CDC-ATSDR Workshop

Issues Regarding Data on Race and Ethnicity: The Census Bureau Experience

NAMPEO R. MCKENNEY, MA
CLAUDETTE E. BENNETT, PhD

Ms. McKenney is the Assistant Division Chief, Special Population Statistics, Population Division, Bureau of the Census. Dr. Bennett is a Statistician (Demography), Racial Statistics Branch, Population Division, Bureau of the Census.

Tearsheet requests to Nampeo McKenney, U.S. Bureau of the Census, Population Division, Room 2312, FOB 3, Washington, DC 20233, tel. 301-763-7445; fax 301-763-3862.

Synopsis

In this paper, the authors describe some of the complexities of collecting and presenting data on race and ethnicity based on the experiences of the Bureau of the Census. Different methods of data collection, different content and format of questions, and different definitions make it difficult to collect consistent race and ethnic data across data systems. The Bureau of the Census experiences have shown that changing ethnic self-identity and concepts, intent of the question, consistency of reporting, and the classification of persons of mixed racial parentage affect the quality of the data. These are some of the issues that must be addressed as statistical agencies and researchers seek to provide comparable race and ethnic data.

DATA COLLECTION on race and ethnicity is complex. This paper identifies some of the complexities and difficulties encountered by the Bureau of the Census in collecting, tabulating, and publishing data on race and ethnicity. In particular, attention is focused on those issues that have implications for the public health surveillance data systems. The complexities of collecting data by race and ethnicity will increase in the future as the United States is experiencing substantial changes in the racial and ethnic diversity of its population. Decisions about how and what to collect, tabulate, and present for the various racial and ethnic groups definitely are influenced by such factors as demographic changes, immigration trends, changes in ethnic and racial identity, legislative needs, and public policies.

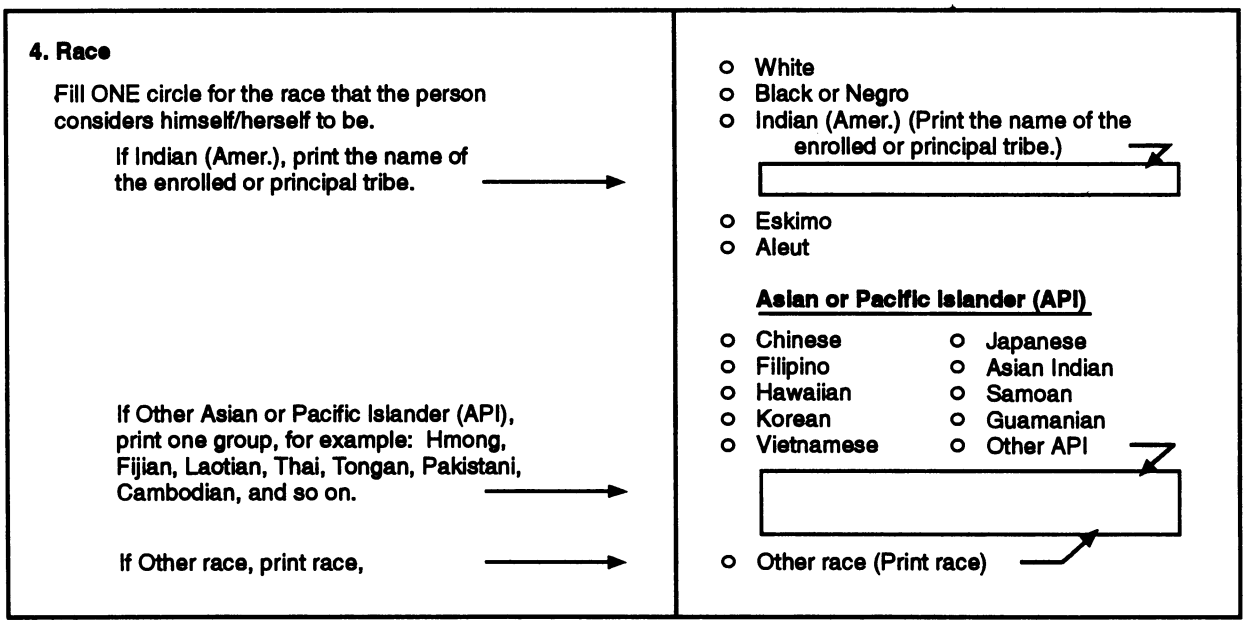
The paper is divided into four sections. Section one describes the Bureau of the Census' concepts of race and ethnicity; section two presents an overview of the demographic trends for the major racial and ethnic groups—White; Black; American Indian, Eskimo, and Aleut; Asian and Pacific Islanders; and Hispanic origin. Section three covers the relationship between the public health surveillance and the census data; and section four focuses on current and future issues on race and ethnicity facing the Bureau of Census. Although the major emphasis of the paper is on the race item, the Hispanic origin item, one of the primary identifiers of ethnicity, is also discussed.

