

RACE/ETHNIC DATA COLLECTION IN THE FEDERAL CIVIL SERVICE

Philip A. D. Schneider, U.S. Office of Personnel Management

In 1977 OMB issued Circular A-46 which levied on Federal agencies new standards for reporting race and ethnicity statistics. Subsequently, this OMB standard was subsumed into Commerce Statistical Policy Directive 15 entitled "Race and Ethnic Standards for Federal Statistics and Administrative Reporting." Since 1972 the Office of Personnel Management (formerly the U. S. Civil Service Commission) has maintained the Central Personnel Data File (CPDF). This automated file contains records on nearly all active Federal civilian employees; currently the file contains approximately 2.9 million records with 41 elements of data for each employee. One of these elements of data is the current Minority Group Designator (MGD) which includes the following categories: Negro, Spanish-surnamed, American Indian, Aleut and Eskimo in Alaska, Oriental, Non-Hispanic in Puerto Rico, Non-designated in Hawaii and Guam, and Non-minority. The data in CPDF are updated monthly by personnel transactions that occur throughout the Federal establishment. The volume of transactions into CPDF is annually in excess of 10 million actions. Most Federal agencies also maintain automated personnel information systems analogous to CPDF for their own employees.

The advent of the OMB A-46 and Commerce Directive 15 Standards required us to revise the MGD to accommodate the new categories and definitions in the standards. We were then faced with three alternative decisions. Which data standard allowed in Commerce Directive 15 should we implement? Commerce Directive 15 allows a separate standard for race and ethnic data elements so that an individual can be reported for race and reported separately for ethnicity, thus allowing cross classifications of both race and ethnicity. Directive 15 also provides for an alternative combined race and ethnicity data element in which all possible combinations of race and ethnicity cannot occur for reporting purposes. The Commerce Directive 15 combined standard includes the categories of Black, Hispanic, Asian and Pacific Islander, Native American and White. Neither Directive 15 separate or combined standards allow for an "Other" category of reporting. The second decision area we were faced with is the collection methodology. Commerce Directive 15 does not require either self-identification or visual observation, but leaves the choice up to the implementing agency. The MGD data currently in CPDF were obtained and are maintained on a visual observation basis. Our decision then required us to determine whether we were to continue with visual observation or move to a self-identification mode in implementing Directive 15. The third decision area involves the methodology of converting a major government-wide system like CPDF (and local systems in each Federal agency) to accommodate the Commerce Directive 15 standards. Our alternatives here were to either resurvey the entire 2.9 million Federal work force to collect data under the new definitions, or to computer convert the current data to the new standards if

statistical accuracy would support such a decision.

Prior to making these decisions we determined that we should develop a pilot test to more fully explore what the implications of these decision alternatives might be for statistical consistency, accuracy and the continuation of meaningful trends on minority employment in the Federal civilian work force. Our initial inclinations prior to the pilot test were to go with the separate standard as allowed for in Directive 15 because it theoretically provides more data and is easily expandable for additional ethnic classes beyond Hispanic. We also felt that self-identification was the appropriate collection methodology since previous speculations about errors in the process -- particularly for those employees who were "Spanish-surnamed," suggest that the employee is the best judge of his or her race and ethnic category. Finally, our inclination was to go with a complete resurvey because of unknown errors in computer conversion to data previously determined on a visual basis and particularly because members of the Spanish-surnamed group, for obvious reasons, do not appear to well lend themselves to accurate visual identification.

Our research methodology for developing the pilot test included the following items. We decided to stratify the sample of employees to be included in the pilot test by those factors which we could articulate might affect how employees self-identify their race and ethnicity, particularly vis-a-vis the previous visual reporting base. To this end we developed the following seven strata: (1) minority groups currently in the CPDF, i.e., Negro, Spanish-surnamed, American Indian, Orientals, and Other (Non-minority); (2) geographic location (east, midwest, west); (3) occupation (professional/administrative, and technical/clerical/other occupational groups); (4) salary (to 11599, 11600 to 28499, and over 28499); (5) sex; (6) age (to age 40, over 40); and (7) education level (through high school, and beyond high school). We then developed a stratified random sample from CPDF which involved approximately 100,000 current employees. We used a reporting instrument which requested the pilot test employees to self-identify their race and ethnicity according to both Commerce Directive 15 separate and combined standards. We also in this stratification allowed for a 50% inflation factor due to potential but unknown non-responses, and to what might be categorized as frivolous responses (a significant problem when self-identification was tried in the late 1960's). The aim of the instrument was to provide the necessary data to allow matches to be developed for each sample employee between the previously obtained visual determination of minority group designator and the self-identification under the new Commerce Directive 15 separate and combined standards. Management of the pilot test data collection exercise was very tightly controlled through approximately 1100 operating Federal personnel offices around the country. This is

one of the very few cases to my knowledge in which these sensitive data are available on both a pre and post change situation.

