## DISCUSSION

## Joseph Steinberg, Social Security Administration

To users of census data, the two papers presented this morning are welcome additions to thowledge. We are a good deal closer today to some worthwhile goals as a result of this work--the goals of improved postcensal estimates of population, by age and sex--and of knowledge where errors in the count occur.

The Marks-Waksberg paper summarizes the results of a large volume of case-bycase evaluative work on coverage of the 1960 Census. It is, probably, as close as anyone can come to a final summary of substantial direct evaluative efforts. It is interesting to note that the final reinterview estimate of net undercoverage in the 1960 Census is 1.9 percent of the population; and that record checks give a range (with a "bolder attitude") of 2.5 to 3.1 percent. The Siegel-Zelnik paper takes account of recent evaluations of the assumptions in some analytic techniques and presents the results of those analytic methods now considered as preferable for the 1960 Census situation. The estimated net underenumeration by demographic analysis is 3.1 or 3.2 percent.

In working toward the goal of a "corrected" distribution of the population by age, sex, and color, the Siegel-Zelnik paper presents three experimental sets of estimates. They state that "the absolute level of the adjusted census or current estimates at each age (except possibly 65 and over) would be closer to the theoretical truth." But, they conclude that they have so far been unable to arrive at a set of figures to be recommended for general use.

Users of decennial census data have had available "some estimates of net undercounts for 1960 by age, sex, and color" in the United States Summary volume of the 1960 Census of Population. The analytic results in the Siegel-Zelnik paper are quite different for both white males and females 65 and over, (and by age group) for nonwhite females, and to some lesser degree, for nonwhite males. These substantial changes reflect the differences between the analytic methods used in the current presentation and the results in the Summary volume of the 1960 Census. Users have also had results in the Current Population Reports, Series P-25, No. 310, which also differ from the currently presented analytic results. As we examine the Marks-Waksberg paper and related matters, a few additional observations seem pertinent.

1. Many users might agree that not only is the most important statistic from

the Census the population count, but so, too, in the postcensal period is this true--and by age and sex.

2. Useful information is available from the case-by-case approach on the relative contributions to coverage errors from field work and imputation procedures. About 10 percent of the undercount and 40 percent of the overcount arose in the processing operations. Net undercoverage due solely to field work is estimated as 1.9 percent.

3. The paper states that the reinterview results on the measures of age reporting are rather poor estimates of the bias in age reports. The user hopefully might look forward to some final summarization and greater elaboration suggesting the basis for this statement--and also similarly in regard to other items such as income, etc.

4. While data are not yet published, one might look toward the results of the recent registration for Medicare as an additional basis for evaluating the estimates of net undercoverage for the equivalent age groups in the 1960 Census. Likely, one will find some undercoverage. The counts of aged registrants would provide a lower bound to the estimates. The next few years will indicate the extent of underregistration for Medicare--as nonregistered aged persons apply for hospitalization benefits, the validity of their age allegations are examined and substantiated through appropriate means, and a range of estimates of underregistration are derived. At that point, a more reasonable estimate of undercoverage will be possible.

As we examine the Siegel-Zelnik paper, a few additional observations also seem pertinent.

1. One must agree that the techniques of demographic analysis often provide a strong basis for judging the demographic reasonableness of census results or other methods of evaluation. However, as the authors point out, there are a number of limitations to these techniques (albeit there are limitations--whether smaller or greater--in other evaluative techniques). One might, therefore, assume that evaluative analytic techniques (and various derived results) may continue to be presented at intervals based upon alternatives not as yet considered. One would hope that despite the important research considerations, that one of the existing results (modified as necessary) can be recommended for use.

2. It is interesting to observe, that the several levels of estimates of relative undercoverage, indicate that efforts to improve coverage in 1960 by various means do appear to have worked to some degree--percentage-wise up to 0.7 or 0.8 percent better than 1950.

3. The authors mention an anomaly between the Census of 1950 and 1960--an apparent large net overcount of persons 65 and over in 1960--an anomaly which has now appeared in three successive censuses. For 1960, could this more likely reflect the better counting of persons at 65 and over than errors in age reporting? For 1960, the unpublished evidence from the Medicare registration in 1965-1966 may perhaps support this hypothesis.