The Bureau's Concepts of Race and Ethnicity

The Bureau of the Census has traditionally treated race and ethnicity as two separate concepts. The racial and ethnic classifications used by the Bureau of the Census generally adhere to the guidelines of the Federal Statistical Policy Directive No. 15 which was issued by the Office of Management and Budget in 1978.

This directive stipulates that Federal agencies are to collect and present data on at least four racial groups—American Indian or Alaska Native, Asian or Pacific Islander, Black, and White; and one ethnic group—Hispanic (*1*).

Race. The Bureau of the Census has collected information on race since the first census in 1790. Over time, the Bureau has used different question formats, content, and terminology. For example, the race response categories on the census forms have included a mixture of various principles and criteria such as national origin, tribal affiliation and membership, and physical characteristics. In addition, the Bureau of the Census has used both the enumerator's observation and self-identification



to collect data on race. Information on race is now obtained through self-identification. Prior to 1960, information on race was primarily based on observation by the enumerator.

The question on race in the 1990 census was asked of all persons. The race concept reflects self-identification by the respondents. Persons were asked to report the one race with which they most closely identified. The Bureau of the Census did not provide a definition of race for the respondents. Evidence from census studies showed that respondents would answer according to their own self-perceptions of race.

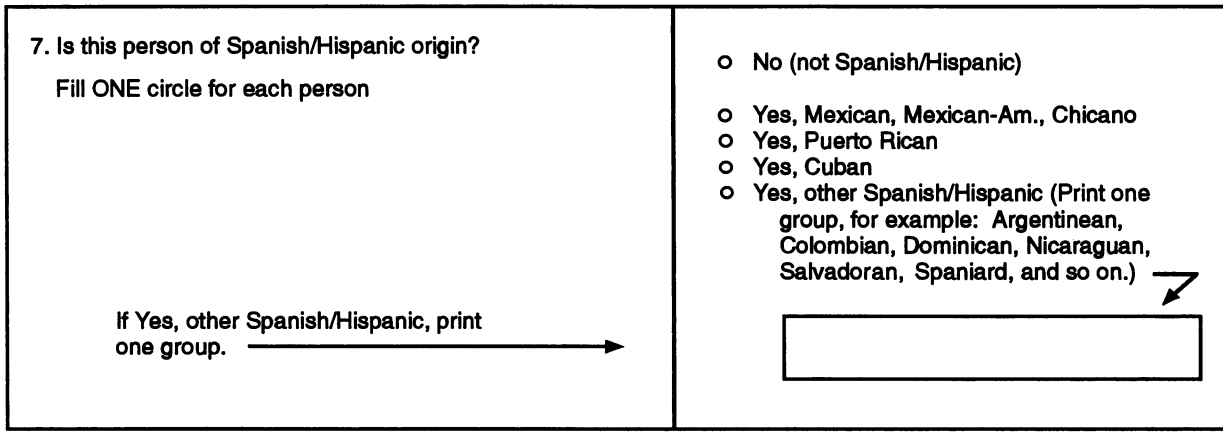
The 1990 race question, as in previous censuses, included a number of sociocultural (national origin) groups (see box). The question had 14 specific categories—White; Black or Negro; Indian (Amer.); Eskimo; Aleut; and 9 Asian and Pacific Islander groups; as well as two residual categories, "Other API" (Asian or Pacific Islander) and "Other race." Three categories required write-in entries. Persons reporting Indian (Amer.) were asked to write in their enrolled or principal tribe. Those reporting "Other API" or "Other race" were asked to write in their race. Although the Bureau of the Census generally adheres to the guidelines of the directive, the race item included an "Other race" category, which is not specified in the directive. Persons who did not identify with any of the specific race categories, especially those of mixed racial parentage, could report in the "Other race" category.

Hispanic origin. The Hispanic origin concept was first introduced in 1970 using the self-identification approach. Prior to 1970, the Bureau of the Census identified portions of the Spanish-Hispanic population through indirect measures based on birthplace of the person and the parents, mother tongue, and Spanish surname items.

The Spanish-Hispanic origin question included in the 1990 census also was asked of all persons (see box, page 18). The question listed a "No (not Spanish-Hispanic)" category followed by four Hispanic categories—"Yes, Mexican, Mexican-Am., Chicano"; "Yes, Puerto Rican"; "Yes, Cuban"; and "Yes, other Spanish-Hispanic." Persons who marked "Yes, other Spanish-Hispanic" were asked to write in their specific origin, such as Argentinean, Colombian, Dominican . . . and so on.

The Bureau of the Census is evaluating the race and ethnic data collected in the 1990 census. This evaluation is part of a larger extensive research program on race and ethnicity for the 2000 census. Its objective is to develop race and ethnic identifiers that will produce high quality data for the 2000 census and will meet major data needs. The need to improve the quality of the race and ethnic data is driven by the many important uses of the data, such as redrawing political boundaries, implementing legislation and programs, and funding governmental programs.

