DISCUSSION

Katherine K. Wallman, Office of Federal Statistical Policy and Standards

The presentations in this session have focussed primarily on recent initiatives by a number of Federal agencies to implement changes in the collection of racial and ethnic data required by Statistical Policy Directive No. 15. To put these efforts in context, I would like to spend a few minutes discussing the background of this statistical standard, and will then proceed to comment more directly on the findings presented by the authors.

Directive No. 15 was issued initially to serve two purposes. One of those purposes was to improve the comparability of data across Federal agencies, so that when data were available from the Social Security Administration, from the Office of Personnel Management, from the Equal Employment Opportunity Commission, and other agencies, they could be analyzed in conjunction with basic Census data or in other combinations. The second driving force, and the one that brought the issue of standardizing the racial and ethnic categories to the foreground, was an interest in reducing respondent burden. This was burden not only on individuals but more importantly, I think, burden on institutions and businesses who were required, primarily for various Federal compliance programs, to keep these data in just slightly different ways for the various agencies. The required categories were basically the same, but a little different in construction or in definition of who was included in a particular category. For those two reasons, that is, to increase comparability and to reduce burden, our Office, in response to requests by Federal agencies as well as by the public, agreed to issue a standard for racial and ethnic categories and definitions.

The categories that are in Directive 15 represent, by and large, population groups of interest to the Federal Government. For the most part, both the major groups and their respective components are those about whom the various Federal agencies historically had collected data. A choice was given in the Directive with respect to the format for collecting racial and ethnic data. The preferred alternative for collection of data under Directive 15 was, in fact, the two-question format, which had separate questions on race and ethnicity. The reason for that was very simple—it gave the analyst the greatest amount of power in using the data. Further, it provided the opportunity to get full counts of Blacks and Whites.

The combined alternative, on the other hand, was presented to meet the need which strongly influenced our issuing the Directive, that is, to reduce respondent burden—not only for individuals, who are respondents in the cases which have been discussed here this morning, but more importantly for institutions and for businesses who must keep records on an ongoing basis to report to EEOC, the Office for Civil Rights, and a variety of other Federal agencies. These respondents indicated that it would be simpler for them to keep their records in this way. Thus, this alternative addressed the burden reduction aspect of providing racial and ethnic data.

With this background, I would like to turn now to a discussion of the papers which have been presented this morning. I would like first to make two general comments with regard to the findings which have been outlined. I think the most striking thing that has come out of all this has been the fact that the combined standard has proven to be a better question in terms of increasing response rates. I mentioned earlier that our primary purpose in having the combined alternative was burden reduction; the findings of the agencies indicate that there may be an unanticipated benefit in using the combined response option. This alternative does suffer from the problem of not providing the ultimate in analytical opportunity; further, as Phil Schneider noted the combined alternative makes it difficult to expand the categories and subcategories. These are obviously tradeoff kinds of problems. The other general comment I would like to make is to recognize the contributions which these testing activities are making not only to the agencies who are conducting them right now, but also to the many other agencies who will be using this standard in various ways either in statistical surveys or in administrative and other kinds of compliance reporting activities.

In terms of the specifics of the papers, I would like to highlight a few of the things that have emerged from the findings presented here this morning. The need to better inform respondents of the use of the data is something which the Social Security Administration papers have highlighted for us. Another thing that was not mentioned from the platform this morning, but was mentioned at some length in the papers, is the need to put whatever detail the respondent should have concerning the categories and their definitions directly on the form. To have details concerning the components of the categories in the instructions essentially means this information will not be taken into consideration by the respondent.

Certain dilemmas which we had anticipated when the standard was developed continue, and were evidenced in the testing which has been done by these agencies. One of these is the dilemma created between reducing burden by asking only about Hispanics in the ethnic question (in the two-question format) and the concerns of respondents who are confused when only Hispanic heritage is questioned. Another dilemma which has been pointed out in several of the papers concerns the effect of the use of the word "voluntary" on response rates. On the one hand, we want, and in fact are required, to fully

inform the respondent of the voluntary nature of the question; at the same time, this information impacts on the respondent's desire to answer the question.

A dimension added by the OPM paper which I think contributes particularly to our knowledge at this time concerns computer conversion of racial and ethnic information. There are a number of administrative record systems where there may, in fact, be a desire to do some further work in trying to convert prior records (records kept prior to the implementation of the new standard) to get more historical, longitudinal information. What OPM has done will be useful in setting the ground for others who want to investigate that area further.

The need for visual identification to supplement self-identification has been brought out not only in OPM's testing, but also in the Social Security testing. Obviously, there is no readily available opportunity for visual identification if the administering agency is not present when the applicant completes the form. There will have to be some tradeoff analysis concerning the need to supplement and improve the information versus the cost of so doing.

In the Fernandez-McKenney paper, which contains a wealth of insights on the collection of data on the Hispanic population, there are also some very basic findings which have broad applicability. One that seems so simple, but which is a lesson for all of us, concerns the placement of the "no, not of Hispanic origin" option in the item responses. Another general finding concerns the problems that the Census Bureau is having in trying to use the open-ended questions on ethnicity. Many of us have been desirous of having the question that way, but clearly there are problems in terms of response which argue against the open-ended format in most cases.

One thing that I personally find very interesting in the Fernandez-McKenney findings is the misunderstanding of the term Mexican-Americans, and the related implications of this outcome. When we were putting Directive 15 together initially, we had terms such as Native-Americans and Mexican-Americans and Asian-Americans. It was the sense of the representatives from the Federal agencies who were contributing to developing this Directive that those kinds of terms would tend to cause over-reporting, because of the tendency for people to say, "I am an American." Apparently, our sense was correct.

Lois Alexander has raised some more general questions about the voluntary versus mandatory collection of racial and ethnic information. This has been a tradeoff problem in many areas. We continue to be concerned about the conflicts inherent in the Privacy Act and the Freedom of Information Act, and other constraints which are brought on by legal and other concerns. I don't know that any of us has a quick answer to this problem. It clearly will be an area for further investigation. One concern I have about the possibility of increasing the release of this information, either as a consequence of the items being made mandatory or for some other reason, is the possible misuse of this data down the road. We have heard a specific case here this morning about the EEOC's interest in having this data, which fortunately has been resolved. As data of this kind become more available, the possibility for misuse increases. One of the things that we find is that when people find a body of data, they tend to go ahead and use it without thinking too much about whether it is really appropriate to their purposes. There are certain problems in using Social Security data for some of the purposes which have been suggested. For example, when one tries to determine applicant pools using Social Security data, there is frequently no information on an employee's current occupation, nor is there information on his availability for a particular positon. Thus, it is only with great tenderness and respect that this kind of data should be used in the compliance arena.

In conclusion, let me simply say that the work which has been completed at Social Security, OPM, and Census has contributed significantly to an understanding of this very complex and sensitive area of data collection. The findings of these researchers will be of substantial use to Federal agencies, as well as to others in both the public and private sectors. We look forward to continued progress in this very difficult area.