whether or not such comparisons can be pertinently drawn.

(d.) 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?

Even if the sampling methodologies did not share the census's reliance on error-prone efforts to resolve difficult cases, the issues raised in the response to Question 12 above would still be pertinent. While there may be a place for sampling in improving the census, the particular procedure proposed for sampling nonrespondents appears to have some serious shortcomings.

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% in 1990. The reasons for this decline are in many instances outside of the Census Bureau's 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 rates?

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

Taking the last question first, some of the appropriate and inappropriate uses of sampling with respect to non-response are addressed in the answer to Question 12 above.

I agree that a high level of public cooperation and a high response rate are keys both to obtaining an accurate census and to holding down costs. While I have not reviewed the Census Bureau's publicity plans, I understand that they involve improvements to both the quality and the timing of the publicity efforts. (See also the answers to Question 2 and Question 5 above regarding the success of "traditional methods" in improving census participation.)

It should be noted that the issue of undercount adjustment also has very significant implications for levels of public cooperation and response:

- On the one hand, there is reason to believe that a decision to adjust the census would have a very serious negative effect on census participation. If people expect the census count to be adjusted, they may not think that the effort required to complete their census form is necessary. Similarly, the critical involvement of public officials and temporary census employees in securing high participation rates might be jeopardized by a decision to adjust the census. In the "Notice of Final Decision" on adjustment of the 1990 Census, then-Secretary of Commerce Robert Mosbacher wrote:

> I am worried that an adjustment would remove the incentive of states and localities to join in the effort to get a full and complete count. The Census Bureau relies heavily on the active support of state and local leaders to encourage census participation in their communities... If civic leaders and local officials believe that an adjustment will rectify the failures in the census, they will be hard pressed to justify putting census outreach programs above the many other needs clamoring for their limited resources.

Without the partnership of states and cities in creating public awareness and a sense of involvement in the census, the result is likely to be a further decline in participation. [*Federal Register*,, 7/22/91, page 33584.]

There is a real risk that, with an expectation of a correction through adjustment, the field staff would not have the same sense of commitment and public mission in future censuses and, as a result, careless and incomplete work would increase, thereby decreasing the quality of census data. These are the workers the Bureau depends on to collect the data from the groups that are hardest to enumerate. If these data suffer, the information lost at the margin is information that is especially important to policy development. [*Federal Register*, 7/22/91, page 33605.]

- On the other hand, the current controversy over adjustment may play a positive role in encouraging census participation. This controversy has increased awareness of the importance of being included in the census on the part of civic leaders, local government officials, civil rights organizations, and the general public. It might be possible to translate this awareness into something that everybody will find superior to an adjustment for undercount: a census in which people get counted the first time.

Thank you again for the opportunity to address these questions. I hope that these answers promote a greater understanding of the issues surrounding census undercount adjustment and that the resulting dialog will lead to a better census.

Sincerely,

Kenneth J. Darga, Senior Demographer
Michigan Department of Management and Budget

---

[1]Kenneth J. Darga, "Straining Out Gnats and Swallowing Camels: The Perils of Adjusting for Census Undercount," Office of the State Demographer, Michigan Information Center, Michigan Department of Management and Budget. Paper submitted as testimony to the House Subcommittee on the Census, May 5, 1998.

[2]Kenneth J. Darga, "Quantifying Measurement Error and Bias in the 1990 Undercount Estimates," Office of the State Demographer, Michigan Information Center, Michigan Department of Management and Budget. Paper submitted as testimony to the House Subcommittee on the Census, May 5, 1998.