Responses to Representative Maloney --

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?

A. I have been a hard-nosed advocate of "getting it right the first time" for many years, and won over many of my colleagues at OMB. This approach emphasizes planning that not only covers what you expect, but also is robust with respect to things that may go wrong. It works well even when it doesn't work perfectly.

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?

A. The undercount is by its nature a nonresponse problem (or in many cases a <u>refusal</u> problem). A refusal rate as low as 1% would be considered an outstanding achievement for most surveys. There are two factors that draw attention to this particular problem in the Census:

>   a) The problem of missing data in the census data base is much larger than this, but these other gaps are susceptible to "imputation," which, by its nature, leaves little or no evidence of the alteration (imputed items, for better or worse, are designed to look like the actual data on which they are based). The evidence of the undercount gap is external. The comparative estimates may not be completely accurate, but they are highly visible.

>   b) The requirement for a complete enumeration goes substantially beyond the arena where statistical methods are most effective (and beyond the arena where the usual statistical standards for an acceptable level of error can be applied).

On this second point, comparison with the "voluntary" income tax system is instructive. IRS has used statistical principles in its approach to tax returns (e.g., it offers taxpayers the option of selecting an imputed average minimum deduction, the "standard deduction," in lieu of documenting actual deductions), but an explicit statistical adjustment to the tax liability of willing taxpayers to offset the loss due to those who refuse to file would be considered arbitrary and capricious. Rather we live with the consequences (reduced revenue) and IRS spends an extraordinary amount of research and auditing effort to discover those who refuse to pay or underpay their taxes and correcting these problems on a case by case basis. Such a statistical adjustment strategy could eliminate the estimated revenue shortfall and make the revenue per person with tax liabilities more "accurate" on the average, but it would be less accurate for almost every individual who actually pays his or her taxes. Most statistical methods do not deal well with the issues of <u>individual fairness</u> that are critical to administrative systems such as the tax system or

to an "actual enumeration" intended to provide "fair" representation for every individual.

Some lessons of the tax system are also useful for a census. Look for classes of individuals who represent a disproportionate share of the refusals (the differential undercount), and use this information to refine and focus your strategy for bringing those individuals into the system. Demographic Analysis has provided some of the most important insights into the gross characteristics of census refusals, but Dual System estimates of those characteristics (because of the large confounded bias component and the substantial inconsistency with DA results) may be counterproductive for such a strategy (e.g., if attributes that arise from this very large bias are misinterpreted as actual attributes of the refusal population, efforts and funds may be substantially mistargetted).

"Traditional" efforts have missed some important opportunities to improve response in the past and I have commented on these in my responses to questions #8, #23, and #25 below.

3. You have mentioned your concerns about block level accuracy. Can you discuss you (sic) 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?

A. I did not comment on block-level accuracy per se. Three important attributes of an enumeration that must not be discarded lightly are 1) simple robustness, 2) uniform accuracy in both large and small areas and 3) additivity. As a practical matter, block level accuracy is one way these properties can be substantially preserved. I have commented on the shortcomings of compromise accuracy targets in my response to Miller question #1. As to the performance of DSE, I only note that it failed even the test of state-level accuracy in the one large scale evaluation we have (the 1990 PES) -- this is elaborated in my response to question #4 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?

A The statement by Secretary Mosbacher in 1991 was based on the original Census Bureau adjustment estimate, which was later found by the Census Bureau to be substantially inaccurate. Indeed the Secretary noted that the tally of 29 more accurate versus 21 less accurate, based on the original 2.1% adjustment, had already been reversed (to about 21-23 more accurate and 27-29 less accurate) based on the findings of independent analysts. A year later the census Bureau acknowledged that the original adjustments were substantially in error, revising the overall PES undercount estimate downward to 1.58%. By August of 1992, the CAPE report had been completed and it showed that even the revised estimate overstated the PES measured undercount by an amount much larger than the July 1992 "correction"; in other words the revised undercount

2

data set reflected in about equal parts the characteristics of the undercounted and the characteristics of measured bias.

The 1991 statement by the Secretary reflected what the Census Bureau had told him at the time. Due in large part to the efforts of the Secretary in commissioning the comprehensive CAPE evaluation, the 1991 data set on which these 1991 claims were based turned out to be more error than fact.

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?