4. The finding of Coale and Zelnik that "females 15-29 appeared to be the most fully enumerated group," in 1950 is strikingly inconsistent with the 1960 preferred composite result based on demographic analysis. And, the 30-44 group among white females is now best! Did the method produce the result, or are we to assume that the same cohort was enumerated well in 1950 and 1960.

5. If one places credence in reinterview net coverage error results for nonwhite females under 5, 5-9, and 10-14, what would the rate of underregistration of births be that would be consistent? Would this offer some added possibility anent nonwhite males?

6. The present analytic technique involves the use of expected sex ratios. The authors state that their research indicates that these expected sex ratios are rather sensitive to the level of net undercounts at the higher ages. Possibly the authors may wish to suggest in later discussion the implications of this comment with respect to the levels of estimates that may be derived for age sex groups--and the reasonable bounds on their estimates of the degree of sensitivity in their present use of the technique at such higher ages.

7. One can certainly join in their concern about the need for measurement of differential coverage error. For overall evaluation, there is need for adequate estimates with respect to geographic variation. There would also seem to be substantial need for measurement and reporting by socio-economic classifications. Users of census counts (and census data classified by a variety of characteristics) often assume them to be accurate. As noted later, major policy perhaps may have somewhat inadequate factual underpinning if unadjusted census results (whether unadjusted for undercoverage or biases in, say, age reporting, are the basis.

8. This user would prefer that a single "best" estimate be recommended together with a technical note on the range of undertainty. While the range would not be a probability confidence interval, it could nevertheless furnish useful information. In this way, if a user chose to use the upper limit for cost analysis or for a target, it would of necessity be so identified. This would still leave the needs of the general audience untouched and uniformly served.

Evaluative work of the type reported in these papers is important to the producers and the users of data.

To the producers, the challenge is:

1. To find improved techniques for doing a better job next time.

2. To find improved techniques for doing a better evaluative job.

3. To take advantage of the synthesized knowledge for providing the user with "better" population estimates. Would that this third challenge could provide the user, at an earlier date, with uniformly available, authoritatively backed, results.

The Census staff continues to examine methods for improved census taking. Problems of census taking and survey operations in slum areas have been identified since, at least, the 1950 Census as of a much higher order than in many other situations. Would that the properly motivated local people presently pressing for national recognition of the need for improvement in the lot of the poor could be effectively challenged to understand the need for and to help work toward improved data on coverage and content in their local areas.

The thrust toward improved evaluative methods has been suggested by Marks-Waksberg as coming perhaps through greater use of matching of special lists. This effort is one which deserves support-with substantial resources made available (perhaps some as early as this fiscal year) so that as many as possible of the identified limitations in the use of this technique (i.e., size of sample, or failure to get current addresses) can be overcome.

Perhaps, consideration should also be given to a method currently in use in another federal agency--record checks after an evaluation reinterview.

The suggestion to provide "best estimates" of postcensal population by age and sex is made with full realization that this creates a major problem for the provider. But certainly one might agree that the absence of such authoritative estimates leads to chaotic multiple estimates or the use of "comparable to Census" estimates when, in fact, the evidence we have heard today is that significant underestimates (in program decisions) may be the result of use of unadjusted census data. The relevant current data for many classes of decisions involve in large part the accuracy of current estimates of population by age, sex, and color. The number of persons who are unemployed and employed, as estimated from the Current Population Survey, are now and have virtually always been tied to "comparable to Census" estimates of population by age, sex, and color. As we examine the estimates of undercoverage in the census, shouldn't we be concerned as to the impact on this use, especially when decisions based upon whether we have full employment or the likely current volume of unemployment, are involved? (Besides the CPS survey results are already being adjusted for undercoverage relative to "comparable to Census" estimates.) Similarly, estimates of the current size and change in the number of the poor and their characteristics, and the possible levels of cost of various alternatives under proposals for a negative income tax are likely to be changed. Other estimates such as the number of persons eligible for Government programs such as Medicare; the number of persons living in substandard housing or eligible for rent subsidies, etc., get changed. It would appear useful for statisticians to face up to the question, and to answer affirmatively, that there is need for publication of a single set of official authoritative figures on "best estimates" of the population by age and sex, not only for the cnesus period, but also on a postcensal basis--and that these be used in the preparation of all important series.