Responses to Chairman Miller --

1. In your testimony you indicated some reservations about the "ground rules" of the CAPE evaluation. What were these reservations?

A. I have two reservations, both concerning the standards of accuracy being applied.

First, there is the emphasis on "state-level" accuracy. It is true that accuracy of the state totals alone would permit the Census Bureau to provide information that discharges its obligation for supporting apportionment (assigning a number of seats to each state). But this disregards (irresponsibly in my opinion) the need for accurate substate data with which to actually construct congressional districts. Important rights are at stake in the accuracy of this process and there is no other data base that will serve.

Second, even the standard of state-level accuracy is treated as an "average" standard. The CAPE expert panel found that adjusted totals looked all right on the average, with some exceptions. In other words the biases (most of the problem) were judged to be acceptable in most cases. This disregards the concept of individual fairness (discussed at length in my response to Maloney #2). Given what is at stake, one must question whether individual states or voters would (or should have to) live with this kind of judgment.

2. You stated that the 1.58% revised estimate of the 1990 undercount released by the Census Bureau still contained bias and that using the Census Bureau estimate of measured bias and their assumptions concerning the offsetting "correlation bias" from the CAPE report, their estimate not (sic) of bias was 1.2% undercount. How does this compare to the similar figure attributed to the 1980 Census?

A. I believe the phrase I used was "net of bias." The published estimate of 1.58% is the sum of measured undercount (the target group) and measured bias (DSE error that exaggerates the undercount). To get at the "best estimate" of overall undercount that can be inferred from the CAPE analysis, the "measured bias" must be subtracted out and the unmeasured undercount (the missing piece represented by the estimated correlation bias) must be added back in. The result is 1.58 - 0.73 + 0.38 = 1.23 or about 1.2%. Each of these terms has been revised (and they are still being revised) so this estimate is certainly not precise to hundredths of a percent, in fact, even the tenths position is soft. Thus the actual similarity between this 1.2% (plus or minus something) and the 1.2% (also plus or minus something) estimated for the 1980 Census is surely coincidence. I attended one of the near riots that followed release of the 1980 figures (with the huge "closure error") -- acceptance of the 1990 figures was relatively sedate by comparison. Attempts to draw fine comparisons based on numbers ten years apart from evolving methodologies in the face of unmeasured uncertainties is, at best, naive.

In terms of quality control, closure error (see Maloney question # 21), and the DSE measured undercount (0.85%), 1990 appears better than 1980. But if one selects other measures, 1980

looks better. When one considers the size of the uncertainties and the fact that the variance (error) of the difference between two (independent) estimates is the SUM of variances (errors) in the two components, one would be hard pressed to find any significant difference in overall quality between 1980 and 1990.

3. From your experience as a professional statistician, to what extent are errors experienced in an actual enumeration likely to appear in a sample as well?

A. The errors that occur in an enumeration are what statisticians call "nonsampling errors." The introduction of sampling per se has no effect on these errors from sources other than sampling, it adds a new source of error (sampling error) to those sources that already exist. If no other changes are made, then the total error (consisting of both sampling and nonsampling error) increases. There is, however, an exception to this rule -- when the sample is small enough to permit use of improved enumeration methods that directly reduce these other errors.

The ASA Blue Ribbon Panel on the Census discussed this in its technical attachment to the 1996 Report --

> "The achievement of greater accuracy depends on how much more accurate the refined enumeration procedure is than the standard procedure and how much greater is its cost These factors need to be evaluated for each specific case to determine the comparative accuracy of census and sample results. It should also be noted that more refined enumeration methods can sometimes only be employed with smaller-scale, (i.e., sample), studies. Reasons for this include the need for highly trained enumerators who are available only in limited numbers and the use of burdensome questionnaires that can be employed only with a small sample."

The samples proposed for following up nonresponse are very large (not much smaller than the total nonresponding population that would be treated in full follow-up). This requires the same enumeration procedures (with the same contribution of nonsampling error) in the sample as would apply to the full follow-up alternative. If refinements that would reduce nonsampling error (the kind described by the Blue Ribbon Panel) were feasible on this large scale, then the cost of extending such refinements to full follow-up would be modest. (see also Maloney #24)

4. In response to a question, you indicated that you had seen a lot of "bad" sampling. Can you elaborate? How do you determine when sampling is appropriate or effective?

A. Probably the most blatant case was the use of sampling in furtherance of tax fraud. It is an interesting story, but (fortunately) relatively rare. A much more widespread abuse is the use of methods that produce wide error bands so that preferred interpretations are not excluded as inconsistent with the data. Another variation on this theme is the use of methods that are not robust so that different combinations of assumptions can be tested by the policy shop until some

preferred result is produced.   Statisticians know these risks better than most, but too often they salve their conscience with "caveats" and let the advocates do the damage.

Some of the red flags are:

> 1) "new" methods that have not been evaluated,
> 2) the choice of methods with large variance or bias when tighter methods would make the implicit assumptions more visible, and
> 3) use of non-robust methods (that are sensitive to assumptions) when more robust alternatives are available.

OMB frequently requires a SUCCESSFUL evaluation before approving such methods, but this doesn't always work.

5. In response to a question you implied a relationship between errors in measuring undercount and strategies for reducing the differential undercount. How does the former affect the latter?

A. This is discussed in detail in my response to Maloney #6 and #19.