A. As I indicated in my response to question #2, the undercount is a nonresponse problem more than a design problem. While I was at OMB, I was consistently critical of Census Bureau arguments that revolved around the premise that "we have always done it that way." OMB regulations prevent me from discussing the detailed information that came to me in the course of my employment, but suffice it to say that we made a number of recommendations for improving response to the Census prior to 1990. I am gratified that the Census Bureau under Martha Riche adopted some of these recommendations for the 2000 census. Other recommendations for improving the performance of the count have never been acted on by the Census Bureau (I have commented on some of these in my response to question #8). Because of the high visibility of this undertaking the Census Bureau has been risk-averse for a long time. It has been slow to embrace response theories (due as much to cognitive psychologists as to statisticians) that have proven very effective over the past two decades.

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 that would have been added back into the census by the Post Enumeration Survey?

A. The substantial inconsistencies between the picture of the undercount population implied by demographic analysis (DA) and that implied by the 1990 Post Enumeration Survey (DSE methodology) were of great concern to the expert panel supporting the CAPE evaluation. In comparing the two, the panel drew attention (in Attachment 8) to several sources of error in the DA estimates -- some that would generally exaggerate the number of persons not counted and others that would specifically exaggerate the number of blacks and Hispanics not counted. While DA may be subject to these biases, they pale in comparison to documented biases in the PES. The DSE methodology used in the PES was able to measure undercounted persons accounting

for, at best, about 0.9 percent of the population.

If the undercounted population is any larger than this (as is implied by DA), then the racial and ethnic characteristics of the missing group are <u>unknown</u> to PES (in DSE there is <u>no data</u> on the missing group). This is why the expert panel pressed the Census Bureau so persistently to remove this large remaining bias -- unless and until that large bias is removed, it is impossible to isolate the true racial and ethnic characteristics of the (measured) DSE undercount group from the spurious racial and ethnic characteristics attributable to the millions of spurious undercount cases inferred by DSE but actually contributed by the DSE bias processes. [Note: the previous sentence refers only to the net effect of bias -- the errors due to DSE that produced this measured bias in the net figure actually contributed much larger numbers of spurious undercounts and spurious overcounts inferred by the DSE methodology -- some of these spurious inferences may be related to the numbers that appear in question #19.)

In short, DA may somewhat exaggerate the number of persons undercounted, but the DSE methodology measured a far smaller undercount group than that implied by DA, and could not estimate the racial and ethnic composition of this measured undercount group with any accuracy in the presence of DSE's very large measured bias.

If you could remove the bias from the dual system PES estimates, and you were willing to make the leap of faith (as we are asked to do in the 2000 Census plan) that DSE works, (i.e., its measured undercount group accurately represents the actual undercount), you are left with the unavoidable conclusion that demographic analysis substantially overstated the 1990 undercount (probably by exaggerating black and Hispanic components of the differential undercount) and that the 1990 Census was the most accurate census in history.

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?

A. There are several types of bias in the count. They generally reflect the kind of <u>nonsampling</u> error statisticians classify as "nonresponse," and reflect a level of nonresponse that would be considered trivial in almost any sample survey (remember that the accuracy standards expected of an enumeration vastly exceed those that sample surveys are typically capable of meeting). Based on external benchmarks (demographic analysis), there is an overall downward bias in the count of about one percent. If this bias were uniform, it would make virtually no difference to the objective of supporting an accurate apportionment. The same external benchmark suggests that the bias may not be uniform across all potential Congressional districts. The evidence here is in proxy demographic variables, e.g., race and ethnicity. Once again, if the racial and ethnic characteristics of the populations in each potential Congressional district were uniform, there would be virtually no effect on the accuracy of apportionment. So one must explore the

mechanisms that produce these differences.

If the Census Bureau address listing methodology disproportionately misses black or Hispanic households, then this error affects both the count and the DSE. Likewise, if a disproportionate number of black and Hispanic households deliberately avoid participation the census, they will be missed by both the count and the DSE. DSE is blind to these particular types of errors and impotent to "correct" them. The presence or absence of ICM has no effect on errors of this kind. The bias they produce will persist until the nonresponse problems are addressed directly by such things as better listing, more effective follow-up (refusal conversion), and eliminating root causes of mistrust.

