I should like to commend the Bureau of the Census for the serious efforts it has made and continues to make to evaluate Census data. They have used varied approaches, have undertaken detailed analyses and the magnitude of the effort is very great indeed. They are to be commended for the high standards they set and for the pioneering work that they are undertaking.

At the same time I should like to commiserate with them because their difficulties are enormous. There is a tremendous movement of people each year and this makes it difficult for them to carry out an effective re-interview in order to determine how good the census itself was. For example, as is pointed out in the Marx-Waxberg paper, almost 10% of the people had moved between the census and the PES, although the PES (Post Enumeration Survey) was conducted within five months of the Census.

I was struck with the fact that the enumeration status of college students could not be determined for 7% of those in the sample; and the enumeration status could not be determined for 5.5% of Social Security recipients. Both of these are groups for whom current addresses should be excellent, at least relative to many other groups in our population. In both cases the Bureau of the Census had what appeared to be current addresses shortly before the census, but even so they were unable to determine the enumeration status of significant proportions of both groups. If these groups are so mobile, it seems likely that other groups are even more mobile.

There are a few points on which I would like clarification. In the use of record checks for coverage evaluation four population groups were included:

- 1. Persons enumerated in the 1950 Census.
- Persons missed by the 1950 Census, but picked up by the Post Enumeration Survey.
- 3. Children born during the intercensal period.
- 4. Aliens who registered with the Immigration and Naturalization Service in January, 1960.

In the Marx-Waxberg paper, it is stated "their combined representation is believed to be 98% or more of the entire population." In 1950 Coale estimated that 3.6% of the population was missed; but the Post Enumeration Survey accounted for only 1.4% or 2.2% less than the Coale estimate. The minimum reasonable estimate prepared by the Bureau of the Census was 2.5% missed or 1.1% more than was accounted for by the PES. It seems to me that 2.5% or possibly more of the population was not included in the record check for coverage evaluation. I would prefer, therefore, that the emphasis be on 98% or less rather than 98% or more. This may seem a trivial point but I think that it reflects a philosophy that runs throughout the papers, namely, that those preparing estimates of a corrected count tend to use minimum figures wherever possible rather than maximum or even expected or medium estimates.

Using the record checks for coverage evaluation, one estimates an under-enumeration of 2.6 to 4.7%. But this represents undercoverage only. Net error is estimated by subtracting from the underenumeration the overenumeration, estimated to be 1.3%.

We have been told that "the technique used in 1960 defines both over- and underenumeration relative to a specified small area, say, an enumeration district that contains the usual residence of the person being checked." The estimated gross overenumeration of 2,325,000 persons or 1.3% of the total would appear to have been based on this small area concept. This concept is a perfectly good one if one is deriving the estimate of underenumeration in the same way; however, when the record check is used one is not using the small area concept but rather the entire United States is the basic area. That is to say, a list sample has been prepared and the Bureau of the Census searched every enumeration district in which they thought it might be likely to find the individual. Thus the individuals on the list were counted as being correctly enumerated if they were found in any one of several enumeration districts. Overenumeration, however, was estimated on the small area basis, which overstates overenumeration in the entire United States. I am not entirely sure of the procedures used, but in reading the papers it seems to me that this was the procedure and it is for this reason that I ask for clarification on this point.

One of the puzzling features of the composite estimate is that more males than females were missed in every age group up to 45 years of age. The really surprising thing about this is why male babies and young male children in the age group 0-4 and 5-9, for example, would be missed more than females. The re-interview data do not show that more males are missed. Thus the apparent undercount of males in the 0-4 group, for example, is the result of the estimate of the expected number of males. The estimating procedure includes an allowance for under-registration of births and possibly this leads to an inflation in the number of males. Basically the procedure is to take the number of females in the specified age group, apply an expected sex ratio to this number and thus estimate the number of males. It is quite possible that the expected sex ratio is incorrect. The magnitude of the sex ratios would appear to be all right, but when they lead to an inconsistent result of this type one wonders whether the basic assumption is correct or if the re-interview data should be trusted more for this age group than is the case.

I am also puzzled that the sex ratio for the 5-9 age group is larger -- trivially larger, but still larger -- than for the 0-4 age group among whites.

In reading the paper one finds such terms as the "conservative best estimate" and the "minimum reasonable estimate." For some purposes it is desirable to derive minimum figures and in any case one would be inflating the actual census count. Nonetheless, the philosophy behind this bothers me in that we need a best estimate rather than a "conservative best" or "minimum reasonable" -- why not simply a "reasonable estimate?" I know that for administrative purposes it would be useful to have a single set of figures but, at the same time, it seems to me that there are too many uncertainties in the estimating procedures to permit us to afford the luxury of a single set of corrected figures.

There are two minor points in the Segal-Zelnick paper that I should like to comment on. On page 6A the statement is made that "The assumption of the net undercount of females 15-29 in recent censuses, combined with the fact that births estimated from females 15-29 in one census are approximately equal to births estimated from females 15-29 in the preceding and following censuses, led to the assumption of the uniform net undercounts over time. Literally, this seems unlikely; they probably are talking about ratios being equal rather than the numbers and this should be indicated.

Segal and Zelnick also state that since the number of births in any period of time is considerably larger than the number of deaths and net migrants, errors in the completeness of birth registration are of greater consequence for estimates of net census underenumeration and net census undercounts than errors in the other components. This is true if the magnitude of the errors is the same in each of the components but is not necessarily true if errors are larger in one component than in another and, indeed, in this instance I think that errors certainly are larger in some components than in others.