Let's turn now to the results of the pilot test itself. First, we found no significant impact in terms of reporting response to any of the sub-strata variables. Table I illustrates the match rates that were achieved between the current visually determined MGD's and both the separate and combined standard categories under Commerce Directive 15. A "match" is defined as a comparison between the visual and self-ID responses, e.g., visual "Negro" matching a self-identified "Black." These matches were defined to assess the possibility of a straight conversion from the old to the new standards without necessitating a government-wide recollection of data. It is important to note that we did not establish either the visual or the self-identification responses as a measure of defacto accuracy. We were only interested in the impact of a computer conversion on current statistical trends in accuracy and completeness.

Table I indicates that the "Negro" and "Non-minority" match rates were between .935 and .947. (Note throughout here that matches neglect non-responses on the pilot test which could not be matched against existing MGD categories.) However, the current Spanish-surnamed category matched the Hispanic category at a rate of only .608 on the separate standard; whereas, for the combined standard the match rate between Spanish-surnamed visual and Hispanic self-identification was .859 -- a rather dramatic increase. The match rates for American Indian and Orientals do not show any significant differences between the separate and combined standards. The question then arises as to why the current Spanish-surnamed employees responded far higher in match rates on the combined standard Hispanic than they did to the separate standard Hispanic.

Table II illustrates the distribution of current Spanish-surnamed employees among the various categories on the separate standard (race and ethnicity) and on the combined standard. On the separate standard, of the current Spanish-surnamed employees, 26.9 percent responded as Hispanic of "Other" races. (For the purposes of the pilot test, although not allowed in Commerce Directive 15 implementation, we included an "Other" category for both separate and combined standards responses.) On the other hand, on the combined standard 86.1 percent of the Spanish-surnamed employees responded as "Hispanic." Only 6.4 percent responded as White not-Hispanic, which compares to 6.1 percent responding White, non-Hispanic on the separate standard.

It is also interesting to note that the subgroups of Hispanic employees reporting American Indian or Black races are negligibly small statistically in the Federal work force -- each responding on separate standard matches at less than 1 percent. This puts to rest a previous criticism of the Spanish-surnamed definition for not allowing a combination of Negro/Spanish-surnamed or American Indian/Spanish-surnamed. For completeness purposes we also looked at the match rates of those employees identified visually as Negro or

American Indian to self-identification as Hispanic. On the the combined standard we found 0.7 percent visual Negroes responding as Hispanic and 1.4 percent visual Negroes responding as Hispanic on the separate standard. And 3.9 percent visual American Indians responded to self-identification as Hispanics on the separate standard and 1.4 percent on the combined standard. Therefore, by moving to the combined standard we do not miss significant numbers of employees who are mixed American Indian Hispanics or mixed Black Hispanics.

Looking at the responses in the pilot test on the separate standard, Hispanics did not choose a race in about 26.9 percent of the cases, preferring to respond as Other races (i.e. Non-White, Non-Black, Non-American Indian, etc.). On the combined standard, which contains a forced choice between Hispanic ethnicity and some race, the response of Spanish-surnamed employees is evidently much more complete. For whatever reason, therefore, it appears that Hispanic Federal employees will not respond completely to a request for racial self-identification on race and ethnicity separately. It might also be noted from Table II that the category of Spanish-surnamed employees totally identifying themselves as Hispanic for any race on the separate standard was 90.4 percent. We believe that the 9.6 percent residue of current Spanish-surnamed who do not report themselves as Hispanic represents the error inherent in the definition of "Spanish-surnamed" versus the definition of Hispanic. This so-called naming error results when, for example, Maria Gonzales marries John Smith and ceases to be Spanish-surnamed or Maria Smith marries John Gonzales and begins to be Spanish-surnamed. The end result from this analysis is that we cannot rely on the separate standard for race and ethnicity under Commerce Directive 15 without an "Other" race category to provide reliable and particularly complete reporting. Therefore, we have decided to go with the combined standard under Commerce Directive 15 to insure maximum reporting of Hispanics in consistent categories. We also believe that this pilot test presents a significant piece of information to be considered on the viability of continuing the separate standard as a reliable option for reporting.