The value of coverage measurement is a different story. The DSE methodology requires data and is thus not very useful for dealing with true non-response, but this is not to say that a well-designed coverage measurement program cannot contribute significantly to improving the accuracy of the count. During the period when my OMB colleague, the late Maria Gonzalez, was reviewing plans for the 1990 Census, we discussed the role of coverage measurement at length. I had the temerity to suggest that, with all the expectations placed on the PES, the sample should be larger. Maria gently took me to task, pointing out that a sample small enough to be performed quickly by an expert staff could detect performance problems and errors early enough to correct problems in the count. (These advantages of a small manageable sample are also reflected in the September 1996 Report of the American Statistical Association Blue Ribbon Panel on Uses of Sampling in the Census.) This potential for feedback is the critical difference between effective quality control and simple quality measurement.

8. Do you believe that it is acceptable for the census to consistently miss certain segments of the population -- Africans (sic) 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 you (sic) proposal(s) will reduce the differential undercount?

A. I believe most of these problems must be recognized and addressed for what they are -- deficiencies in performance and highly motivated refusals. The Census Bureau is aware of large differences in response performance even among its regular staff, but has been reluctant to admit or address this problem administratively for a variety of reasons (possible litigation risk?). After the 1980 Census, focus groups were conducted in high undercount areas -- the results were reported by GAO. This research identified some strong, perfectly rational motivations for resisting the Census, most of them related to the numerous questions on the long form. This is a dilemma for the Census Bureau. Local agencies lobby heavily to retain the questions so that they can use block-level data to target such programs as housing code enforcement. But people who have observed the housing inspectors moving into their neighborhoods after the block-level data are released can make the connection for themselves -- and tell their neighbors about it next time. Even those who simply wonder why those questions are there can put two and two together.

The attraction of the rich, geographically detailed Census research data base for enforcement authorities creates some perverse incentives. For anyone who may have reason to avoid the notice of enforcement authorities, the only safe course may be to avoid the census altogether or to file a false report that appears consistent with local regulations (e.g., omit listing some occupants to avoid evidence of overcrowding).

Some steps have already been taken --

> a) Making more forms available for willing respondents may offset differential problems in the mailing list (though I still have some concerns here about double counting or other fictitious reports).

> b) Reducing the number of questions on the long form may help if the changes reduce suspicions among groups with high refusal rates.

One other way to address this problem is to decouple the research component (the long form) of the Census from the count (the short form). The temptation to piggy-back this huge sample (about 20 million) on the basic count has outweighed the possibility of reducing the undercount. The Continuous Measurement program had the potential for accomplishing this, but this has been deferred, trading off potential reductions in undercount for 2000 for one more bite at a huge long form sample.

Yet another possibility that has been proposed is to reduce the size of the long-form sample. Ironically, the long form sample size claimed to be the "minimum" needed to gather research information adequate to make program decisions is about 25 times larger than the sample size proposed to "correct" the constitutionally required count. I suspect these sampling judgments have more to say about the priorities of the research community and various bureaucracies than any reasonable statistical calculation. If the sample size of the research component is reduced across the board, then some resources are freed up for more extensive follow-up of the count. If the research community will not stand for this, even a reduction of the long form sampling rate in high undercount areas would be useful. There are also sound technical reasons for reducing the sampling <u>rate</u> in areas of high population density (what determines sampling error is sample <u>size</u>, but sampling rates are easier to sell to nonstatisticians -- the resulting distortion has affected the sample designs for both the long form and coverage measurement.).

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 males. How does correlation bias affect the accuracy count of those traditionally undercounted, Blacks, Hispanics, Asians, Native Americans, renters?

A. There are two reasons for using the term "correlation bias." The first has to do with the fact that the two samples being compared are not independent. Census argues that this effect is small. Another reason is that cases missing from either the count and/or the PES are correlated. However for the critical "4th cell" (the unobserved cases missing from both), your interpretation is correct -- some unknown number of missed cases (for which there are no data whatever) are assumed to make up this cell. But since there are no data, there are no attributes (black, Hispanic, young, male, etc.) to measure. Correlation bias is a property of the DSE methodology, not the underlying count. It represents that hypothetical portion of the undercount for which neither the count nor the follow-up survey have produced any useful information. Even the size of this bias can only be inferred indirectly by reference to other information. In other words it represents a (hopefully small) chunk of ignorance which adds nothing to our knowledge of the undercount.