As part of the research program, the Bureau of the Census and Statistics Canada (our counterpart organization in Canada) co-sponsored an international conference on the Measurement of Ethnicity



in April 1992. The objective of the conference was to bring together persons in survey operations, academic fields, research, and general data user communities to discuss current and future theoretical and practical issues on ethnic measurement. Some of the themes and ideas that emerged from this conference are discussed in this paper and are available in published proceedings (*2*).

Demographic Trends

"Healthy People 2000" and the 1992 report of the Public Health Task Force on Minority Health Data issued by the Department of Health and Human Services strongly emphasize the need for additional race and ethnic data in the health field (*3,4*). Demographic trends on growth, which show increased racial and ethnic diversity of our population, support that conclusion. This diversity not only creates new challenges for statistical systems, but it underlies the increasing importance of racial and ethnic data for a variety of purposes, including public health surveillance.

In 1970, nearly 88 percent of the U.S. population were identified racially as White; 11 percent were Black; and 1 percent were American Indian, Eskimo, and Aleut, or Asian and Pacific Islander. Nearly 5 percent were of Hispanic origin. (Persons of Hispanic origin may be of any race.) Ten years later, in 1980, one-fifth of the population was either Black; American Indian, Eskimo, and Aleut; Asian and Pacific Islander; or Hispanic origin. That proportion grew faster during the 1980 decade; and by 1990, one of every four persons was one of these populations (table 1). According to the Bureau of the Census Middle Series projections, by 2010, that proportion will increase to one-third, and to one-half by 2050, as shown in the figure (*5*).

During the 1980s, the Black or African American population grew by 13 percent from about 27 million in 1980 to 30 million in 1990 (table 1). Most of the growth in the Black population was due to natural increase, the excess of births over deaths. Immigration from the Caribbean basin and African countries also contributed to the growth. The Black population is expected to increase to nearly 35 million by 2000, and double its current size to 62 million by 2050, as shown in the figure.

The American Indian, Eskimo, and Aleut population also increased substantially during the 1980s, from about 1.4 million in 1980 to nearly 2 million in 1990—a 38 percent increase (table 1). This population is projected to grow steadily during the 1990s to about 2.4 million by 2000, and almost double to 4.3 million in 2050, as shown in the figure.

The Asian and Pacific Islander population is one of the fastest growing populations in the United States. This population has more than doubled in each of the last two decades, and in 1990, it surpassed 7 million (table 1). Immigration from Asian countries, following changes in the immigration laws since 1965, accounted for the majority of the growth. Continued high immigration is expected to result in sustained growth of this population. By the turn of the century, the Asian and Pacific Islander population could expand to more than 12 million, and by the middle of the century reach 40 million, as shown in the figure.

It's important to note that the Asian and Pacific Islander population is comprised of a number of diverse groups that vary substantially in their population size and demographic, social, economic, and health status. Data presented for the entire group can obscure important differences in such

characteristics among the detailed Asian and Pacific Islander groups.

The Hispanic origin population grew by 53 percent during the 1980s, from about 15 million in 1980 to 22 million in 1990. This is an impressive growth for a population of this size (table 2). About half of the growth was due to immigration from Mexico and Latin America. Similar to the Asian and Pacific Islander population, the Hispanic origin population is made up of groups that vary widely in their characteristics. By 2000, the total Hispanic origin population is projected to be 31 million, and to more than triple its current size to exceed 80 million by 2050.

The increased diversity of our nation will make the data collection efforts more challenging in the future for both the census and public health surveillance systems. Each must meet the demands of being able to chronicle and respond to the changing racial and ethnic composition and subsequent data needs.

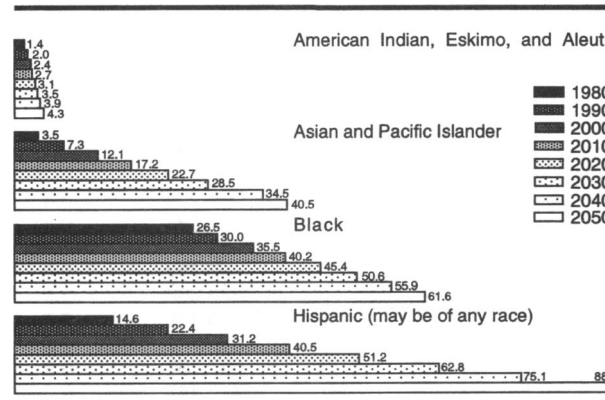
Relationship of Census and Public Health Data

Census data and some of the public health surveillance data systems are closely related and, in fact, interdependent. On one hand, census data are the denominator for birth, mortality, and morbidity rates. Census data also are used for designing sampling frameworks for such surveys as the National Health Interview Survey and National Maternal and Infant Health Survey, which provide data for public health surveillance. On the other hand, data on births and deaths obtained from vital records are essential elements for the development of the Bureau of the Census' population estimates and projections.