Let's turn now to an assessment of the non-responses which we experienced in the pilot test itself. There were four kinds of non-response. The first kind of non-response was due to censoring of the initial stratified sample. The initial sample included a very small number of Federal employees each in several hundred operating personnel offices. Approximately 7500 of these initial sample employees were censored from the survey sample to reduce reporting burden in these personnel offices. An analysis of the characteristics of this censored sample under the initial strata did not indicate any bias due to the censoring. The second sort of non-responses were those employees to whom the pilot test instrument was undeliverable, i.e., they had died, retired, transferred or resigned from the work force subsequent to the month of CPDF used to generate the sample. An analysis of these undeliverable responses did not indicate significant bias under the pilot test strata.

The third kind of non-responses, which we call "voluntary non-responses," were those employees who selected the option provided on the test instrument of "choose not to respond" and provided no information on self-identification of race or ethnicity on either standard. These non-responses are contained in Table III. This fourth kind of error we call "item non-response," and it affects primarily the 26.9 percent of Hispanics who non-responded for race.

If we look at the distribution of non-response rates by current minority group designator for employees in the pilot test, we find that the voluntary non-response is very uneven. Without some action, we would bias the target reporting towards the non-minority groups since these are the largest groups in the Federal work force but their non-response rates are the smallest. By not making some provision to gather data for those employees choosing not to respond, we then would over-magnify the incidence of non-minority and under-represent the true incidence of various minority categories in CPDF. Therefore, in the area of collection methodology we have decided to indeed move to self-identification but, to recognize the non-response problem, we are providing for a visual supplement by the appointing official when a new hire Federal employee chooses not to provide a response to the invitation for self-identification of race and ethnicity. In order that we make no assumptions about the distribution of voluntary non-responses in future collection efforts, we are going to provide an opportunity to each employee to review the conversion methodology from the existing MGD categories to the Commerce Directive 15 combined standard categories and show how those conversions were accomplished. The employee then will be provided the option to learn what category he/she has been converted to and provide

a chance if desired to change that conversion item. This is particularly important for those employees who have origins in the Indian Sub-continent, since prior to Commerce Directive 15 this group was not considered a minority group but is included in Commerce Directive 15 under the category of Asian and Pacific Islanders.

To recap, then, our initial decisions and inclinations were rather totally reversed by the results of the pilot test. Our initial inclination to go with the separate race and ethnic standard in Commerce Directive 15 died in face of the item non-response for Hispanics. Therefore, we have decided to go with the Commerce Directive 15 combined standard for race and ethnicity. Our tentative decision initially to go with self-identification was confirmed, however with a major supplement by needing to provide a visual observation for those employees who choose not to respond. Finally, on the conversion methodology our initial inclination to resurvey the entire work force was found to be unnecessary since, for the largest group of employees, Blacks and Whites, computer conversion from existing MGD categories to Commerce Directive 15 categories will match in 94 percent of the cases; this portion of the work force amounts to about 93 percent of the total work force in the Federal government. We believe that presenting the employee with the option for knowing how the conversion went and for providing changes if the employee wishes is a reasonable accommodation to the 6 percent non-match rates experienced in the pilot test.

The changes I have outlined herein have been approved by top management in the Office of Personnel Management and the Equal Employment Opportunity Commission and will be implemented on January 1, 1981.

Table I: MATCH RATES OF CURRENT MGD'S TO CD 15

	Negro	Spanish-Surnamed	American Indian	Oriental	Non-minority	Total
Separate Standard	.935	.608	.840	.859	.945	.929
Combined Standard	.939	.859	.863	.923	.947	.941

Table II: DISTRIBUTION OF SPANISH SURNAMED TO CD 15 SUBGROUPS

SEPARATE STANDARD

<u>American Indian</u>		<u>Black</u>		<u>Asian/Pacific Islander</u>		<u>White</u>		<u>Other</u>		<u>Total</u>	
H	Non-H	H	Non-H	H	Non-H	H	Non-H	H	Non-H	H	Non-H
0.8	0.5	0.8	0.5	1.0	1.7	60.9	6.1	26.9	0.8	90.4	9.6

COMBINED STANDARD

<u>American Indian</u>	<u>Black, not Hisp.</u>	<u>Asian/Pacific I.</u>	<u>Hispanic</u>	<u>White, not Hispanic</u>	<u>Other</u>
0.6	0.5	2.6	86.1	6.4	3.9
			<u>Separate</u>	<u>Combined</u>	
Visual Black to Self-Identified Hispanic			1.5	0.7	
Visual American Indian to Self-Identified Hispanic			3.9	1.4	

Table III: NON-RESPONSE RATES OF CURRENT MGD'S

<u>Black</u>	<u>Spanish-Surnamed</u>	<u>American Indian</u>	<u>Oriental</u>	<u>Non-minority</u>	<u>Total</u>
.310	.255	.266	.199	.193	.244