Its effect on the accuracy of estimated characteristics of the undercounted population must also be inferred indirectly. The CAPE report breaks down the DSE (revised) estimate of 1.58 percent undercount as 0.85% actual measured undercount and 0.73% measured bias. Note that the correlation bias has not yet appeared. The 0.71% bias reflects the measurable errors made in the DSE analysis, e.g., correct counts incorrectly classified as overcounts or undercounts.   If the characteristics of these bias cases could be removed (as the expert panel urged) then what would be left would be a set of accurate characteristics (race, ethnicity, etc.) of the actual undercount group found by DES. But the Census Bureau determined that the processes available to remove the effect of the bias would add additional error. So the only characteristics of the 1.58% DSE estimate of undercount that can be tabulated consist of the unknown real characteristics of the true undercount group mixed with the unknown spurious characteristics contributed by the bias group. Since the detailed characteristics of the undercount group in DSE are knowable but are, in fact, unknown, any adjustment for the offsetting "correlation bias" (whose characteristics are unknowable within DSE by definition) can only be made at the most aggregate level. Based on other information, the Census Bureau estimated the size of the correlation bias (relative to the 1.58% estimate) at 0.38%, leaving an overall estimate of about 1.2% undercount consisting of 0.85% actual measured undercount within DSE (net of measured bias) plus 0.38% undercount missed by DSE. About the only thing we do know after all this is that the results of DSE are clearly inconsistent with the 1.8% undercount estimate derived from demographic analysis.

On another level, however, we can indirectly infer something about the accuracy of undercount characteristics that are theoretically observable in DSE. If we could remove the measured bias, and if the correlation bias were large enough to account for the difference between the 0.85% measured undercount observed in DSE and the 1.8% total undercount implied by DA, we would be left in the position of trying to estimate the characteristics of the whole undercount population (1.8%) from DSE information that represents only a minor fragment (0.85%) of that population. This is why I asserted that the DSE methodology becomes unreliable if the correlation bias is large.

10. Wouldn't the only risk of correlation bias be minimization of the undercount rather

than an overestimation (sic) the undercount?

A. No. The risk if the correlation bias is large is that you have <u>no information</u> about most of the undercount population. Statistical estimates based on no information are notoriously unreliable.

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.

A. Yes, I agree. Sampling methods have been used in this way for over 50 years.

12. Dr. 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.

A. The statement is too speculative for my tastes and, as Dr. Brown noted, is based in part "on idealized statistical assumptions." In fairness to Dr. Brown, he also noted risk of bias, several real-world problems, and some potentially troublesome interactions, concluding on balance that "if Congress can find the money, I'd prefer to see a full follow-up rather than the current sample response follow-up plan." I fully agree with that conclusion.

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.

A. I believe the statement with ellipsis and insertion(s) does not accurately convey the view expressed by Dr. Brown. There were two slightly different statements, one in his prepared testimony and the other in the transcript. The clearer of the two was in the transcript:

> "As of a month ago, the plans for the first stage were drastically simplified and improved, I believe. And these first stage changes lead me to believe that that stage can work to provide suitably accurate numbers. And some other changes they have announced to the ICM protocol make it possible to believe that that it might work as well as desired." [note -- the written statement also included the words "to correct undercount" at the end of this sentence]

The first part of the quotation is an endorsement of the improvements made to the first stage (sampling for follow-up) -- I agree that improvements have been made. The second part of the quotation is an extremely guarded expression of optimism that ICM may work after all. I am much less sanguine than Dr. Brown. A lot depends on what is meant by suitable accuracy or working "as well as desired." My views are based, not on the highly charged debate in the literature, but on the comprehensive evaluation (CAPE) performed by the Census Bureau itself.

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.

A. Variations on this approach have always held some attraction for OMB, particularly when there was potential for subtle abuses. The results have been mixed. Public or even expert scrutiny may reach consensus long before it has reached the whole truth. Constraints on processes tend to reward conservative methods and inhibit innovative breakthroughs. But the most potent risks arise in the case of procedures that are not inherently robust. All the careful caveats devised by statisticians cannot prevent a motivated advocate from changing the results by altering sensitive assumptions. Some statisticians believe that they have discharged their professional obligations by adding caveats to a frail result. I do not share that view.