There are several possible explanations for results to differ by race and ethnicity when data from the Bureau of the Census and the public health surveillance systems are used as the denominator and numerator: different data collection methods, different content and format of the questions, and different definitions and classifications for race and ethnicity.

Different data collection methods. As noted previously, the Bureau of the Census uses self-identification in collecting data on race and ethnicity. Self-identification information is generally obtained by self-reporting or direct interview. Self-identification generally is the most socially accepted way of adequately collecting representative data on racial and ethnic populations. The Bureau of the

Population by race and Hispanic origin: 1980 to 2050
(numbers in millions)



NOTE: Data for 2000 to 2050 are projections.

Census evaluations indicated that, overall, self-identification results in more consistent reporting of race, particularly for persons of mixed racial parentage, than the enumerator's observation method (*2*). For example, observers may classify race differently from self-reports because of a number of factors, such as geographic familiarity with the racial group, the understanding of what groups are included in which racial category, andso forth. That is not to say that self-identification is not without problems, such as inconsistent reporting and misreporting.

Thacker and Berkelman in 1988 noted that a variety of methods—direct interview, interviewer's observation, reporting by health providers, and so forth—are used when providing data to the public health surveillance system (*6*). For example, a National Health Interview Survey interviewer asks respondents to identify their race. In contrast, for infectious diseases or for death records, for example, information on race may be obtained by observation or from next of kin.

Different methods for obtaining information on race can yield different responses for the same persons. A Bureau of the Census study conducted in 1970 comparing race reported by self-identification with enumerator's observation revealed relatively close agreement (more than 95 percent) for the White and Black populations, but substantial differences (only 73 percent agreement) for the "Other race" populations (Asians and American Indians) (*7*). Similar findings were reported in the areas of public health statistics. For example, Hahn reported that the race recorded by observers may not correspond to the race an individual may choose. His study of infants' birth records showed that in a majority of cases the race of the person

Table 1. Racial distribution of the U.S. population in 1990 and 1980

Race	1990		1980		Number change	Percent change
	Number	Percent	Number	Percent		
All persons	248,709,873	100.0	226,545,805	100.0	22,164,068	9.8
White	199,686,070	80.3	188,371,622	83.1	11,314,448	6.8
Black	29,986,060	12.1	26,495,025	11.7	3,491,035	13.2
American Indian, Eskimo, and Aleut	1,959,234	0.8	1,420,400	0.6	538,834	37.9
American Indian	1,878,285	0.8	1,364,033	0.6	514,252	37.7
Eskimo	57,152	0.0	42,162	0.0	14,990	35.6
Aleut	23,797	0.0	14,205	0.0	9,592	67.5
Asian and Pacific Islander	7,273,662	2.9	[1]3,500,439	1.5	3,773,223	107.8
Chinese	1,645,472	0.7	806,040	0.4	839,432	104.1
Filipino	1,406,770	0.6	774,652	0.3	632,118	81.6
Japanese	847,562	0.3	700,974	0.3	146,588	20.9
Asian Indian	815,447	0.3	361,531	0.2	453,916	125.6
Korean	798,849	0.3	354,593	0.2	444,256	125.3
Vietnamese	614,547	0.2	261,729	0.1	352,818	134.8
Hawaiian	211,014	0.1	166,814	0.1	44,200	26.5
Samoan	62,964	0.0	41,948	0.0	21,016	50.1
Guamanian	49,345	0.0	32,158	0.0	17,187	53.4
Other API	821,692	0.3	NA	NA	NA	NA
Other race	9,804,847	3.9	6,758,319	3.0	3,046,528	45.1

[1] The number of Asian and Pacific Islanders in this table is not entirely consistent with the 1990 counts. The 1980 count of 3,500,439 Asians and Pacific Islanders based on 100-percent tabulations includes only the 9 specific Asian and Pacific Islander groups listed separately in the 1980 race item. The 1980 total Asian and Pacific Islander population of 3,726,440 from sample tabulations is comparable to the 1990 count; the 1980 figure includes groups not listed separately in the race item on the 1980 census form.

NOTE: NA = not applicable.

SOURCE: U.S. Department of Commerce, Bureau of the Census: 1990 Census Summary Tape File 1, United States Summary and 1980 Census of Population, General Population Characteristics, United States Summary.

Table 2. U.S. population by Hispanic origin in 1990 and 1980

Hispanic origin	1990		1980		Number change	Percent change
	Number	Percent	Number	Percent		
Total population	248,709,873	100.0	226,545,805	100.0	22,164,068	9.8
Hispanic origin	22,354,059	9.0	14,608,673	6.4	7,745,386	53.0
Mexican	13,495,938	5.4	8,740,439	3.9	4,755,499	54.4
Puerto Rican	2,727,754	1.1	2,013,945	0.9	713,809	35.4
Cuban	1,043,932	0.4	803,226	0.4	240,706	30.0
Other Hispanic	5,086,435	2.0	3,051,063	1.3	2,035,372	66.7
Not Hispanic	226,355,814	91.0	211,937,132	93.6	14,418,682	6.8

SOURCE: U.S. Department of Commerce, Bureau of the Census: 1990 Census Summary Tape File 1, United States Summary and 1980 Census of Population, General Population Characteristics, United States Summary.

reported on the death certificate differed from that reported by the parent (*8,9*).