I applaud the four minimum principles for effectiveness and objectivity advanced by Dr. Brown in paragraph 2 of his prepared testimony, and I would add robustness to the list for the reasons indicated above. I also share the concern expressed in his notes about ICM procedures:

> "such a procedure <u>violates</u> my principle 2(iii). But this <u>contradiction</u> at present seems <u>unavoidable</u> if one hopes to use reasonable ICM procedures to reduce the differential undercount problem below where it stood in 1990 and 1980." [emphasis added]

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.

A. I agree that there will always be some residual refusals and performance errors, but my response to question #8 above indicated the potential for reducing these below the currently accepted levels. I would certainly not write off the current level of refusals as unavoidable if that

is what is being inferred from Dr. Brown's statement.

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.

A. From the content and context of his statement, I believe Dr. Brown was referring to the shortcomings of the PES as an analytical tool. There are also indications in the CAPE report that additional data might have resolved some of the intractable problems of the evaluation, for example the inability to remove the measured bias. On the other hand, increasing sample size does not generally reduce the size of such biases, and it was the size and relationship of the biases that was the downfall of the PES as a tool for accurately allocating the undercount.

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?

A. It is much too large and unmanageable to provide quality control or even to secure the advantages touted by the ASA Blue Ribbon Panel (e.g., tighter control using expert staff to reduce nonsampling error).   The Census Bureau has experienced differential performance problems using regular staff in samples as small as 20,000 households. And if the DSE methodology fails because the type of nonsampling error known as correlation bias is too large (a distinct possibility based on the PES experience), then any sample size is too large.

18. The results of the PES in 1990 showed that the 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 that the 1980 census. Do you agree with the conclusion that 1990 was less (sic) also less accurate than 1980?

A. As indicated in my response to questions # 6 and #9, the PES results were in substantial conflict with 1990 demographic analysis results. If you believe the DSE methodology worked, then 1990 looks more accurate than 1980. I tend to agree with Dr. Brown -- it was pretty much a wash. I can still remember the wide-spread consternation with the unprecedented "surprises" and errors that occurred in the 1980 Census (the largest "closure error" in history by a wide margin, compromised quality control, and many horror stories from the field). This large deviation from expectations (closure error) probably made it much more difficult to detect and correct other, smaller discrepancies. So I would not be surprised if the performance indicators were a bit soft in 1980. Demographic analysis has provided the most consistent benchmark, but its methods and assumptions have changed over time and I doubt that comparisons are reliable to tenths of a

percent. In 1990, there were some suspicions that the DA figures might be too high. You also have to make allowances for the fact that the most recent census is almost always bad-mouthed in the course of justifying more funds for the next one.

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?

A. I have some difficulty understanding this question, but I will try to respond. First, 26 million would be about 10% of the count, but I am not sure what that figure refers to. Undercount, overcount, and the "net" are less haphazard and less precise than might be inferred from the details of this question. No count is perfect, so it is assumed that there are some undetected double counts and undetected undercounts that are reflected in the total enumeration. Since they are undetected, we don't know how many there are from the count itself. By means of external comparisons.(principally demographic analysis), we can estimate (with some error) how far off the count may be. Since the external comparison doesn't tell us anything about the mix of over- and undercount, we can logically infer things only about the net effect. In the past, coverage evaluation has provided some incomplete estimates of the mix of over- and undercount, but this should not be confused with the 1990 output of DSE (about half of those net errors were DSE errors, which is why they were classified as "measured bias").   (see also the response to question #6 above)

I have always been chagrined that millions of people may not take the census seriously or refuse to participate. OMB gets some misdirected census returns that are really bizarre. OMB also gets both complaints and misdirected hate mail that display distrust of the Census or the Census Bureau or both.

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?

A. Finer geographic stratification is a two-edged sword.   In theory, criteria other than political boundaries should be the deciding factors -- strata that are geographically diffuse can be perfectly valid and may perform better. Consider a case where census staff in some states are much more proficient at converting refusals than staffs in other states. With geographically diffuse strata, the effect tends to average out, but with state-based strata, the effect produces another kind of differential undercount that directly distorts apportionment information. It is entirely possible that state-based strata are a political palliative that imposes real penalties on accuracy.

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 be (sic) in the field for enough days in the 2000 plan. Can you explain the difference in opinion?