Different content and formats. Another explanation is the content of the item. In the census, the race and Hispanic origin items are asked separately, while in some health surveys, the race and Hispanic origin items are combined. The combination item generally lists White, non-Hispanic; Black, non-Hispanic; Native American; Asian or Pacific Islander; and Hispanic. For some persons, responses to a combined race and ethnic question will be different from responses to separate questions. For example, a person who identifies as Filipino race and Hispanic origin may report each identity in separate questions, but would have to choose one identity in the combined question.

Buehler and coworkers conducted a systematic study of racial and ethnic reporting of infectious diseases in 30 areas which indicated that the reporting and method of race-ethnicity identification varies widely by State and disease. Differences were attributed to such factors as the use of a combined race and ethnic question and to incomplete reporting. Buehler and coworkers concluded that race and ethnic categories used in the Epide-

miologic Surveillance Project were broad and encompassed diverse subgroups (*10*).

Differences may also result because of a person's perception of a particular question. For instance, the separate question on race in the census posed problems for some respondents. In the 1990 census, nearly 10 million persons reported in the other race category (table 3). About 98 percent of the 10 million were persons of Hispanic origin who either considered Hispanic as a race and did not identify with any of the specific racial categories or did not understand the race concept. About 2 percent were persons of mixed racial parentage who wanted to identify with more than one race.

Different definitions or classification systems. Different definitions or classifications can also affect the data. As stated previously, Statistical Directive No. 15 identifies four specific racial categories and it does not include an "Other race" category. Although the Census Bureau adheres to the general guidelines of the directive, the 1990 census race item included an "Other race" category for persons who did not identify with the specific racial categories. Most other data systems do not include an "Other race" category.

The 1990 census "Other race" category included write-in entries such as multiracial, multiethnic, mixed, interracial, Wesort (a part Indian, part Black, and part White group in southern Maryland), or Spanish-Hispanic type entries, such as Mexican and Puerto Rican. In contrast, the National Center for Health Statistics (NCHS) codes entries such as Hispanic, Mexican, and so forth, as White in the race item on vital records.

The Census Bureau prepared special 1990 census files with the "Other race" entries reassigned to one of the specified racial categories, that is, White, Black, American Indian, and so forth using information reported on the census form and other variables. These data files provide race data more comparable with NCHS and other statistical systems. However, some inconsistencies between the 1990 census and birth records remain because of different algorithms for classifying race.

Issues Raised by Evaluations of Census Data

In this section we present some of the pertinent issues raised in the evaluations of race and Hispanic origin data by the Bureau of the Census that may be relevant in reviewing, examining, and developing recommendations on surveillance data. The evaluations of the race and ethnic data show

Table 3. Race by Hispanic origin for the U.S. population in 1990

Race	Total population	Hispanic origin	Not of Hispanic origin
Percent by race			
Total	100.0	100.0	100.0
White	80.3	51.7	83.1
Black	12.1	3.4	12.9
American Indian, Eskimo, and Aleut	0.8	0.7	0.8
Asian and Pacific Islander	2.9	1.4	3.1
Other race	3.9	42.7	0.1
Percent by Hispanic origin[1]			
Total	100.0	9.0	91.0
White	100.0	5.8	94.2
Black	100.0	2.6	97.4
American Indian, Eskimo, and Aleut	100.0	8.4	91.6
Asian and Pacific Islander	100.0	4.2	95.8
Other race	100.0	97.5	2.5

[1] Persons of Hispanic origin may be of any race.

SOURCE: U.S. Department of Commerce, Bureau of the Census, 1990 Census Summary Tape File 2 and CP-1, General Population Characteristics. United States Summary.

that, overall, the questions on race and Hispanic origin included in the 1990 census performed satisfactorily. However, the evaluations show problems that affect data for some specific groups or geographic areas.

Identification of race. McKenney and Cresce noted that most of the population of the United States are able to report their race, given the question used in both forms. That is, 98 percent of the population enumerated in the 1990 census responded to the race item. However, evaluations from the 1990 census suggest that more respondents than in 1980 had difficulty in responding to this question (*2*).

Information from telephone inquiries during the data collection stage of the 1990 census, information on nonresponse to the race question, results of content reinterviews, and findings from cognitive research studies suggest that some persons had difficulty answering the race question. For example, telephone inquiries to the Census Bureau information lines during the data collection stages for the 1990 census indicated that a substantial number of respondents did not understand how to answer the race question. The majority of the inquiries were from persons who were confused as to how to report because national origin groups were listed in the race item; persons of Hispanic origin who felt that the race question or its categories were not relevant to them; and persons of mixed parentage, or parents of interracial or

'The Bureau of the Census has traditionally treated race and ethnicity as two separate concepts. The racial and ethnic classifications used by the Bureau of the Census generally adhere to the guidelines of the Federal Statistical Policy Directive No. 15 which was issued by the Office of Management and Budget in 1978.'

multi-racial children, who wanted to report their own race or the race of their children in more than one race category.

Based on an analysis of data from the 1990 Census Reinterview Study, McKenney and coworkers concluded that foreign born persons and Hispanics, both native and foreign born, had difficulty reporting in the race item (*11*). Cognitive research sponsored by the Census Bureau found that some Hispanics, particularly the foreign born, found the race question confusing and therefore, misreported in various categories.

Defining Hispanic. Cresce and coworkers stated that one of the issues that the Bureau of the Census faces is defining "Hispanic." This population, as defined by Federal Statistical Policy Directive No. 15, is composed of diverse groups that share a common language and some common traditions (*12*). However, not all persons who would belong in the universe, according to the directive, identify with the term "Spanish" or "Hispanic." Hayes-Bautista and Chapa argue for the use of "Latino," while others (Gimenez) suggest such terms as "Spanish American" (*13,14*).

Results from evaluation and focus group testing indicated that no one term or set of terms enjoy universal approval or understanding by either the Hispanic or non-Hispanic populations (*15*). For instance, in Massachusetts and Rhode Island, the Hispanic question attracts a positive response from a substantial number of persons of Portuguese descent who identify with the term "Hispanic." In summary, the use of terminology that is generally understood and accepted by a population is especially important in a system that uses self-identification, but it is also important for systems that rely on observation and reporting by an enumerator, health practitioner, or the like.

Consistency of responses. Consistency in reporting is measured in a variety of ways, such as comparing counts over time for a population group and comparing a response at reinterview with the original response from a census or survey for the same group of persons.

The identification of an individual with a particular race or ethnic group can fluctuate over time as a function of various social conditions or changing social realities. McKenney and Cresce noted that the increases in the American Indian population during the last two decades (72 percent between 1970 and 1980 and 38 percent between 1980 and 1990) are much greater than can be attributed to natural increase (*16*). Changing the methodology and the definition, improvements in the census, and improved outreach account for some of the increase in this population. Passel and Berman, in a study of this population, concluded that part of the increase between 1970 and 1980 was also due to a shift in self-identification. Persons who chose to report as White in previous censuses chose to change and report as American Indian in 1980 (*17*). Census information shows that about 6 to 9 million persons who report as White in the race item report American Indian as a single entry or in combination with another group in the ancestry item. Matthew Snipp indicated that the large number of persons with Indian ancestry, along with political factors, make it difficult to obtain consistent reporting on this population (*18*).

An evaluation study comparing responses in the 1990 census with those reported in a 1990 census reinterview for identical persons revealed considerable inconsistent reporting in the American Indian category. Most of the inconsistency among American Indians involved persons who identified as White in either the census or the reinterview (*11*).

At the Conference on Ethnicity in 1992, Stanley Lieberson and several other scholars suggested that the inconsistent reporting may be attributable, in part, to ethnic flux, which reflects the dynamic nature of the ethnicity and race concepts (*19*). Statistical agencies should determine whether the inconsistent reporting is attributable to bad question design or other failures.

Misreporting. Misreporting can affect data especially for relatively small populations or small geographic areas. Misreporting can occur in any data collection system because the respondents do not understand the intent or wording of the question. In the 1980 census, some non-Hispanic persons misreported in the "Mexican origin" or the

"Other Spanish/Hispanic" categories of the Hispanic origin item because they did not understand the terms "Spanish" or "Hispanic." Also, some non-Hispanics viewed the term "Mexican-Amer." in the Hispanic origin question as asking if they were "Mexican" or "American" and reported in the category to indicate they were "American." This misreporting occurred primarily in southeastern and northeastern States where the Hispanic population was sparse and, therefore, had a negligible effect upon the national data for this population. The Bureau of the Census modified the 1990 question on Hispanic origin to reduce this type of misreporting. Evaluation results from the 1990 census suggest that the misreporting in the Hispanic origin categories has declined from the 1980 levels, but it still occurs.

Evaluation results of the 1990 data also suggest some misreporting in the American Indian category of the race item. A review of household records for two States shows instances where parents reported themselves as Asian Indian, but reported their children as American Indian to indicate that the children were born in the United States. Forbes noted that Asian Indians may report in this category because they are adopting the label "Indian-American" (*20*).

Future Issues

As the Bureau of the Census plans the race and ethnic questions for the 2000 census, it is facing a number of theoretical and practical issues. Several issues that tend to affect all data systems are identified subsequently.