A. The phenomenon described by Representative Sawyer is not a matter of opinion. This pattern is well known -- it is not unique to the Census enumeration, it occurs in sample surveys as well (and for the same reasons). As the achieved response rate rises, productivity tends to fall and error rates tend to rise. Early respondents are self motivated to cooperate. Reluctant or forgetful respondents tend to pay less attention to the task and thus make more errors. Resistant respondents reached late in the process are often distracted by irritation and have less motivation to consider questions carefully. The stress of dealing with respondent irritation or the pressure of a final close-out process may also cause data collectors to make additional errors of their own.

But while the error rate (per observation) goes up, total error is generally reduced by filling in gaps that contribute to non-response error with 70% or 80% or 90% accurate information. Depending on error characteristics, there may be a break-even point at some very high response rate, but this point is rarely reached in most sample survey designs because follow-up is terminated at a lower level due to cost considerations. Surveys to capture very rare attributes or those which are bias-sensitive may spend the extra money to reach the break-even point. The census enumeration probably qualifies on both counts. There is a world of difference between flirting with the break-even point (as may be happening in some areas in the full enumeration) and calling it quits after a couple of tries.

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?

A. The GAO suggestion is sound. OMB has made similar recommendations. Though many other factors influence respondent cooperation, questionnaire length is one of a handful of factors that has consistently shown a correlation with response rates. Other attributes of a "streamlined" approach (visual simplicity, user-friendliness) also have a salutary effect on response. If this is incorporated into one of the "decoupling" strategies described in my response to question #8, there are opportunities for a much more sophisticated research program (e.g, more frequent measurement, much more powerful and efficient sample designs, etc.) which can make the data-rich research component more useful as well.

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 tradition (sic) '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.

A. I would go farther. I would not simply supplement the traditional headcount approach, I would replace most of it with a modern headcount approach. Until recently, only a few innovators in the Census Bureau paid much attention to the extraordinary improvements that have been made in mail survey methods or the lessons of cognitive psychology. But there are some cracks in the traditional conservative edifice. In 1990, the Census Bureau was persuaded to use a stratified design for the long form sample (only 5 decades after Neyman demonstrated the power of this technique). The Bureau has been listening seriously to some of the architects of the methodological advances of the 1970's and 1980's. Modern, simpler form designs that tested very well but were rejected in the 1980's may have made a comeback for 2000.

But the basic Census 2000 plan still represents 1980's thinking. Fundamental changes such as decoupling strategies that would liberate both the enumeration and the research component of the traditional census approach have been deferred because of risk averse client groups. I agree with GAO -- the time for these fundamental changes is now.

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, 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?

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?

A. I believe that GAO is discussing the same phenomenon discussed in my response to question #21, i.e., the increase in the per-observation error rate when pursuing high response. There

would only be 3% total error at stake in the decision whether to pursue the last 3% of the count, so the 30% figure cited by GAO must be the error rate per observation.

If cost considerations are set aside, then the (quality) break-even point is higher than the trade-off point implied by the GAO statement. A per-observation error rate of 30% is clearly preferable, from a quality standpoint, to the 100% per-observation error rate of completely missing an observation (some accurate data is usually better than no data). But error phenomena are not so well behaved as this, so there usually is a (quality) break-even point short of 100% response.

How is this affected by sampling? If you select a large sample and then pursue a 100% response rate among those selected for the sample, you will see the same rise in error rates as before. And then the small component due to sampling error must be added to this error. Sampling does not reduce the total error produced by pursuing high response rates, it increases it.

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 Bureaus (sic) 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?

A. Some of the trend in mail response rates is due to the fact that the mail portion of each successive census covered a larger fraction of the population. In the earlier censuses, the target population for mail was more selective and easier to reach successfully. By 1990 the mail portion was virtually the whole census and none of the problems could be avoided.   But now that 100% has been reached, this element of the decline should plateau. Response to telephone surveys has been hit hard by telemarketing and call-screening technologies. Response to traditional mail survey methods (like the census) has also declined, but more modern mail survey methodologies have bucked the trend (see my response to question #23).

One of the most effective elements in the modern revival of mail methodologies is the multiple contact strategy. These are almost always personalized contacts, but a "public education" campaign that drew attention and raised interest in the census might produce some of the same effect. If the campaign is as bureaucratic and condescending as its title, it probably wont produce

that effect.

Modernization of mail methods is the best bet for reversing the "decline" experienced by traditional mail methods. Don Dillman ("Total Design Method") argues that there is a synergy among the various elements of his method that cannot be achieved piecemeal. Before he started updating his book, he was making fairly regular visits to Suitland. I hope the right people were listening.