Overlapping concepts of race and ethnicity. The separate questions on race and ethnicity used by the Bureau of the Census allow persons to report both their "ethnic" and "racial" identity. This approach provides the most complete set of data for the racial and ethnic groups. However, some researchers and data users have suggested that we use a question that combines the race and Hispanic origin item. Some respondents do not view race and ethnicity as distinct. Rather, they view the race and ethnic questions (in particular, Hispanic origin) as asking for the same identity, and therefore, the race or Hispanic origin item is perceived as confusing or subject to various interpretations.

As discussed previously, the Bureau's research showed that some Hispanics identify themselves racially as Hispanic; others find the race question confusing and do not know how to respond (*15*). Yet, other Hispanics consider Hispanic as an ethnic group and their race as White, Black, or relate more to their Indian roots and identify their race as Indian or Mestizo. The two separate questions allow these persons to report both their ethnic identity and their racial identity.

Concept of race. McKenney and Cresce noted that some scholars and general data users argue that the race concept should be abandoned from all classifications because it is racist, ambiguous, and vague (*2*). The issue of the validity of a race question was thoroughly discussed at the Conference on Ethnicity, with considerable division on the issue. Yet there was a consensus that it was necessary to identify and collect data on those groups that have been traditionally identified in the race item. The experts provided several suggestions, ranging from maintaining the question as is, using alternative terminology for the question, to combining the race question with the ethnic questions. It is clear, however, that the Bureau of the Census will continue to collect data on the racial groups to meet the very important data needs.

Classification of persons of mixed race. One of the major issues facing the Bureau of the Census is the classification of persons of mixed racial parentage. Organizations representing persons of this population argue that census procedures, in fact all government statistical and administrative systems, do not allow them to report their true identity. Some researchers assert that persons of mixed racial parentage are an emerging racial-ethnic group and should be recognized in official statistics (*2*).

The Bureau of the Census asks all persons, including those of mixed racial parentage, to report one race. For the 1990 census, persons who could not provide a single response had the option of marking the "Other race" category and reporting entries such as "interracial" or "biracial." However, write-in responses of specific multiple entries such as "White-Black" or "Chinese-White" were assigned according to the first write-in. (In direct interviews, persons were asked to first self-report one race. If they could not, then the mother's race was suggested.)

Census data clearly indicated that the number of interracial couples has increased in recent decades, from 157,000 in 1960 to more than 1 million in 1992. The number of children in these unions also has increased from 100,000 to more than 1 million in 1991. The implication of the increasing number

'Census data clearly indicated that the number of interracial couples has increased in recent decades, from 157,000 in 1960 to more than 1 million in 1992. The number of children in these unions also has increased from 100,000 to more than 1 million in 1991.'

of interracial unions and of children in these unions are far reaching. For example, in the 1980 census, there were 229,000 children in American Indian and White unions. Of those children, nearly half identified with the race of the mother and half with the race of the father. That is, about 50 percent identified as White and 50 percent identified as American Indian (*21*). This is an example of how the classification can increase or decrease the number of children of a particular group. It's important to note that the National Center for Health Statistics now uses the mother's race to tabulate natality data. However, the Center does provide information on both the mother's and father's race so data users can use the race of either parent for natality data.

The classification of persons of mixed racial parentage is quite a challenge, considering the programs and data bases that require persons to be classified in one category.

Conclusion

This study reveals at least three relevant findings on the comparability of race and ethnic data across data systems. First, data from the Census Bureau and the public health surveillance data system are not always comparable because the systems do not always use the same data collection methods. The Census Bureau uses self-identification to obtain information on race and ethnicity. In the health systems, data for race and ethnicity are collected using a variety of methods, such as direct interview, observations, and third party identification. Different methods of collecting race and ethnic information yield results that are not always comparable.

Second, data are also not comparable because of the question content and format used. The Census Bureau uses two separate questions to gather data on race and Hispanic origin, while many of the health data systems use a combined race and Hispanic origin question. For some persons, responses to a combined question will be different from those to separate questions.

Third, data are not comparable because of different definitions or classifications. The 1990 census race item included an "Other race" category for persons who do not identify with one of the specified categories of the race item. The health systems generally do not include an "Other race" category.

To provide data more comparable to other data systems, the Census Bureau prepared a special 1990 census file with persons in the "Other race" category reassigned to a specific race category. However, some inconsistencies remain between census and other data sets because of different algorithms for assigning race.

This paper also describes some of the complexities of collecting and presenting data on race and ethnicity, based on the experiences of the Census Bureau. It discusses a number of issues that should be examined as the Bureau prepares for the 2000 census, and as statistical agencies seek to make data on race and ethnicity comparable across data systems. These issues include, for example, changing ethnic identity, consistency of responses for race and ethnicity, misreporting in racial and ethnic categories, overlapping concepts of race and ethnicity, and classification of persons of mixed race. Another factor to consider is the growing racial and ethnic diversity of the U. S. population.

Finally, an understanding of the relevant findings and issues is important because of the interdependence of the two systems. Because of this interdependence, changes or problems in one of these systems could affect the quality of the data in the other. In preparing for the 2000 census, the Census Bureau plans to conduct research and consult with a wide array of governmental and private data users to discuss issues raised in this paper as well as other relevant issues. Within the last several years, governmental agencies held a number of conferences and workshops, and Congress held hearings to address issues on race and ethnicity. All of these efforts are needed to ensure that race and ethnic data are of high quality and meet the changing data demands of our nation.

References

1. U.S. Department of Commerce, Office of Federal Statistical Policy and Standards: Federal Statistical Policy Directive No. 15—Race and ethnic standards for Federal statistics and administrative reporting. U.S. Government Printing Office, Washington, DC, May 1978.

2. McKenney, N. R., and Cresce, A. R.: Measurement of ethnicity in the United States: experiences of the U.S. Census Bureau. *In* Challenges of measuring an ethnic world: science, politics and reality. Proceedings of the Joint Canada—United States Conference on the Measurement of Ethnicity. April 1-3, 1992. U.S. Government Printing Office, Washington, DC, 1993, pp. 173-222.
3. Public Health Service: Healthy people 2000: national health promotion and disease prevention objectives. DHHS Publication No. (PHS) 91-50212. Office of the Assistant Secretary for Health, Office of Disease Prevention and Health Promotion. U.S. Government Printing Office, Washington, DC, 1990.
4. U.S. Department of Health and Human Services, U.S. Public Health Service: Improving minority health statistics. Report of the PHS Task Force on Minority Health Data. U.S. Government Printing Office, Washington, DC, May 1992.
5. U.S. Department of Commerce, U.S. Bureau of the Census: Population projections of the United States, by age, sex, race, and Hispanic origin: 1992 to 2050. Current Population Reports, Series P25-1104. U.S. Government Printing Office, Washington, DC, November 1993.
6. Thacker, S. B., and Berkelman, R. L.: Public health surveillance in the United States. Epidemiologic Rev 10: 164-190 (1988).
7. U.S. Department of Commerce, U.S. Bureau of the Census: 1980 census of population. Accuracy of data for selected population characteristics as measured by the 1970 CPS- Census match, PHC (E)-1. U.S. Government Printing Office, Washington, DC, 1982.
8. Hahn, R. A.: Inconsistencies in coding of race and ethnicity between birth and death in U.S. infants: a new look at infant mortality, 1983 through 1985. JAMA 267: 259-263, Jan. 8, 1992.
9. Hahn, R. A.: The state of federal health statistics on racial and ethnic groups. JAMA 267:268-271, Jan. 8, 1992.
10. Buehler, J. W., Stroup, D. F., Klaucke, D. N., and Berkelman, R. L.: The reporting of race and ethnicity in the national notifiable diseases surveillance system. Public Health Rep 104:457-465, September-October 1989.
11. McKenney, N. R., Bennett, C. E., Harrison, R. J., and del Pinal, J.: Evaluating racial and ethnic reporting in the 1990 census. Paper presented at the annual meeting of the American Statistical Association, San Francisco, CA, Aug. 8-12, 1993.
12. Cresce, A. R., Lapham, S. J., and Rolark, S. J.: Preliminary evaluation of data from the race and ethnic origin questions in the 1990 Census. Paper presented at the annual meeting of the American Statistical Association, Boston, MA, Aug. 10-13, 1992.
13. Hayes-Bautista, D. E., and Chapa, J.: Latino terminology: conceptual bases for standardized terminology. Am J Public Health 77:61-68 (1987).
14. Gimenez, M. E.: Latino/'Hispanic'-who needs a name: the case against a standardized terminology. Int J Health Serv 19:557-571 (1989).
15. McKenney, N. R., Cresce, A. R., and Johnson, P. A.: Development of the race and ethnic items for the 1990 census. Paper presented at the annual meeting of the Population Association of America, New Orleans, LA, Apr. 21-23, 1988.
16. McKenney, N. R., and Cresce, A. R.: Identification of ethnicity in the United States: The census bureau experience. Paper presented at the annual meeting of the Population Association of America, Toronto, Canada, May 3-5, 1990.
17. Passel, J. S., and Berman, P.: Quality of 1980 census data for American Indians. Soc Bio 33:163-182 (1986).
18. Snipp, M. C.: Who are American Indians? Some observations about the perils and pitfalls of data for race and ethnicity. Popul Res Policy Rev 5:237-252 (1986).
19. Lieberson, S., and Waters, M.: Ethnic groups in flux: the changing ethnic responses of American whites. Ann Am Acad Political Soc Sci 487:79-91 (1986).
20. Forbes, J. D.: The manipulation of race, caste, and identity: Classifying Afro-Americans, Native Americans and red-black people. J Ethnic Studies 17:1-51 (1990).
21. Bennett, C. E., and Robinson, B. G.: Racial classification issues concerning children in mixed race households. Paper presented at the annual meeting of the American Statistical Association, Fort Lauderdale, FL, Jan. 3-5, 1993.